# When does native language input affect phonetic perception? The precocious case of lexical tone

H. Henny Yeung[a,1], Ke Heng Chen[a,2], and Janet F. Werker[a]

[a]Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, B.C. V6T 1Z4, Canada

[1]Corresponding author, presently at the Laboratoire Psychologie de la Perception (CNRS UMR 8158), Université Paris Descartes, Paris Sorbonne Cité, 45 rue des Saints-Pères, 75006 Paris, France. Tel. phone: +33(0)1.42.86.21.96 and fax: +33(0)1.42.86.33.22

[2]Presently at the School of Audiology and Speech Sciences, University of British Columbia, 2177 Wesbrook Mall, Vancouver, B.C. V6T 1Z3, Canada

E-mail addresses: henny.yeung@parisdescartes.fr (H. Yeung), mokochen@gmail.com (K. Chen), jwerker@psych.ubc.ca (J. Werker).

**Abstract**

Previous studies have suggested that the perception of vowels and consonants changes from language-universal to language-specific between 6 and 12 months of age. This report suggests that language-specific perception emerges even earlier for lexical tones. Experiment 1 tested English-learners' perception of Cantonese tones, replicating declines in tone discrimination from 4 to 9 months of age. Experiment 2 tested infants learning *non-native* versus *native* tone systems (Mandarin-learners versus Cantonese-learners). All Chinese-learners discriminated the tones, but showed language-specific differences in tone *preferences* at both ages. Indeed, English-, Mandarin-, and Cantonese-learning 4-month-olds all exhibited distinct preferences. With other work, this shows that language-specific speech perception emