Cognition 159 (2017) 139–155

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Spoken word recognition in young tone language learners: Age-dependent effects of segmental and suprasegmental variation

Weiyi Ma [a,b,*,1], Peng Zhou [c,1], Leher Singh [d], Liqun Gao [e]

[a] ARC Centre of Excellence in Cognition and Its Disorders, Macquarie University, Sydney 2109, Australia
[b] School of Linguistics and Literature, University of Electronic Science and Technology of China, Chengdu 610000, China
[c] Department of Foreign Languages and Literatures, Tsinghua University, Beijing 100084, China
[d] Department of Psychology, National University of Singapore, 117570, Singapore
[e] Centre for Speech, Language and the Brain, Beijing Language and Culture University, 100066 Beijing, China

ARTICLE INFO

Article history:
Received 16 September 2015
Revised 16 November 2016
Accepted 28 November 2016
Available online 10 December 2016

Keywords:
Speech
Language acquisition
Word recognition
Lexical tone
Vowel

ABSTRACT

The majority of the world's languages rely on both segmental (vowels, consonants) and suprasegmental (lexical tones) information to contrast the meanings of individual words. However, research on early language development has mostly focused on the acquisition of vowel-consonant languages. Developmental research comparing sensitivity to segmental and suprasegmental features in young tone learners is extremely rare. This study examined 2- and 3-year-old monolingual tone learners' sensitivity to vowels and tones. Experiment 1a tested the influence of vowel and tone variation on novel word learning. Vowel and tone variation hindered word recognition efficiency in both age groups. However, tone variation hindered word recognition accuracy only in 2-year-olds, while 3-year-olds were insensitive to tone variation. Experiment 1b demonstrated that 3-year-olds could use tones to learn new words when additional support was provided, and additionally, that Tone 3 words were exceptionally difficult to learn. Experiment 2 confirmed a similar pattern of results when children were presented with familiar words. This study is the first to show that despite the importance of tones in tone languages, vowels maintain primacy over tones in young children's word recognition and that tone sensitivity in word learning and recognition changes between 2 and 3 years of age. The findings suggest that early lexical processes are more tightly constrained by variation in vowels than by tones.

1. Introduction

Children are exposed to numerous sources of phonemic and non-phonemic variation in speech, such as vowels, consonants, tones, and intonations. However, only a subset of this variation distinguishes the meanings of words in a language, placing onus on the learner to discern lexically relevant cues. There is an extensive body of literature on how learners accomplish this, most of which has focused on children learning vowel-consonant languages such as English and French (e.g., Mani & Plunkett, 2007; Nazzi, 2005; Quam & Swingley, 2010; Swingley & Aslin, 2000, 2002; White & Morgan, 2008; but see Mattock & Burnham, 2006; Singh & Foong, 2012). By contrast, approximately 60–70% of the world's languages are tone languages, relying on both segments (vowels, consonants) and lexical tones (hereafter, tones) to distinguish word identity (Yip, 2002) or using tones to indicate grammatical features such as tense, aspect, and case (e.g., Demuth, 1993). Furthermore, at least half of the world's population speaks a tone language natively (Fromkin, 1978). Thus, extant evidence and theories on early language development are based on a minority of languages and language learners. A major theoretical gap therefore exists in our understanding of children's sensitivity to tones in relation to segments. This study examines children's sensitivity to tone and segment variation in spoken word recognition in tone languages.

Mandarin, the most widely spoken tone language, distinguishes words using four basic tones (in addition to a fifth, "neutral" tone, which is typically used on weak syllables) (Liu & Samuel, 2004). Thus, *ma* can mean mother (T1: high level), hemp (T2: rising), horse (T3: dipping), and curse (T4: falling). Although several acoustic attributes, such as vowel duration (Gandour & Harshman, 1978), amplitude (Garding, Kratochvil, & Svantesson, 1986), and voice quality (Gottfried & Suiter, 1997) may define changes in tone, the primary cue to tone perception is the fundamental frequency

* Corresponding author at: ARC Centre of Excellence in Cognition and Its Disorders, Department of Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia.
E-mail address: weiyi.ma@mq.edu.au (W. Ma).
[1] WM and PZ contributed equally to this work.

http://dx.doi.org/10.1016/j.cognition.2016.11.011
0010-0277/

contour (Khouw & Ciocca, 2007). At the outset, we review past studies on the acquisition of lexical tones (see Table 1 for a summary of research on tone acquisition in Chinese-reared young children).

Tone sensitivity is evident very early in development. Recent research has evinced tone sensitivity in Mandarin- and Cantonese-reared infants as early as 4 months of age (Yeung, Chen, & Werker, 2013). Even earlier than that, neural responses to tone distinctions have been observed even in newborns (Cheng et al., 2013). Furthermore, both English- and Mandarin-reared 6-month-olds are able to discriminate Thai tones, but later in infancy, tone discrimination was only observed in Mandarin-reared infants (Mattock & Burnham, 2006). A similar pattern of narrowing for tones in non-tone language-exposed infants was found in French-reared infants (Mattock, Molnar, Polka, & Burnham, 2008, but see Liu & Kager, 2014 on Dutch-reared infants). In a study of spoken word recognition of tone-bearing words, Singh and Foong (2012) demonstrated that bilingual Mandarin-English infants integrated lexical tones in Mandarin but not in English at 11 months of age. These findings are aligned with a widely reported development – perceptual narrowing – where young infants initially perceive differences between a wide range of phonetic contrasts, then selectively narrow down their sensitivity to native contrasts (e.g., Gervain & Mehler, 2010; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). However, in a study of tone discrimination between 5 and 18 months, Liu and Kager (2014) reported a different pattern of attunement not previously observed for segments: Dutch learning infants showed a transient decline in tone sensitivity between 5 and 9 months, followed by an increase in sensitivity to tone contrasts between 9 and 18 months. This suggests that the decline in tone sensitivity observed at 9 months (e.g. Mattock & Burnham, 2006; Yeung et al., 2013) may reflect a transient process that may later reverse. These findings suggest that normative developmental processes, such as perceptual narrowing, predominantly drawn from segments may not apply to lexical tones.

Do Mandarin-learning children integrate tones in word learning and word recognition? One approach to understanding monolingual Mandarin-learning children's knowledge of tone contrasts has been to scrutinize their production of tones. Based on adults' coding of children's tone production, Hua and Dodd (2000) reported early mastery of tone contrasts in Mandarin-learning 1.5- to 4.5-year-olds. Furthermore, according to parental reports of their children's productive vocabulary using a word checklist that contains common words children may produce, monolingual Mandarin learners produced minimal pairs of words differing by tones in their vocabularies before the age of 3 (Tardif, Fletcher, Zhang, Liang, & Zuo, 2008). However, adult listeners' interpretations (Hua & Dodd, 2000) and parental reports (Tardif et al., 2008) of tone production may have been influenced by their use of semantic, syntactic, and contextual cues in identifying children's tone production.

Research in the production of tones suggests that while Mandarin tones emerge early in production, the maturation of tone

Table 1
A summary of research on tone acquisition in Chinese-reared young children.

Study	Participants	Stimuli	Tasks	Major findings
Burnham et al. (2011)	5-, 7-, and 11-month-olds (Cantonese learners)[a]	Phone block (e.g., /jok3/, /jot3/, /jap3/), Tone block (e.g., /p^hak3/, /sak3/, /kak1/)	Odd-one-out task (aurally presented; the odd one is underlined)	Children were more sensitive to segment variation than tone variation
Cheng et al. (2013)	Newborns, 6-month-olds	T1/3 (highly distinctive), T2/3 (less distinctive)	Mismatch responses[b]	The T1/3 pair elicited mismatches response in both age groups, while the T2/3 pair did so only in the 6-month-olds
Ciocca and Lui (2003)	4-, 6-, and 10 years old (Cantonese learners)	Tone minimal pair contrasts	Picture-image matching task	Performance improved with age. Children's performance reached adult levels at 10 years of age
Hua and Dodd (2000)	1.5- to 4.5-year-olds	Elicited speech production	Picture naming and description tasks	Tones were acquired before segments
Li and Thompson (1977)	1.5- to 3-year-olds	Elicited speech production (longitudinal)	Picture naming and description tasks	Tones were acquired before segments. Rising and dipping tones were acquired later
Mattock and Burnham (2006)	6- and 9-month-olds	/ba-rising/, /ba-falling/ (Thai tones)	Conditioned head-turn procedure[c]	Both 6- and 9-month-olds were sensitive to the tone contrasts
Singh and Foong (2012)	7.5-, 9-, and 11-month-olds (bilinguals)	Familiar words (e.g., /bei1/, /tou2/) and tone changes	Conditioned head-turn procedure[b]	Children started to use tones in distinguish word identity in Mandarin at 11 months of age
Singh et al. (2014)	18- and 24-month-olds (bilinguals)	Novel words (e.g., /leng2/), tone and vowel changes	Mispronunciation paradigm in IPLP	Both age groups were highly, equivalently sensitive to vowel and tone variation
Singh et al. (2015)	2.5- to 4.5-year-olds (bilinguals)	Familiar words and tone and segment changes	Mispronunciation paradigm in IPLP	Older children were more sensitive to segments than tones
Wong (2012a, 2012b, 2013); Wong et al. (2005)	3-, 4, and 5-year-olds	Elicited speech production in a picture naming task	Analysis of the low-pass filtered speech production	Children did not produce adult-like tones. The ease of tone acquisition differed across tones based on both perception and production data
Yeung, Chen, and Werker (2013)	4- and 9-month-old Mandarin and Cantonese learners	/$tɕ^h$i/-rising tone, /$tɕ^h$i/-mid-level tone (Cantonese tones)	Children were first familiarized with a tone, and then were exposed to tone changes in the test	Sensitivity emerged first for tones then vowels and consonants

[a] Unless otherwise indicated, the children are Mandarin learners.
[b] A component of the ERP to an odd stimulus in a sequence of stimuli.
[c] This procedure makes use of an infant's differential preference for a given side as an indication of a preference for, or a familiarity with, the input or speech associated with that side.

production continues well beyond 3 years of age, sometimes extending to 10 years of age (Ciocca & Lui, 2003). Studies investigating Mandarin tone productions have demonstrated that 3-, 4-, and 5-year-olds did not produce adult-like tones as indicated by analyses of the low-pass filtered tone production, which eliminates lexical information but retains tone information (Wong, 2012a, 2012b, 2013). It is potentially revealing that even tones correctly identified by adult listeners were phonetically distinct from adults' tone productions (Wong, 2012a). Furthermore, research on 1.5-, 3-, 4-, and 5-year-olds' tone productions consistently revealed exceptional difficulty in acquiring T3 compared with other tones (e.g., Li & Thompson, 1977; Wong, 2012a, 2012b, 2013). Research on tone perception has demonstrated a complementary finding: Mandarin-learning 3-year-olds presented with a minimal pair of words differing by tones (e.g., *tang1* [soup] vs. *tang2* [candy]) (Wong, Schwartz, & Jenkins, 2005) were much less accurate in recognizing T3 relative to T1, T2, and T4.

Taken together, these findings point to several conclusions about tone development. First, children exhibit early perceptual sensitivity to tones attested across different tone languages. This sensitivity is maintained through infancy and beyond (Mattock & Burnham, 2006; Yeung et al., 2013). Production and perception data suggest that Mandarin learners can use most Mandarin tones to distinguish words at 3 years of age, although T3 is exceptionally difficult to acquire. Tone acquisition is therefore a protracted process. However, all these studies have focused on the acquisition of tones, and have generally not directly compared sensitivity to tones with sensitivity to segments.

Given that tone languages rely on both segments and tones to define word identity, it is possible that tone speakers are comparably sensitive to segments and tones in word recognition. However, Mandarin-reared infants have been shown to discriminate tones before attuning to native segment contrasts (Yeung et al., 2013, see Table 1) in an auditory discrimination paradigm. Research on Mandarin-reared 1.5- to 4.5-year-olds' speech production suggests earlier mastery of tones than segments (Hua & Dodd, 2000; Li & Thompson, 1977). By contrast, research comparing phonological and tonological awareness in school-aged children suggests greater segmental than tone awareness in 5-, 7-, and 11-year-old Cantonese speakers' word processing (Burnham et al., 2011). Furthermore, experimental studies with adult tone language speakers are divided 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). It remains to be seen whether the development of sensitivity to tones and segments proceeds asynchronously.

To our knowledge, there are only two published studies to date investigating relative sensitivity to segments and tones in early lexical processing. Using the Intermodal Preferential Looking Paradigm (IPLP – Golinkoff, Ma, Song, & Hirsh-Pasek, 2013), Singh, Hui, Chan, and Golinkoff (2014, Experiment 2a) recruited 18- and 24-month-old Mandarin-English bilingual children, exposed primarily to Mandarin language input. Children were first familiarized with the associations of novel objects and labels (e.g., *It is a leng2!* [The speech stimuli were presented in Mandarin]). The objects were then presented side-by-side in a split screen display accompanied by pre-recorded speech stimuli, directing children to look at one of them. The target word was either correctly pronounced (CP) or mispronounced with tone variation (MP-tone: *leng4*) or vowel variation (MP-vowel: *ling2*). In this paradigm, sensitivity to tone and vowel variation was indicated by a significant reduction in visual fixation to target images on MP trials compared to CP trials (Mani & Plunkett, 2007; Swingley & Aslin, 2000; White & Morgan, 2008). Both age groups preferentially fixated the target images on CP trials but not on MP-tone or MP-vowel trials, suggesting that children recognized correct pronunciations and rejected tone and vowel substitutions in equal measure. Thus, the children appeared to be equivalently sensitive to vowels and tones in word recognition.

Singh, Goh, and Wewalaarachchi (2015) recently examined 3- and 4.5-year-old Mandarin-English bilinguals' sensitivity to tone and segment variation in recognizing *familiar* words. Using the IPLP to assess sensitivity to mispronunciations of known words, children were shown two images in a split screen display. One of the words was labeled correctly or with a tone, vowel, or consonant substitution. Each test trial was divided into a pre-target phase and a post-target phase (e.g., Look! That is a [target word]). A significant increase in looking time at the target across phases was interpreted as a measure of children's understanding of the target word. While younger children were more sensitive to tone variation than vowel and consonant variation, older children were more sensitive to vowel and consonant variation than tone variation. Thus, older children demonstrated a reduction in sensitivity to tones in word recognition compared to younger children. Furthermore, vowels and consonants constrained word recognition more readily than tones in 4.5-year-old tone learners, providing support for the primacy of segments over tones as children mature.

Although there have been previous investigations of tone sensitivity relative to segments in children, several questions remain. First, older children might be more familiar with the test items than young children in Singh et al.'s (2015) study. Thus, it is unclear whether this finding was due to older children's phonological and lexical knowledge or whether the 'downweighting' of tones over time reflects an independent developmental process. To address this confound, research should examine word recognition using both novel and familiar words. Second, the age-related reduction in tone sensitivity reported by Singh et al. (2015) may have been due to the fact that participants were bilingual in English and Mandarin (Singh et al., 2015). It is unclear whether *monolingual* tone language learners would demonstrate the same course of development. The present study investigated tone and vowel sensitivity in monolingual Mandarin learners, using both familiar and novel words.

In the present study, we sought to investigate whether the previously reported decline in sensitivity to lexical tones (Singh et al., 2015) would be observed in monolingual tone language learners. Children were recruited at 2 and 3 years of age to document developmental trends during a particularly aggressive phase of vocabulary growth (Fenson et al., 1994). Furthermore, tone knowledge at 3 years of age is present but fragile for particular tones (e.g., T3), making it an ideal age to explore emergent developmental sensitivity to tones (Wong et al., 2005). We compared vowel and tone processing primarily because tones are instantiated on vowels (Lin, Wang, & Shu, 2013) and bear several structural similarities to vowels (Howie, 1976).

This study consists of three experiments. Experiment 1a examined children's sensitivity to vowels and tones in recognizing novel words. Experiment 1b examined factors that may facilitate effective integration of tones in word learning in 3-year-olds, specifically, whether exposure to minimal pairs differing in tone facilitates tone word *learning*. Furthermore, Experiment 1b also examined the learnability of different tones. Finally, using familiar words, Experiment 2 investigated whether the reduced sensitivity to tones in older children was influenced by their familiarity with target words.

Separate groups of monolingual Mandarin learners were tested at Beijing Language and Culture University (BLCU) Kindergarten and Nursery. All children were healthy and had no history of auditory or visual impairments. This study used the IPLP, where children's visual fixation time served as the dependent variable. Experiments 1a and 2 used a mispronunciation paradigm, a standardly used experimental paradigm used to investigate children's phonological sensitivity. 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, 2005; Havy & Nazzi, 2009; Mani & Plunkett, 2007; Swingley & Aslin, 2000, 2002; White & Morgan, 2008).

2. Experiment 1a

Experiment 1a examined young tone learners' sensitivity to vowels and tones in learning novel words in 2- and 3-year old children.

3. Methods

3.1. Participants

The sample of participants consisted of 24 2-year-old (M = 25.20 months, range = 23.5–27.00, 12 girls) and 28 3-year-old children (M = 38.42 months, range = 32.33–42.10, 13 girls). Five additional children were excluded from the final sample because of failure to complete the test (n = 3), inattentiveness (n = 1, overall attention during an experimental session was less than 20% of the length of the session), and 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).

3.2. 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 39in LED TV monitor at a distance of 1 m from the center of the screen. Visual stimuli were displayed in the center (during training) or to the left and right of the screen (during test) at eye level. Images of four novel objects were used as stimuli (Horst, 2009) (Table 2). Auditory stimuli were presented through internal speakers within the TV monitor. A hidden camera recorded children's visual fixation to the display. Video recordings were then coded offline.

3.2.1. Speech stimuli

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). Stimuli consisted of two sets of minimal pairs, /sɪe/ *vs.* /gɪe/ and /tɪa/ *vs.* /pɪa/, all of which are non-words in Mandarin, although all are phonotactically legal. The labels were pronounced with T2 (rising) in the training and on correct pronunciation (CP) trials (/sɪe2/, /gɪe2/, /tɪa2/, /pɪa2/). On tone mispronunciation (MP-tone) trials, /sɪe2/ and /gɪe2/ were mispronounced as /sɪe4/ and /gɪe4/. T2 and T4 (falling) were used because they are highly discriminable for native adult (Shen & Lin, 1991; So & Best, 2010) and child (Li & Thompson, 1977) speakers of Mandarin. On vowel mispronunciation (MP-vowel) trials, /tɪa2/ and /pɪa2/ were mispronounced as /tɪu2/ and /pɪu2/. Both the tone and vowel changes are also non-words in Mandarin. Fig. 1 depicts the tone contours of the word tokens.

To validate the stimuli, 20 adult native Mandarin speakers completed an auditory perception task, consisting of four trials. On each trial, participants heard a pair of sounds (CP and MP-tone, or CP and MP-vowel) presented sequentially, separated by a 1 s pause. They rated (1) the perceptual discriminability between the two sounds on a 1–7 scale (1 = not discriminable, 4 = discriminable, 7 = easily discriminable), and (2) *if* the two sounds were Mandarin words, adult listeners rated the likelihood of their requiring two different labels in Mandarin on a 1–7 scale (1 = highly unlikely, 4 = likely, 7 = highly likely). Both MP-tone (M = 5.30, SD = 1.42) and MP-vowel sounds (M = 5.55, SD = 1.23) were discriminable from CP sounds. A paired-samples t test showed that discriminability ratings did not differ between the CP–MP-tone pair and the CP–MP-vowel pair (p = 0.53), suggesting that vowel and tone changes were similarly discriminable compared to correct pronunciation. However, the MP-vowel sound (M = 5.85, SD = 0.99) was rated more likely to induce a change of label than the MP-tone sound (M = 3.70, SD = 1.72; $t(19)$ = 4.85, p = 0.0001, Cohen's d = 1.08), suggesting that vowels constrained the definition of word identity more readily than tones in Mandarin even when perceptual discriminability was controlled. A second group of 20 adult native Mandarin speakers completed a tone perception task, where they heard a list of target words and were asked to identify its tone by marking T1, T2, T3, or T4. All participants correctly identified the tones of the eight sounds, verifying their tone assignment.

3.3. Procedure

The experimental procedure was almost identical to that of Singh et al. (2014). An experiment consisted of a tone block and a vowel block. The presentation order of the two blocks was counterbalanced across participants. Each block consisted of four phases (Table 2). Before each trial, participants saw an attention getter (e.g., a giggling child) in the center of the screen that drew their attention there so that they were well positioned for voluntary left–right gaze shifts during each test trial.

During the *salience phase* (one trial), participants were shown two novel objects side-by-side without accompanying auditory stimuli. During the *training phase* (four trials), participants were shown animated videos of the same two novel objects as in the salience phase, each paired with one of the two training sentences. Each object was displayed twice in alternating order. During the salience phase, objects were static but during the training phase, objects moved around the screen to engage the children's attention. This was followed by a test phase. During the *test phase*, participants were presented with the static version of the two novel objects side-by-side and were directed to look at one of them. The left-right position of the objects was consistent across the salience and test phases within participants and counterbalanced across participants. There were two CP trials and two MP trials in each block. The presentation order of CP and MP test trials was counterbalanced. The CP and MP trials were separated by a *reminder phase* (two trials) that allowed participants additional opportunity to learn the pairings of the novel objects and words. During each reminder trial, participants were presented with one of the novel objects and a reminder of the object's label. The order of the two reminder trials was counterbalanced across participants. The experimental session lasted approximately 5.5 min.

3.3.1. Coding and data analysis

Using SuperCoder (Hollich, 2005), participants' eye movements were coded by a blind coder frame-by-frame at a frame rate of 30 frames per second. A second coder recoded 20% of the data. Intercoder agreement was 98%. Trials where participant's attended to the screen for 20% of the trial length or less were excluded from

Table 2
An example (the tone block) of trial sequences for the visual and speech stimuli in Experiment 1a.

	Left side	Center	Right side	Speech stimuli
Salience				No audio
Training: Animations of objects (4 trials, 24 s each). The 2 trials repeat				看！这是一个sɪe2. 这里有个sɪe2！看！这是个sɪe2. 看sɪe2在动. 这个sɪe2从这边过来了. Sɪe2, 这个东西叫sɪe2. 这是一个sɪe2. (Look! This is a sɪe2. There is a sɪe2. Look! This is a sɪe2. Look at Sɪe2 moving. This sɪe2 is moving over here. Sɪe2, it is called sɪe2. This is a sɪe2.)
				看！这是一个gɪe2. 这里有个gɪe2！看！这是个gɪe2. 看gɪe2在动. 这个gɪe2从这边过来了. Gɪe2, 这个东西叫gɪe2. 这是一个gɪe2. (Look! This is a gɪe2. There is a gɪe2. Look! This is a gɪe2. Look at gɪe2 moving. This gɪe2 is moving over here. Gɪe2, it is called gɪe2. This is a gɪe2.)
CP Test: (2 trials, 7 s each)				**CP trial 1**: Sɪe2！快看, sɪe2. 看！哪一个是sɪe2? 快看 sɪe2！(Sɪe2! Look quickly at sɪe2. Look, which one is sɪe2? Look quickly at sɪe2.)
				CP trial 2: Gɪe2！快看, gɪe2. 看！哪一个是gɪe2? 快看gɪe2！(Gɪe2! Look quickly at gɪe2. Look, which one is gɪe2? Look quickly at gɪe2.)
Reminder: (2 trials, 7 s each				Sɪe2, 这个东西叫sɪe2. 这是一个sɪe2. (Sɪe2, it is called sɪe2. This is a sɪe2.)
				Gɪe2, 这个东西叫gɪe2. 这是一个gɪe2. (Gɪe2, it is called gɪe2. This is a gɪe2.)
MP Test: (2 trials, 7 s each)				**MP trial 1**: Sɪe4！快看, sɪe4. 看！哪一个是sɪe4? 快看 sɪe4！(Sɪe4! Look quickly at sɪe4. Look, which one is sɪe4? Look quickly at sɪe4.)
				MP trial 2: Gɪe4！快看, gɪe4. 看！哪一个是gɪe4? 快看gɪe4！(Gɪe4! Look quickly at gɪe4. Look, which one is gɪe4? Look quickly at gɪe4.)

Note: The vowel block followed the same procedure except that another pair of novel objects and word tokens (/tɪa2/ - /pɪa2/) were used, and the word tokens were mispronounced as /tɪu2/ and /pɪu2/ on MP trials.

data analysis (Quam & Swingley, 2010). This resulted in a total of 10 trials being excluded across all participants. A total of 406 test trials were analyzed. Based on analysis standards set by prior research (e.g., Mani & Plunkett, 2007; Swingley & Aslin, 2000), visual fixation on a test trial was calculated from 367 ms after the onset 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. It should be noted that the results reported here remain the same when other minima (e.g., 200 and 400 ms) were used. Following Singh et al. (2014), the target word was presented first at the very beginning of a test trial and a subsequent three times within each test trial. Thus, participants' visual fixation was calculated until the end of the trial to capture all four repetitions of the target word.

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 total look to the target (target fixation) 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. Response latency and target fixation are measures of efficiency and accuracy of word identification.

4. Results and discussion

4.1. Latency

We analyzed a subset of test trials (105 trials (55%) for 2-year-olds; 132 trials (59%) for 3-year-olds) where children's initial visual fixation was directed to the distractor 367 ms after the onset of the target word. For each participant, we calculated an average latency for CP trials across blocks and average latencies for MP-tone and MP-vowel trials respectively. Participants whose initial visual fixations were on the target image within a pronunciation type were excluded from the analysis. A 3 (pronunciation type: CP/MP-tone/MP-vowel) × 2 (age group) repeated-measures ANOVA on latency revealed a main effect of pronunciation type ($F(2,76) = 5.75$, $p = 0.005$, $\eta_p^2 = 0.13$) (Fig. 2). The main effect of age ($p = 0.83$) and pronunciation type × age interaction ($p = 0.57$) were not significant. With age groups combined, planned paired-samples t tests were conducted to compare latency for each pronunciation types. Bonferroni correction was applied for multiple comparisons. Latency was shorter on CP ($M = 0.65$ s, $SD = 0.57$) than on MP-tone ($M = 1.38$ s, $SD = 1.59$; $t(47) = 3.06$, $p = 0.004$, Cohen's d = 0.44) and MP-vowel trials ($M = 1.68$ s, $SD = 1.74$; $t(42) = 4.45$, $p = 0.0001$, Cohen's d = 0.69). However, latency did not differ between MP-tone and MP-vowel trials ($p = 0.55$). Thus, participants oriented to target images faster when they heard correct pronunciation than when they heard tone or vowel mispronunciation. Tone and vowel mispronunciation hindered word recognition efficiency in both age groups.

4.2. Proportion of total look to the target (target fixation)

For each participant, we calculated an average target fixation for CP trials across the two blocks, and average target fixations for MP-tone and MP-vowel trials respectively. A 3 (pronunciation type: CP/MP-tone/MP-vowel) × 2 (age group) repeated-measures ANOVA on target fixation revealed a main effect of pronunciation type ($F(2,100) = 7.63$, $p = 0.001$, $\eta_p^2 = 0.13$) and a pronunciation type × age group interaction ($F(2,100) = 3.24$, $p = 0.04$, $\eta_p^2 = 0.06$). This suggests that target fixation differed by pronunciation type and that the two age groups had different patterns of target fixation (Fig. 3). Thus, the two age groups were analyzed separately.

4.2.1. Two-year-old participants

Planned paired-samples t tests compared target fixation among pronunciation types. Target fixation was greater for CP ($M = 0.60$, $SD = 0.09$) than for MP-tone ($M = 0.47$, $SD = 0.15$; $t(23) = 3.68$, $p = 0.001$, Cohen's d = 0.75) and MP-vowel trials ($M = 0.50$, $SD = 0.14$; $t(23) = 3.00$, $p = 0.006$, Cohen's d = 0.61),

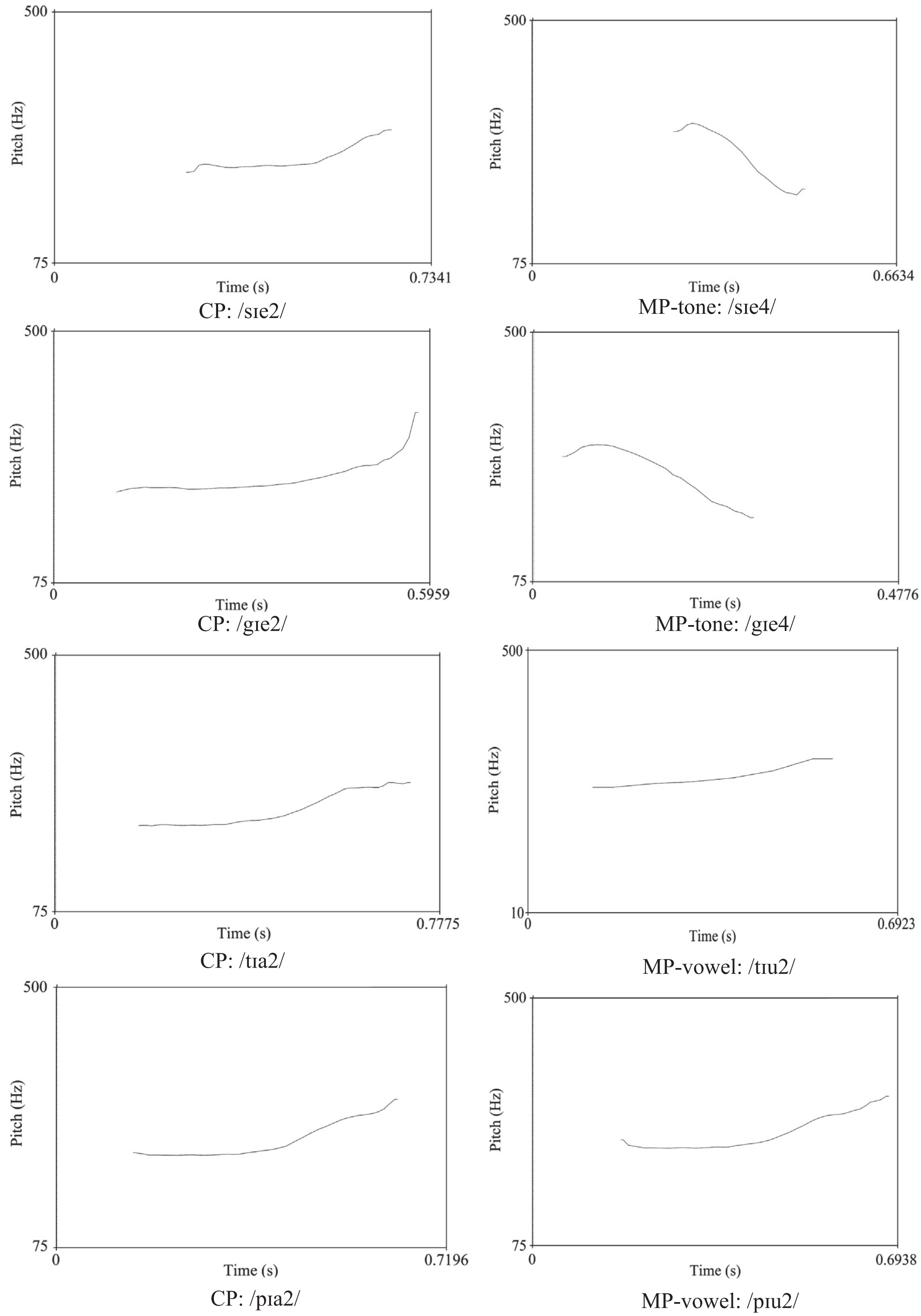


Fig. 1. Tone contours of the exemplar word tokens used in Experiment 1a.

suggesting that both tone and vowel mispronunciation hindered word recognition accuracy. Then, target fixation was compared to chance (0.50) for each pronunciation type. Successful word learning is conventionally indicated by fixation to target that exceeds chance (0.50). Separate one-sample t tests found that target fixation was greater than chance on CP trials ($t(23) = 5.26$, $p = 0.0001$, Cohen's $d = 2.15$), but did not differ from chance on either MP-tone ($p = 0.28$) or MP-vowel trials ($p = 0.94$). Thus, 2-year-olds recognized the words when they were correctly pronounced but not when they were mispronounced by tone or by vowel, suggesting that they were equivalently sensitive to vowel and tone variation.

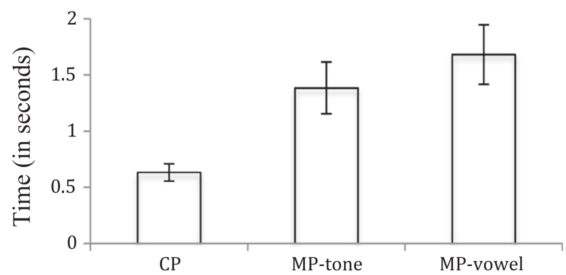


Fig. 2. Latency (with the two age groups combined) in Experiment 1a. Children oriented faster to target images on CP than on MP-tone and MP-vowel trials. Error bars reflect SEM.

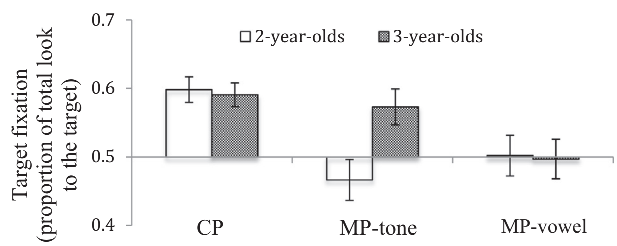


Fig. 3. Target fixation on CP, MP-tone, and MP-vowel trials in Experiment 1a. In the 2-year-olds, target fixation was greater on CP than on MP-tone and MP-vowel trials. Target fixation was greater than chance only on CP trials. In the 3-year-olds, target fixation was greater on CP than MP-vowel trials. Target fixation was greater than chance on CP and MP-tone trials. Error bars reflect SEM.

4.2.2. Three-year-old participants

Planned paired-samples t tests revealed that target fixation was greater for CP ($M = 0.59$, $SD = 0.09$) than for MP-vowel trials ($M = 0.50$, $SD = 0.15$; $t(27) = 3.10$, $p = 0.004$, Cohen's d = 0.42). However, target fixation did not differ for CP and MP-tone trials ($M = 0.57$, $SD = 0.14$; $p = 0.55$). Thus, word recognition accuracy was hindered by vowel mispronunciations but not by tone mispronunciations. Separate one-sample t tests revealed that target fixation was greater than chance on both CP ($t(27) = 5.27$, $p = 0.0001$, Cohen's d = 1.98) and MP-tone trials ($t(27) = 2.80$, $p = 0.009$, Cohen's d = 1.06), but did not differ from chance on MP-vowel trials ($p = 0.92$). Thus, while the 3-year-olds recognized words when they were correctly pronounced, they rejected vowel mispronunciations as correct labels for visual targets. In contrast to 2-year-olds, 3-year-olds associated tone mispronunciations with visual targets demonstrating a relatively reduced sensitivity to tone in comparison to younger children.

Experiment 1a demonstrated that tone and vowel variation hindered the efficiency of word recognition at both 2 and 3 years of age. In terms of the probability of fixating the target in response to mispronunciations, the two age groups appeared different from one another with older children appearing relatively insensitive to tone variation. However, children at both age groups were sensitive to vowel variation. Thus, despite the importance of vowels and tones in Mandarin, Mandarin learners demonstrated a robust sensitivity to vowel variation at 2 and 3 years of age, but not to tone variation at 3 years of age.

One might argue that vowel variation in the present study might be more salient than tone variation. Additionally the vowel shift (/e/-/u/) may have been more complex given that it is also associated with a frequency change. However, differences due to salience are unlikely as the MP-vowel token did not have a different tone contour to the corresponding CP token (Fig. 1). Furthermore, adults rated vowel and tone variation as similarly discriminable relative to correct pronunciations. They also correctly identified the tone of the MP-vowel token. Finally, if vowel variation was more salient than tone variation, perceptual salience should have influenced the performance of both age groups similarly, which was not observed. Alternatively, it is possible that word recognition was influenced by the onset consonant preceding tone or vowel variation. However, infants did not reliably fixate the target when vowel substitutions were made, suggesting that tones and vowels differ the extent to which they each constrain word recognition at 3 years of age.

Experiment 1a suggests that 3-year-olds did not use tones to distinguish word identity in a word learning task. This finding is unexpected given the phonological status of tones in Mandarin Chinese and given that 3-year-olds can use tones to recognize familiar words (Wong et al., 2005). Thus, it is important to examine factors that facilitate the integration of tones in word representation. First, presumably due to the diversity of the functions fulfilled by tones (e.g., emotional expression, stress, question/statement distinction), learners may need additional support to guide their interpretation of tones in word learning. Such support could include exposure to a minimal pair of words, contrasted by tone, that highlights the lexical function of tones. Research has shown that 3-year-olds' tone perception in familiar word recognition can reach ceiling when minimal pairs contrasting in tones are used (Wong et al., 2005). Second, mispronunciation paradigms – which tests tone knowledge on the MP trial where neither of the two images provide an exact match for the label – may encourage children to look at the image that is more likely to be the target. This may drive an insensitivity to tone variation in older children for similar reasons that adult rated tone variation as less likely to result in a new label in our stimulus validation procedure. Facilitative effects of tone variation on target fixation may be more evident in older children as they may have acquired some of same biases that drive adults' interpretations of tone changes. Thus, it is possible that 3-year-olds could use tones to learn new words when the function of tones is highlighted within the task. Furthermore, since Experiment 1a used only T2 and T4, it is unclear whether other tone pairs (in particular, the tone pair containing T3) would elicit different results. These questions are examined in Experiment 1b.

5. Experiment 1b

In Experiment 1b, we examined whether Mandarin-learning 3-year-olds can use tones to learn new words when additional support is provided. Specifically, their ability to recognize words was tested with correct pronunciations matching the target word. In addition, children were tested on minimal pairs differing only in tones contrasted during an initial training phase. In addition, this experiment examines whether T3 words are more difficult to learn than other tone words.

5.1. Participants

Participants were 26 3-year-old children ($M = 37.89$ months, range = 35.47–41.77; 13 boys). Three additional children were excluded from the final sample because of inattentiveness.

5.2. Stimuli

Images of four novel objects were used (Horst, 2009). The novel labels were /ʃɔ/ and /tɪu/, each produced in three versions with T2, T3, and T4 respectively (Fig. 4). These labels are non-words in Mandarin. To validate the stimuli, 15 adult native Mandarin speakers completed an auditory perception task. On each trial, participants

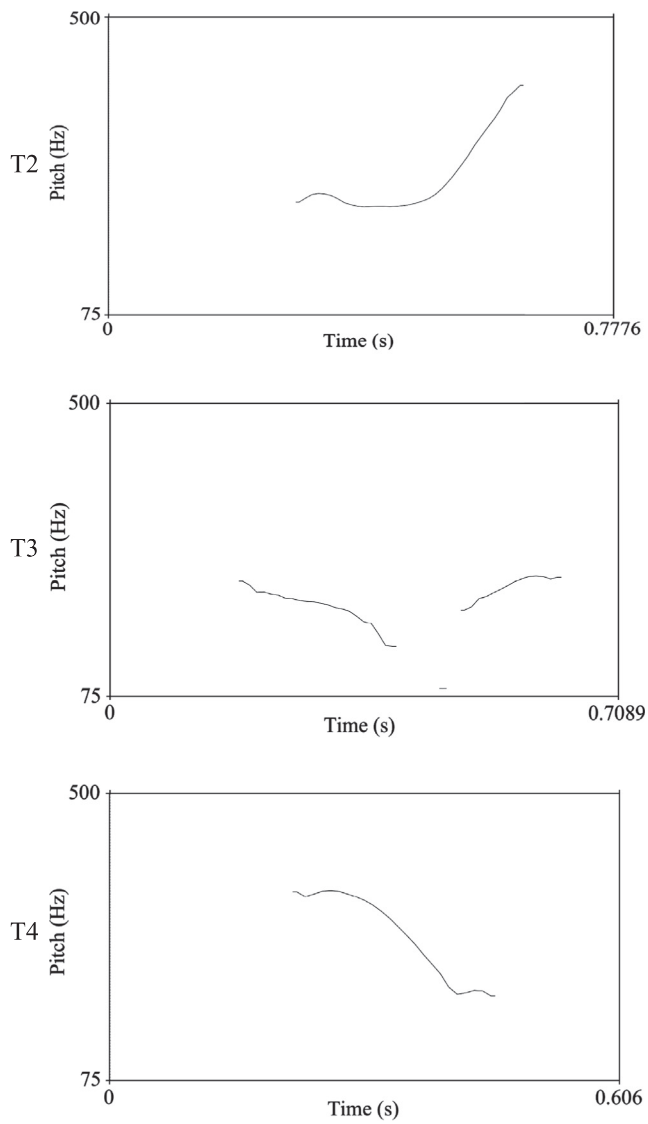


Fig. 4. Sample tone contours for the stimuli used in Experiment 1b (/ʃɔ/).

heard a pair of sounds differing only in tone (e.g., /ʃɔ2/-/ʃɔ4/). Participants rated their perceptual similarity on a 1–7 scale (1 = not similar, 3 = somewhat similar, 7 = highly similar) and then were asked to identify the tone of each sound by selecting T1, T2, T3, or T4. Since adults' ratings did not differ between /ʃɔ/ and /tɪu/, ratings were collapsed across labels. A paired-samples *t* test revealed that T2/3 pairs ($M = 5.89$, $SD = 1.12$) were rated as perceptually more similar than T2/4 pairs ($M = 3.54$, $SD = 1.98$; $t(14) = 2.65$, $p = 0.005$, Cohen's d = 0.68). This is consistent with the finding that T3 is less distinctive than other tones (Blicher, Diehl, & Cohen, 1990; Gottfried & Suiter, 1997; Hao, 2012; Liu & Samuel, 2004; Shen & Lin, 1991; So & Best, 2010; Whalen & Xu, 1992). Furthermore, all participants correctly identified the tones of the word tokens.

The experimental procedure was similar to that of Experiment 1a except that the training included *tone* minimal pairs (e.g., /ʃɔ2/, /ʃɔ4/). Furthermore, the test phase only contained CP trials (Table 3). The experiment consisted of two blocks: a T2/4 block and a T2/3 block. In the T2/4 block, children learned a T2 word and a T4 word (e.g., /ʃɔ2/, /ʃɔ4/). In the T2/3 block, they learned a T2 word and a T3 word (e.g., /tɪu2/, /tɪu3/). The block assignment of /ʃɔ/ and /tɪu/ and the presentation order of blocks were counterbalanced across participants.

Based on the relatively late acquisition of T3 (Wong, 2012a, 2012b; Wong et al., 2005), we predicted difficulty in learning T3 words. Furthermore, testing T2 words in both the T2/4 and T2/3 blocks enables us to evaluate whether the ease of mapping tones onto words is influenced by perceptual similarity across tones. If perceptual similarity plays a role, the T2/3 block should be associated with reduced learning of *both* T2 and T3 words. However, if the difficulty of T3 acquisition is specific to T3 regardless of contextual cues provided by training, children's performance on T2 words should not differ between blocks.

5.3. Coding and data analysis

Coding was identical to the Experiment 1a. Recording of 20% of the subjects by another coder yielded an inter-coder agreement of 99%. Trials with attention less than 20% were excluded from data analysis, which resulted in a total of 6 trials being excluded across all participants. The final data sample contained 202 test trials.

6. Results and discussion

6.1. Latency

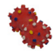
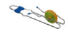
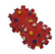
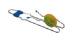
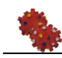
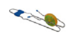
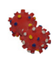
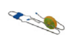
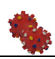
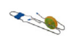
Each block had four test trials, two for each word. We collapsed latencies across trials to generate an average latency for each word within a block. Separate paired-samples *t* tests demonstrated that in the T2/4 block, latencies did not differ between T2 ($M = 0.21$ s, $SD = 0.21$) and T4 words ($M = 0.37$ s, $SD = 0.38$; $p = 0.12$) (Fig. 5). However, in the T2/3 block, latencies were shorter for T2 words ($M = 0.24$ s, $SD = 0.41$) than for T3 words ($M = 1.18$ s, $SD = 0.82$; $t(15) = 4.15$, $p = 0.001$, Cohen's d = 1.04). Furthermore, both T2 ($t(17) = 3.94$, $p = 0.0001$, Cohen's d = 0.93) and T4 words ($t(17) = 2.51$, $p = 0.02$, Cohen's d = 0.59) in the T2/4 block had a shorter latency than T3 words in the T2/3 block. Thus, only T3 words were associated with a processing cost with respect to latency. Furthermore, the latency of T2 words did not differ between blocks ($p = 0.48$), suggesting that training that highlighted tone contrasts did not affect the efficiency of word recognition of T2 words.

6.2. Proportion of total look to the target (target fixation)

For each child, we calculated an average target fixation for each word within a block. Target fixation was compared to chance (0.50) in order to evaluate children's abilities to learn new words differentiated by tones. In the T2/4 block, planned one-sample *t* tests revealed that target fixation was above chance on both T2 ($M = 0.64$, $SD = 0.14$; $t(25) = 5.07$, $p = 0.0001$, Cohen's d = 1.99) and T4 trials ($M = 0.61$, $SD = 0.20$; $t(25) = 2.97$, $p = 0.006$, Cohen's d = 1.16) (Fig. 6). In the T2/3 block, target fixation was above chance on T2 trials ($M = 0.65$, $SD = 0.15$; $t(25) = 5.35$, $p = 0.0001$, Cohen's d = 2.01) but not on T3 trials ($M = 0.48$, $SD = 0.12$; $p = 0.36$). Thus, when presented with T2 and T3 words, children learned T2 words but not T3 words. Three-year-olds therefore used T2 and T4 but not T3 to learn new words. This is consistent with the commonly reported difficulty of T3 acquisition (Wong et al., 2005). Was the ease of tone word learning influenced by the perceptual discriminability within a particular word pair? A paired-samples *t* test argued against this possibility by demonstrating that target fixation to T2 words did not differ between blocks ($p = 0.65$).

Experiment 1b demonstrated that the 3-year-olds integrated tones in word learning when supportive cues were provided that highlighted the lexical function of tones. Combined with Experiment 1a, the findings suggest that Mandarin-learning 3-year-olds do *not* automatically use tones as cues to word identity and that vowels constrain word learning to a greater degree than tones at

Table 3
An example (the T2/4 block) of trial sequences for the visual and speech stimuli in Experiment 1b.

	Left side	Center	Right side	Speech stimuli
Salience				No audio
Training: Animations of objects (4 trials, 24 s each). The 2 trials repeat				看! 这是一个ʃɔ2. 这里有个ʃɔ2! 看! 这是个ʃɔ2. 看ʃɔ2在动. 这个ʃɔ2从这边过来了. ʃɔ2, 这个东西叫ʃɔ2. 这是一个ʃɔ2. (Look! This is a ʃɔ2. There is a ʃɔ2. Look! This is a ʃɔ2. Look at ʃɔ2 moving. This ʃɔ2 is moving over here. ʃɔ2, it is called ʃɔ2. This is a ʃɔ2.)
				看! 这是一个ʃɔ4. 这里有个ʃɔ4! 看! 这是个ʃɔ4. 看ʃɔ4在动. 这个ʃɔ4从这边过来了. ʃɔ4, 这个东西叫ʃɔ4. 这是一个ʃɔ4. (Look! This is a ʃɔ4. There is a ʃɔ4. Look! This is a ʃɔ4. Look at ʃɔ4 moving. This ʃɔ4 is moving over here. ʃɔ4, it is called ʃɔ4. This is a ʃɔ4.)
Test: (4 trials: 7 s each)				**Trial 1**: ʃɔ2! 快看, ʃɔ2. 看! 哪一个是ʃɔ2? 快看ʃɔ2! (ʃɔ2! Look quickly at ʃɔ2. Look, which one is ʃɔ2? Look quickly at ʃɔ2.)
				Trial 2: ʃɔ4! 快看, ʃɔ4. 看! 哪一个是ʃɔ4? 快看ʃɔ4! (ʃɔ4! Look quickly at ʃɔ4. Look, which one is ʃɔ4? Look quickly at ʃɔ4.)
Reminder: (2 trials, 7 s each)				ʃɔ2, 这个东西叫ʃɔ2. 这是一个ʃɔ2. (ʃɔ2, it is called ʃɔ2. This is a ʃɔ2.)
				ʃɔ4, 这个东西叫ʃɔ4. 这是一个ʃɔ4. (ʃɔ4, it is called ʃɔ4. This is a ʃɔ4.)
Test: (the above 2 test trials repeated)				**Trial 3**: ʃɔ2! 快看, ʃɔ2. 看! 哪一个是ʃɔ2? 快看ʃɔ2! (ʃɔ2! Look quickly at ʃɔ2. Look, which one is ʃɔ2? Look quickly at ʃɔ2.)
				Trial 4: ʃɔ4! 快看, ʃɔ4. 看! 哪一个是ʃɔ4? 快看ʃɔ4! (ʃɔ4! Look quickly at ʃɔ4. Look, which one is ʃɔ4? Look quickly at ʃɔ4.)

Note: The T2/3 block followed the same procedure except that another pair of novel objects and target words (/tiu2/, / tiu3/) were used. The word assignment into the T2/4 or T2/3 blocks was counterbalanced across participants.

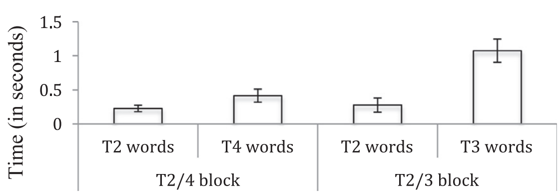


Fig. 5. Latency in Experiment 1b. Only T3 words were associated with a low level of word recognition efficiency. Furthermore, T2 words did not differ in latency between blocks. Error bars reflect SEM.

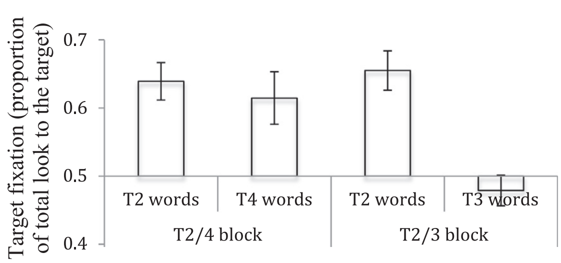


Fig. 6. Target fixation in Experiment 2 (this experiment has only CP test trials). Target fixation was greater than chance on both words in the T2/4 block, but only on T2 words in the T2/3 block. Furthermore, T2 words did not differ in target fixation between blocks. Error bars reflect SEM.

3 years of age. This raises the question as to why tone sensitivity in word learning declines with age. It is possible tones bear a weaker relationship to word identity than vowels in tone languages. This phonological knowledge may have already been acquired by 3-year-olds but not by 2-year-olds. Alternatively, given the various functions of pitch in tone languages, it is possible that accurate tone interpretation requires additional cues, such as within-task training that highlights the lexical relevance of tone. Novel words do not have established phonological representations, potentially rendering tone interpretation ambiguous given that tone changes serve non-lexical functions as well. In addition, younger children tend to store highly specific acoustic features (Hollich, 2006; Newman, 2006). The goal of Experiment 2 was to examine whether 3-year-old children demonstrate a greater sensitivity to tones when they are familiar with the meanings of target words.

7. Experiment 2

Experiment 2 examined 3-year-olds' recognition of familiar words, which were either correctly produced or mispronounced with tone or vowel variation. This experiment evaluated whether the primacy of vowels over tones observed in the 3-year-olds' novel word learning still holds when familiar words are used.

7.1. Participants

The participants were 28 3-year-old children (M = 38.01 months, range = 32.50–40.00; 14 boys). One additional child was excluded from the final sample because of overall inattentiveness.

7.2. Stimuli

Four criteria were used in selecting tested word pairs: (1) the 3-year-olds were highly familiar with the words and the relevant images, (2) words paired together had the same onset consonants ensuring that participants could not use the onset to identify the target image, (3) the tone and vowel changes were also possible Mandarin words, and (4) the word pairs contained all 6 possible tone pairs for 4 tones (regardless of the order). As a result of these constraints, only six pairs of words were used (*bao1* [bag] – *bi3* [pen], *dao1* [knife] – *dan4* [egg], *chuan2* [boat] – *che1* [car], *shan1* [mountain] – *she2* [snake], *shou3* [hand] – *shu1* [book], *tao2* [peach] – *tu3* [soil]). Eleven of these words (except *tao2*) have age of acquisition (AoA) estimates (M = 18.55 months, SD = 2.16) based on the Mandarin MacArthur Communicative Developmental Inventory (Tardif et al., 2008). The AoA of a word is defined as the age at which at least 50% of children have a word in their productive vocabulary, normed across a large sample of children. Thus,

Table 4
Carrier phrases used in Experiment 2.

	Carrier phrases	Character by character English translation	Duration (s)
1	Kan4! Na3 yi 2 ge4 shi4 看！哪一个是	Look! Which one *classifier*[a] is	2.31
2	Ni2 qiao2! Na3 er shi4 你瞧!哪儿是	You look! Which *particle* is	2.37
3	Kuai4 qiao2! Na3 er shi4 快瞧!哪儿是	Quickly look! Which *particle* is	2.35
4	Kan4 kan4! Na3 ge4 shi4 看看!哪个是	Look Look! Which *classifier* is	2.17
5	Kuai4 kan4! Na3 er shi4 快看!哪儿是	Quickly look! Which *particle* is	2.25
6	Qiao2! Na3 yi 2 ge4 shi4 瞧!哪一个是	Look! Which one *classifier* is	2.34
7	Ni2 kan4! Na3 ge4 shi4 你看!哪个是	You look! Which *classifier* is	2.14

[a] Functional words are italicized.

3-year-olds were highly familiar with these words. Visual stimuli consisted of 12 images for each of the 12 words. To maintain children's attention in the task, seven different carrier phrases were used in which the target words occurred in utterance final position (Table 4).

For tone mispronunciation, *bao1*, *chuan2*, *dao1*, *tao2*, *shou3*, and *shu1* were pronounced as *bao2*, *chuan3*, *dao3*, *tao4*, *shou4*, and *shu4*, containing all 6 possible pairs of tone variation for 4 tones regardless of the order (Fig. 7). The vowel mispronunciations were realized by nasalizing the vowels. Thus *bi3*, *che1*, *dan4*, *shan1*, *she2*, and *tu3* were pronounced as *bĩ3*, *chẽ1*, *dãn4*, *shãn1*, *shẽ2*, and *tũ3*.[2] Vowel mispronunciations were therefore created in the same way for all words. Prior research suggests that Mandarin-learning 3-year-olds are highly familiar with nasalization as a means of varying word identity (Hua & Dodd, 2000; Tardif et al., 2008). Separate one-way ANOVAs revealed no significant differences in duration, intensity, or average pitch level among CP, MP-tone, and MP-vowel trials (Table 5). Target words were randomly paired with carrier phrases and then synthesized into a phrasal unit using Audacity 2.0.3.

To validate the stimuli, 15 adult native Mandarin speakers completed an auditory perception task. On each trial, they heard a pair of word tokens (CP and MP-tone *or* CP and MP-vowel) presented sequentially and rated their perceptual discriminability on a 1–7 scale (1 = not discriminable, 4 = discriminable, 7 = easily discriminable). They were asked to identify the tone of each token. For each participant, we calculated average discriminability ratings for CP–MP-vowel pairs and CP–MP-tone pairs respectively. A paired-samples t test showed that CP-MP-tone pairs ($M = 5.06$. $SD = 0.78$) were as discriminable as CP-MP-vowel pairs (M = 5.18, SD = 0.87), ($p = 0.56$), a pattern of results similar to that of Experiment 1a. Furthermore, all participants correctly identified the tones of these word tokens.

7.3. Procedure

The experimental procedure was almost identical to that of Singh et al. (2015) and used a standard procedure to investigate children's sensitivity to familiar words (e.g., Mani & Plunkett, 2007). In this paradigm, children saw two images side-by-side for 6 s on each trial. The onset of the target word began 2633 ms into a trial (e.g., Look! Where is the [target word]?). As children's responses to the target word were calculated from 367 ms after the onset of the target word, each trial was segmented into two 3-s phases: a pre-target phase and post-target phase. A significant increase in looking time to the target image across phases indicates that the participant has mapped the verbal label onto the visual target.

The experiment contained 24 trials: 12 CP trials, testing recognition of the 12 target words, as well as 6 MP-tone trials and 6 MP-vowel trials. No pictures or trials of the same pronunciation type (CP, MP-tone, MP-vowel) were presented consecutively. Four stimulus orders were created. The left/right position of target images was counterbalanced across subjects.

7.4. Coding and data analysis

Coding was identical to previous experiments. Intercoder agreement on 20% of the data set was 98%. Trials with attention coded as less than 20% were excluded from data analysis, which resulted in a total of 24 trials being excluded. Based on the analysis standards set by prior research (e.g. Mani & Plunkett, 2007), only trials where children fixated both the target and the distractor during the pre-target phase were included, which resulted in another 20 trials being excluded. A total of 628 trials were analyzed. As before, we examined both latency and target fixations.

8. Results and discussion

8.1. Latency

For each child, we calculated average latencies for CP, MP-tone, and MP-vowel trials. A repeated measures ANOVA revealed a main effect of pronunciation type ($F(2,54) = 29.60$, $p = 0.0001$, $\eta_p^2 = 0.52$), suggesting that latency differed by pronunciation type (Fig. 8). Planned paired-samples t tests revealed that latency for CP trials ($M = 0.47$ s, $SD = 0.26$) was shorter than that for MP-tones trials ($M = 0.70$ s, $SD = 0.34$; $t(27) = 4.96$, $p = 0.0001$, Cohen' d = 0.94), which in turn was shorter than that for MP-vowel trials ($M = 1.00$ s, $SD = 0.48$; $t(27) = 4.34$, $p = 0.0001$, Cohen' d = 0.82). Therefore, both tone and vowel variation hindered the efficiency of spoken word recognition, but vowel variation hindered the efficiency more than tone variation.

8.2. Target fixation

We then calculated average target fixation for pre- and post-target phases by pronunciation type. A 2 (phase: pre-/post-target) × 3 (pronunciation type: CP/MP-tone/MP-vowel) ANOVA revealed a main effect of pronunciation type ($F(2,54) = 6.03$, $p = 0.004$, $\eta_p^2 = 0.18$) and phase ($F(1,27) = 57.03$, $p = 0.0001$,

[2] Due to the issue of practicality, one of the vowel mispronunciations (*bĩ3*) used in Experiment 2 is not a legal word in standard Mandarin. However, *bĩ* is a legal word with T1 or T4 in Mandarin. Furthermore, children's performance (either latency or target fixation) on the *bĩ3* trial was similar to other MP-vowel trials.

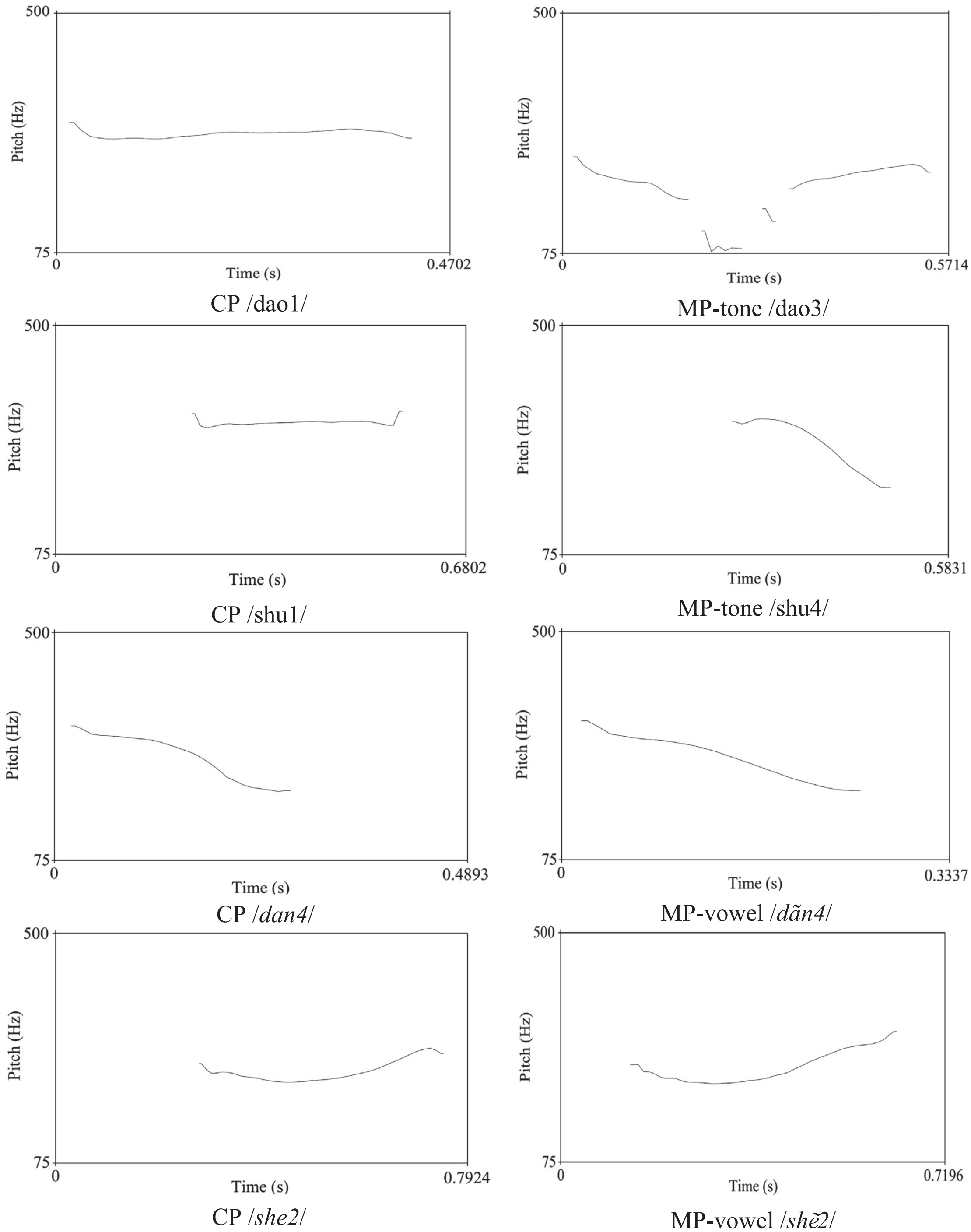


Fig. 7. Sample tone contours of the CP, MP-tone, and MP-vowel versions of the words used in Experiment 2.

Table 5
The ANOVA analyses comparing the average duration, intensity, and pitch of words across the three pronunciation types in Experiment 2.

	CP	MP-tone	MP-vowel	
Duration	0.60 s (±0.11)[a]	0.59 s (±0.14)	0.60 s (±0.13)	$F = 0.21$, $p = 0.76$
Intensity	66.55 dB (±2.80)	67.88 dB (±2.26)	65.03 dB (±2.06)	$F = 1.03$, $p = 0.38$
Pitch	265.24 Hz (±45.67)	259.77 Hz (±38.77)	260.47 Hz (±45.11)	$F = 0.29$, $p = 0.70$

[a] Mean and standard deviation.

$\eta_p^2 = 0.68$), and a pronunciation type × phase interaction ($F(2,54) = 7.68$, $p = 0.001$, $\eta_p^2 = 0.22$). To disaggregate this interaction, planned paired-samples *t* tests compared target fixation between phases for each pronunciation type (Fig. 9). On both CP and MP-tone trials, target fixation was greater in the post-target phase (CP: $M = 0.69$, $SD = 0.09$; MP-tone: $M = 0.67$, $SD = 0.22$) than the pre-target phase (CP: $M = 0.49$, $SD = 0.05$; MP-tone: $M = 0.46$, $SD = 0.14$), (CP: $t(27) = 10.13$, $p = 0.0001$, Cohen's d = 1.91; MP-tone: $t(27) = 4.97$, $p = 0.0001$, Cohen's d = 0.94). However, on MP-vowel trials, target fixation did not differ between post- ($M = 0.52$, $SD = 0.16$) and pre-target phases ($M = 0.47$, $SD = 0.11$; $p = 0.19$). Thus, children fixated the visual targets preferentially

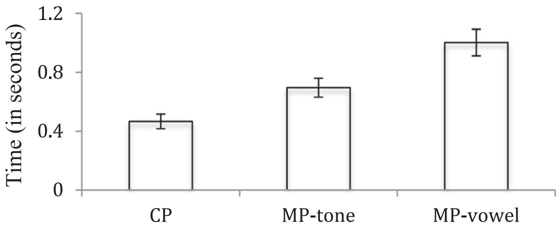


Fig. 8. Latency in Experiment 2. Both tone and vowel variation hindered word recognition efficiency, while vowel variation hindered this efficiency more than tone variation. Error bars reflect SEM.

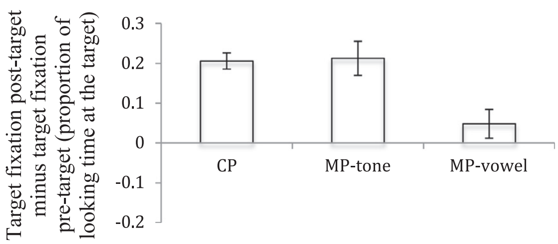


Fig. 9. Target fixation in the pre- and post-target phases in Experiment 2. Target fixation significantly increased across phases on CP and MP-tone trials but not on MP-vowel trials. Error bars reflect SEM.

when presented with correct pronunciations or tone mispronunciations, but not when presented with vowel mispronunciations.

Proportion of fixation to target was then analyzed for each of the six tone pairs (MP-tone trials) via a series of paired-samples *t* tests. Results revealed significant or marginally significant increases in target fixation across phases in five tone pairs (Fig. 10). Children were highly insensitive to tone variation in the T2-3 and T3-4 pairs, where T2 and T3 were mispronounced as T3 and T4 respectively. However, target fixation did not differ between phases for the T1-3 pair, where T1 was mispronounced as T3, suggesting that children were sensitive to the distinction between T1 and T3. Thus, the difficulty in acquiring T3 seems be related to its perceptual similarity to T2 and T4 in particular.

In Experiment 2, 3-year-olds fixated the target preferentially in the context of tone variation even when words were familiar to them. Experiments 1a and 2 revealed that the 3-year-olds were sensitive to vowel variation but not to tone variation when recognizing both newly learned and familiar words. This suggests that vowels maintain primacy over tones in constraining word learning and word recognition. In the absence of supportive cues highlighting the lexical contrastiveness of tones (Experiment 1b), native Mandarin learners were relatively insensitive to tone variation in tone at 3 years of age.

9. General discussion

This study examined monolingual Mandarin-learning 2- and 3-year-olds' sensitivity to vowels and tones in word recognition. Three major findings emerged. First, vowel and tone variation hindered the efficiency of word recognition, suggesting that both age groups were highly sensitive to vowel and tone variation. Second, vowel variation hindered the accuracy of word recognition in both age groups, but tone variation hindered the accuracy of word recognition only at 2 years of age. This suggests that vowels constrained word recognition more significantly than tones at 3 years of age. Finally, 3-year-olds used tones to learn new words only when additional support was provided. Even under these conditions, T3 was still difficult to acquire. Thus, despite the functional importance of tones in tone languages, vowels appear to have primacy over tones in the 3-year-old tone learners' word recognition.

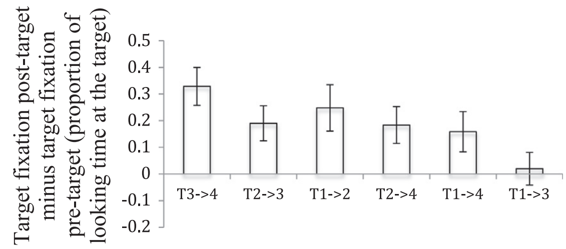


Fig. 10. The target fixations in the pre- and post-target phases for each tone pair. T*x*->*y*: when Tone *x* was mispronounced as Tone *y*. Target fixation significantly or marginally significantly increased across phases on T3->4 ($p < 0.0001$), T2->3 ($p = 0.008$), T1->2 ($p = 0.010$), T2->4 ($p = 0.014$), and T1->4 ($p = 0.047$) trials. The adjusted *p* value was used for the multiple comparison analyses. Error bars reflect SEM.

Our findings revealed that tone variation hindered the efficiency of word recognition in the 3-year-olds, suggesting that their acceptance of tone variation was not due to a failure to discriminate tones. Furthermore, tone sensitivity was reduced in recognizing both newly learned and familiar words, suggesting that tones were subject to similar constraints whether words were known or not. It appears then that tone learners demonstrate a 're-weighting' of tone sensitivity between 2 and 3 years of age. Combined with previous findings (Burnham et al., 2011; Quam & Swingley, 2010; Singh et al., 2015), this trend has been observed (1) across multiple age ranges from toddlerhood to early adolescence, (2) across different tone inventories, (3) in both monolingual and bilingual tone language learners, and (4) across different tone contrasts within a tone language. These findings support a generalized account where tone sensitivity declines with age – a reversal of what one might expect to happen with development – and suggest that tones bear a weaker relationship to word identity than segments across languages in child language processing. This is consistent with tone/segment processing asymmetries previously attested in adults (e.g., Wiener & Turnbull, 2016). The high priority assigned to segments in word recognition may be a universal feature of language in spite of the functional importance of tones in defining word identity in tone languages.

9.1. Why might tones have a weaker relationship to word identity than segments?

We propose three explanations for this: the origin of tones in relation to segments, the variety of communicative functions shared by pitch, and the lexical load carried by tones versus segments. First, lexical links between tones and meaning may be related to the origin and evolution of speech and music. Charles Darwin hypothesized that speech and music originated from a common emotional signalling system based on the imitation and modification of sounds in nature (Darwin, 1871, also see Brown, 2000; Fitch, 2010; Mithen, 2006). This hypothesis is supported by recent evidence that speech and music share underlying cognitive and neural resources (Levitin & Menon, 2003; Maess, Koelsch, Gunter, & Friederici, 2001; Musacchia, Sams, Skoe, & Kraus, 2007; Musso et al., 2015) and draw on a common code of prosodic attributes when used to communicate emotional states (Bowling, Gill, Choi, Prinz, & Purves, 2010; Ilie & Thompson, 2006; Juslin & Laukka, 2003; Thompson, Marin, & Stewart, 2012). Changes in frequency spectrum, intensity, and tempo that evoke emotional responses in music and speech also trigger emotions when perceived in nonlinguistic, non-musical environmental sounds (Ma & Thompson, 2015). Perhaps the vast number of potential meanings used in linguistic communication may have necessitated a high degree of differentiation in language relative to music. Indeed, tones are thought to have originated in part out of a consonant system that was 'bursting out of its seams' such that tone contrasts evolved as a way to maintain lexical distinctiveness (Matisoff, 1973). It is possible that segments permit greater lexical differentiation than tones due to the multiple concurrent ways in which they can be contrasted (e.g., consonants: aspiration, place, manner of articulation; vowels: backness, roundedness, height, etc.). Thus, segments may be better optimized for lexical differentiation and may therefore carry more lexical weight as suggested by some studies on adult language processing (Tong et al., 2008; Wiener & Turnbull, 2016).

A question of great theoretical importance then arises: If a language component shares certain properties with another domain, will that component have a weaker function in language? This possibility is supported by the current findings and by a recent ERP study revealing earlier neural responses to vowel rather than to tone variation in auditory word recognition in Mandarin (Hu et al., 2012). However, it is unclear whether the weak responses to pitch variation (i.e. tones) found in this study was driven by participants placing a greater premium on segments as an influential source of contrast or whether this was a by-product of the cross-domain functions served by pitch. It should be noted that there are factors determining vowel quality that are also shared with musical systems. For example, timbre is a key property differentiating vowels in speech as well as instruments in music (Lotto & Holt, 2011). Therefore, cross-domain sharing of resources is not unique to tones and is evident in some vowel contrasts as well, yet participants appeared highly sensitive to vowel variation in the present study. Finally, cross-domain sharing may even enhance cognitive processing as indicated by a recent ERP study showing that emotional prosody evoked faster neural responses in non-linguistic vocalizations (e.g., laughter) than in speech (Pell et al., 2015). However, there has been little comparison of the processing of different acoustic attributes with the speech code based on whether they share properties with non-linguistic auditory stimuli, such as music. Future research could examine whether the degree of vowel qualities (frequency spectrum, tempo, duration, timbre, etc.) that are shared with other domains (e.g., music, environment sounds) influences phonological acquisition. This will help us understand how human auditory cognitive system receives and registers incoming acoustic detail and processes this information accordingly to fulfill distinct, specialized cognitive functions. This will also inform broader conclusions on how auditory cognition is affected by attention, context, prior knowledge, and acoustic properties shared across domains, a topic that remains somewhat unresolved (Grosjean, 2001; Peretz & Coltheart, 2003).

Second, tones may have a weaker relationship to word identity than segments because of the functional diversity of tones in speech. In contrast to segments, pitch variation serves other communicative functions beyond defining word identity. For example, in non-tone languages, 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 tones) (Quam & Swingley, 2010), tones may have a weaker relationship to word identity than segments even in tone languages on account of the functional complexity of pitch cues in tone languages.

Based on this explanation, the relative insensitivity to tone variation in Mandarin word recognition should be accompanied by an enhanced sensitivity to intonation. Do children demonstrate an enhanced sensitivity to intonation between 2 and 3 years of age? Although little research has examined this question in Mandarin, it has been investigated in studies on non-tone-learning children. Research with German-learning 6-, 12-, 24-, and 36-month-olds has demonstrated that only 36-month-olds reliably used intonation to detect and anticipate the end of a speaker's turn (Keitel & Daum, 2015; Keitel, Prinz, Friederici, Hofsten, & Daum, 2013). Furthermore, electrophysiological correlates of the processing of intonational phrase boundaries (e.g., syntactic boundaries) were observed in German-learning 3-year-olds but not in 21-month-olds (Männel & Friederici, 2011). These findings provide preliminary support for a developmental advance in intonation knowledge between 2 and 3 years of age in non-tone-learning children and provides suggestive evidence for a growing awareness of pitch as a source of intonational contrast as children mature.

Intonation serves communicative functions in tone languages as well (e.g., Liu & Pell, 2012; Yuan, 2011), making attention to intonational variation just as important for tone learners. Thus, Mandarin-learning 3-year-olds may be developing a greater sensitivity to intonation, triggering a recalibration of tone categories in word processing. However, as the current study explicitly tested word processing, children's knowledge about intonation remains undetermined. Future research could examine whether Mandarin learners' intonation sensitivity is robust enough to influence their pitch perception in general at 2–3 years of age or whether the processing of intonation precedes that of tones. In a study pitting intonation contour against tone contour, Singh and Chee (2016) found that children only resolved intonation-tone relations as late as 4–5 years of age. Furthermore, a recent study reported a rebound in tone sensitivity in Dutch-learning 17- and 18-month-olds, presumably due to an emergent sensitivity to pitch as a source of intonation contrast (Liu & Kager, 2014). Given that pitch functions are even more complex in tone languages (i.e. distributed over lexical and non-lexical functions), it is possible that an increase in sensitivity to intonation is evident later in Mandarin speakers than Dutch speakers. In Mandarin, the acquisition of intonation contrast may lead children to hypothesize that pitch variation does not always define word identity, which may result in a decline in tone sensitivity as children develop. Furthermore, the importance of tones in Mandarin and the acoustic similarities between tones and intonation may complicate this process (Beckman & Venditti, 2010; Ladd, 1996) and pose a challenge for young tone learners. Intonation acquisition continues well beyond the age of 3, given that 3-year-old non-tone learners have difficulty in understanding sarcasm (Laval & Bert-Erboul, 2005) and do not show reliable electrophysiological evidence of processing pauses as prosodic boundary markers (Männel, Schipke, & Friederici, 2013). Furthermore, both tone and non-tone learners start to reliably attend to intonation cues in conjunction with lexical content rather late in development, approximately between 4 and 5 years of age (Friend, 2003; Morton & Trehub, 2001; Quam & Swingley, 2012; Singh & Chee, 2016).

Why does intonation have little influence on the 2-year-olds? This may be related to the fact that acquisition of intonation is a protracted and cumulative process that involves integration of knowledge about pitch contours, grammatical structures, socio-emotional and pragmatic meanings, which may not yet be acquired and fully appreciated by 2-year-olds. For example, before the ages 3–4, children have difficulty (1) matching facial expressions to emotional prosody in speech (Pons, Harris, & de Rosnay, 2004; Sauter, Panattoni, & Happé, 2013),[3] (2) processing intona-

[3] It should be noted that infants are *sensitive* to socio-emotional cues early in life as indicated by (1) their preference of infant-directed speech (Trainor, Austin, & Desjardins, 2000), (2) their ability to associate pitch with their mothers' happy and sad faces at 3.5 months of age (Grossmann, Striano, & Friederici, 2006), (3) their discrimination amongst happy, angry, and sad prosody at 5 months of age (Flom & Bahrick, 2007), and (4) the influence of the valence of vocal sounds on 12-month-olds' attention to objects and exploratory behavior (Vaish & Striano, 2004). However, this early sensitivity to socio-emotional prosody cannot be equated with a finely differentiated appreciation of the various communicative functions fulfilled by prosodic variation.

tional phrase structure (Männel & Friederici, 2011), and (3) reliably resolving tone and intonation relations (Singh & Chee, 2016). Thus, an emergent appreciation for the complex functions of pitch in human languages lead to a transient uncertainty in interpreting pitch movements, which may compromise tone sensitivity in lexical tasks. A longitudinal investigation of the co-evolution of tone and intonation sensitivity in early childhood would inform this possibility.

Third, tones may bear a weaker relationship to word identity than segments in Mandarin because there are more segments than tones in Mandarin. Mandarin has six vowels and 19 consonants but only four basic tones. Statistically, there are consequently more possible segment variations than tone variations. Thus, segment variation may be used more often to induce a change in word identity than tone variation. This may have a stronger influence on older children since they have more language exposure than younger children. Furthermore, tones may be acoustically more variable than segments given that they are also subject to intonation variation. Therefore, tones and segments may differ in terms of how much deviation leads to the perception of a mispronunciation, with tones perhaps associated with more lenient phonemic boundaries. In a corpus analysis of Mandarin Chinese, Tong et al. (2008) demonstrated that the information value, referring to the lexical load carried by phonological constituents, was greater for vowels than for tones. It is possible that the 'division of labor' carried by vowels versus tones in Mandarin renders tones more lexically dispensable than vowels.

9.2. Facilitating tone sensitivity in young children

The present study showed that 3-year-old Mandarin learners used tones to learn new words only when tone minimal pairs were provided, suggesting that highlighting the function of tones can facilitate tone word learning. However, it only facilitated sensitivity to T2 and T4 but not T3. Furthermore, the efficiency and accuracy of word recognition for T2 words did not differ between T2/4 and T2/3 blocks. Thus, the provision of minimal pairs could enhance but not ensure accurate interpretations of tones in word learning. It should be noted that in recognizing familiar words (Experiment 2), 3-year-olds were highly sensitive to tone variation when T1 was mispronounced as T3 but not when T3 and T2 were mispronounced as T4 and T3 respectively. Future research could examine whether children can learn new T3 words in a T1/3 pair. Additionally, it would be interesting to investigate whether the use of tone minimal pairs leads to the lexical integration of tones even in young non-tone learners to determine whether this reflects a language-general tendency to extract determinants of meaning from training provided within a task versus true mastery of a native phonology.

In Mandarin child-directed speech, parents do not often offer tone minimal pairs when naming words (Tardif, 1996). How do children then learn tone words without minimal pairs in the real world? Perhaps children's growing vocabulary and language exposure enables them to retrieve familiar words from their vocabularies and represent words in relation to one another, such as minimal pairs. For example, children's knowledge that *ma1* refers to "mother" may encourage their integration of tones when hearing *ma3* being used to refer to "horse", even when the minimal pair is not provided by the caregiver. Thus, it appears to be helpful to provide minimal pairs when speaking in a tone language to young learners. Furthermore, children may more readily use tones in word learning when the new word forms a tone minimal pair with a familiar word. This prediction could be examined in future research.

When considering that children at the age of 3 do not reliably integrate tones in word learning or word recognition, it should be noted that even 5-year-old Mandarin speakers do not produce adult-like tones (Wong, 2013). Does tone mispronunciation necessarily result in communication failure in tone languages? The current study argues against this possibility. In natural discourse, there are multiple mechanisms available disambiguating referents when tone mispronunciation occurs. For example, the naming environment and the use of nonverbal communicative cues (e.g., pointing gestures, eye-gaze) can eliminate many possible referents of a label (Gliga & Csibra, 2009; Morales, Mundy, & Rojas, 1998). Furthermore, the syntactic structure in which a word appears may facilitate tone word recognition. For example, *bei* means "to carry on the back" with T1 and "the back" with T4 and its meaning can therefore be determined via syntactic bootstrapping or similar mechanisms. In line with this possibility, Mandarin child-directed speech typically provides reliable morpho-syntactic cues to form class distinctions (Ma & Golinkoff, 2016), suggesting that a decline in sensitivity to tones in isolated words may not have catastrophic effects for communication.

9.3. Implications for perceptual narrowing

Prior research has widely reported a feat of early phonological development – perceptual narrowing – where young infants initially discriminate many phonetic contrasts, both native and non-native, then narrow down this sensitivity to only native contrasts (e.g., Gervain & Mehler, 2010; Kuhl et al., 1992; Polka & Werker, 1994). Indeed, infants reared in vowel-consonant languages can discriminate tones at 6 months of age; some studies have revealed that non-tone learners lose this ability by 9 months (Mattock & Burnham, 2006; Mattock et al., 2008, but see Liu & Kager, 2014). The current study has revealed a later period of phonological reorganization within phonemic contrasts, which may be driven by their relative strength in constraining word processing. Even in a language where both vowels and tones define word identity, sensitivity to some sources of phonological variation, such as tone, declines as children mature whereas sensitivity to other sources, such as vowels, remains comparatively stable.

The reduced tone sensitivity applies to the accuracy rather than the efficiency of word recognition, suggesting that this phonological reorganization represents a more developed form of phonological knowledge. This reduction may reflect a growing understanding of the functional diversity of pitch in tone languages. This study invites a more comprehensive definition of the mastery of lexical tones, which is not simply to integrate lexical tones when accessing meaning, but also to differentiate lexical tones from similar pitch movements that serve other communicative functions, such as affect, intonation, and the communication of stress.

The present study raises the question of whether lexical tone sensitivity would rebound after 3 years of age. Arguing against this possibility is the findings that a reduction in sensitivity to tone variation in word processing was observed in 4.5-year-old English-Mandarin bilinguals and 5- to 11-year-old Cantonese speakers (Burnham et al., 2011; Singh et al., 2015), and even in adult Mandarin speakers based on the rating data obtained in Experiment 1a and past studies that showed a primacy of vowel over tone processing in Mandarin word recognition (e.g., Taft & Chen, 1992; Wiener & Turnbull, 2016; Ye & Connine, 1999). Therefore, we hypothesize that lexical tones become increasingly dispensable with age, although such an account merits further investigation.

9.4. A delayed emergence of the reduced tone sensitivity in bilingual children

An age-dependent effect of tone and segment variation emerged between 3 and 4.5 years of age in Singh et al.'s (2015) but between 2 and 3 years of age in this study. This difference may be due to children's language exposure. While Singh et al. tested Mandarin-English bilingual children who were fully competent in Mandarin, this study tested monolingual Mandarin learners. This delayed reorganization of tone sensitivity is consistent with the hypothesis (i.e., the delay hypothesis in Paradis & Genesee, 1996) that bilingual acquisition may lead to a slower rate of acquisition of a particular linguistic feature. Perhaps compared with monolingual Mandarin learners, English-Mandarin bilingual children are exposed to more tone variation (in the form of greater overall pitch variation), which may delay the establishment of Mandarin tone categories. The bilingual environment may therefore provide a more complex set of pitch movements across languages, contributing to a slower rate of tone acquisition (but see Singh, Poh, & Fu, 2016). Nevertheless, the trajectory of tone acquisition is similar across the two studies pointing to a reduction in tone sensitivity in early childhood. Future research should examine the relationship between the amount of exposure to non-tone languages and tone sensitivity in bilingual children. However, it should be noted that two recent studies suggest that tone sensitivity may be enhanced on account of bilingual experience. First, Liu and Kager (in press) showed that bilingual infants learning two non-tone languages showed greater perceptual sensitivity to non-native tones than their non-tone monolingual peers. Second, Singh et al. (2016) tested monolingual Mandarin and bilingual English-Mandarin infants on their sensitivity to tones in novel word learning at 13 months. They report evidence of a bilingual advantage in incorporating tones into words taught in Mandarin Chinese over Mandarin monolingual infants. Further research could investigate whether monolinguals and bilinguals differ in their tone sensitivity in toddlerhood and early childhood.

10. Conclusion

This study examined Mandarin-speaking 2- and 3-year-olds' sensitivity to vowels and tones in recognition of newly learned words and familiar words. Results showed an age-dependent sensitivity to vowels and tones between 2 and 3 years of age featuring a relative insensitivity to tone variation in 3-year-old children. Thus, although vowels and tones distinguish words in Mandarin Chinese, vowels may maintain primacy over tones in word processing, reflecting a vowel processing advantage which may emerge between 2 and 3 years of age.

Acknowledgments

W.M. is supported by Macquarie University ARC Centre of Excellence in Cognition and its Disorders (CCD) through a CCD Research Fellowship. P.Z. is supported by a National Social Science Foundation of China (16BYY076). L.S. is supported by a Singaporean Ministry of Education Tier 1 Academic Research Fund (FY2013-FRC2-009).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2016.11.011.

References

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17*, 1265–1282.
Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1; 2. *Journal of Child Language, 32*, 159–173.
Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*, 614–636.
Beckman, M. E., & Venditti, J. J. (2010). Tone and intonation. In *The handbook of phonetic sciences*, pp. 603–652). .
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.
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*, 37–49.
Bowling, D. L., Gill, K., Choi, J. D., Prinz, J., & Purves, D. (2010). Major and minor music compared to excited and subdued speech. *The Journal of the Acoustical Society of America, 127*, 491–503.
Brown, S. (2000). The "musilanguage" model of music evolution. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 271–300). Cambridge, MA: MIT Press.
Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknecht, C., & Kasisopa, B. (2011). Are tones phones? *Journal of Experimental Child Psychology, 108*, 693–712.
Cheng, Y. Y., Wu, H. C., Tzeng, Y. L., Yang, M. T., Zhao, L. L., & Lee, C. Y. (2013). The development of mismatch responses to Mandarin lexical tones in early infancy. *Developmental Neuropsychology, 38*, 281–300.
Ciocca, V., & Lui, J. Y.-K. (2003). The development of the perception of Cantonese lexical tones. *The Journal of Multilingual Communication Disorders, 1*, 141–147.
Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*, 1584–1595.
Cutler, A., & Chen, H. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics, 59*, 165–179.
Darwin, C. (1871). *The descent of man and selection in relation to sex* (1st ed.). London: John Murray.
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.
Demuth, K. (1993). Issues in the acquisition of the Sesotho tonal system. *Journal of Child Language, 20*, 275–301.
Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5) (Serial No. 242).
Fernald, A. (1985). Four-month-olds prefer to listen to motherese. *Infant Behavior and Development, 8*, 181–195.
Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*, 279–293.
Fitch, W. T. (2010). *The evolution of language*. Cambridge, UK: Cambridge Univ. Press.
Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology, 43*, 238–252.
Friend, M. (2003). What should I do? Behavior regulation by language and paralanguage in early childhood. *Journal of Cognition and Development, 4*, 161–183.
Fromkin, V. (1978). *Tone: A linguistic survey*. New York: Academic Press.
Gandour, J., & Harshman, R. A. (1978). Cross language differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1–33.
Garding, E., Kratochvil, P., & Svantesson, J. (1986). Tone 4 and tone 3 discrimination in modern standard Chinese. *Language and Speech, 29*, 281–293.
Gervain, J., & Mehler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Review of Psychology, 61*, 191–218.
Gliga, T., & Csibra, G. (2009). One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science, 20*, 347–353.
Golinkoff, R. M., Ma, W., 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.
Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics, 25*, 207–231.
Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1–22). Oxford: Blackwell.
Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Developmental Science, 9*, 309–315.
Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40*, 269–279.
Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy, 14*, 439–456.
Hollich, G. (2005). *Supercoder: A program for coding preferential looking (Version 1.5)* (Computer Software). West Lafayette, IN: Purdue University.
Hollich, G. (2006). Combining techniques to reveal emergent effects in infants' segmentation, word learning, and grammar. *Language & Speech, 49*, 3–19.
Horst, J. S. (2009). Novel Object and Unusual Name (NOUN) database Retrieved from: <http://www.sussex.ac.uk/wordlab/noun>.
Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones* (Vol. 6). Cambridge University Press.
Hu, J., Gao, S., Ma, W., & Yao, D. (2012). Dissociation of tone and segment processing in Mandarin idioms. *Psychophysiology, 49*, 1179–1190.
Hua, Z., & Dodd, B. (2000). The phonological acquisition of Putonghua (modern standard Chinese). *Journal of Child Language, 27*, 3–24.
Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23*, 319–330.
Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770–814.

Keitel, A., & Daum, M. M. (2015). The use of intonation for turn anticipation in observed conversations without visual signals as source of information. *Frontiers in Psychology, 6*, 108.
Keitel, A., Prinz, W., Friederici, A. D., Hofsten, C. V., & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology, 116*, 264–277.
Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*, 104–117.
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.
Ladd, D. R. (1996). *Intonational phonology* (Cambridge studies in linguistics 79). Cambridge: Cambridge University Press.
Laval, V., & Bert-Erboul, A. (2005). French-speaking children's understanding of sarcasm: The role of intonation and context. *Journal of Speech, Language, and Hearing Research, 48*, 610–620.
Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage, 20*, 2142–2152.
Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language, 4*, 185–199.
Lin, C. Y., Wang, M., & Shu, H. (2013). The processing of lexical tones by young Chinese children. *Journal of Child Language, 40*, 885–899.
Liu, L., & Kager, R. (in press). Perception of tones by bilingual infants learning non-tone languages. *Bilingualism: Language and Cognition*.
Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition, 133*, 385–394.
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.
Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and speech, 47*, 109–138.
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.
Lotto, A., & Holt, L. (2011). Psychology of auditory perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 2*, 479–489.
Ma, W., & Golinkoff, R. (2016, November). *Syntactic bootstrapping for form class distinction in Mandarin child-directed speech.* In *The Boston University conference on language development*, Boston, MA.
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.
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.
Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4*, 540–545.
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.
Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language, 57*, 252–272.
Männel, C., & Friederici, A. D. (2011). Intonational phrase structure processing at different stages of syntax acquisition: ERP studies in 2-, 3-, and 6-year-old children. *Developmental Science, 14*, 786–798.
Männel, C., Schipke, C. S., & Friederici, A. D. (2013). The role of pause as a prosodic boundary marker: Language ERP studies in German 3-and 6-year-olds. *Developmental Cognitive Neuroscience, 5*, 86–94.
Matisoff, J. A. (1973). *The grammar of Lahu* (Vol. 75). Univ of California Press.
Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*, 241–265.
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.
Mithen, S. J. (2006). *The singing Neanderthals: The origins of music, language, mind and body*. Cambridge, MA: Harvard Univ. Press.
Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development, 21*, 373–377.
Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development, 72*, 834–843.
Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences U.S.A., 104*, 15894–15898.
Musso, M., Weiller, C., Horn, A., Glauche, V., Umarova, R., Hennig, J., ... Rijntjes, M. (2015). A single dual-stream framework for syntactic computations in music and language. *NeuroImage, 117*, 267–283.
Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition, 98*, 13–30.
Newman, R. S. (2006). Perceptual restoration in toddlers. *Perception & Psychophysics, 68*, 625–642.
Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children. *Studies in Second Language Acquisition, 18*, 1–25.
Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S., & Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology, 111*, 14–25.
Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688–691.
Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421–435.
Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1*, 127–152.
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.
Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development, 83*, 236–250.
Repp, B., & Lin, H. (1990). Integration of segmental and tonal information in speech perception: Across-linguistic study. *Journal of Phonetics, 18*, 481–495.
Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31*(1), 97–113.
Schirmer, A., Tang, 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.
Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech, 34*, 145–156.
Singh, L., & Chee, M. (2016). Effects of tone and intonation on spoken word recognition in early childhood. *Journal of Phonetics, 55*, 109–118.
Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition, 124*, 128–142.
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*, 94–109.
Singh, L., Poh, F. L. S., & Fu, C. S. L. (2016). Limits on monolingualism? A comparison of monolingual and bilinguals infants' abilities to integrate lexical tone in novel word learning. *Frontiers in Psychology, 7*. http://dx.doi.org/10.3389/fpsyg.2016.00667.
So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*, 273–293.
Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166.
Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13*, 480–484.
Taft, M., & Chen, H. (1992). Judging homophony in Chinese: The influence of tones. In H. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 151–172). Oxford, England: North-Holland.
Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin-speakers early vocabularies. *Developmental Psychology, 32*, 492–504.
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.
Thompson, W. F., Marin, M. M., & Stewart, L. (2012). Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proceedings of the National Academy of Sciences, U.S.A., 109*, 19027–19032.
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.
Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science, 11*, 188–195.
Vaish, A., & Striano, T. (2004). Is visual reference necessary? Contributions of facial versus vocal cues in 12-month-olds' social referencing behavior. *Developmental Science, 7*, 261–269.
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.
Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*, 25–47.
White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114–132.
Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. *Language and Speech, 59*, 59–82.
Wong, P. (2012a). Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics, 40*, 141–151.
Wong, P. (2012b). Monosyllabic Mandarin tone productions by three-year-old children growing up in Taiwan and the U.S.: Inter-judge reliability and perceptual results. *Journal of Speech, Language, and Hearing Research, 55*, 1423–1437.

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*, 434–443.
Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-year-old Mandarin-speaking children. *Journal of Speech, Language, and Hearing Research, 48*, 1065–1079.
Ye, Y., & Connine, C. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14*, 609–630.
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.
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.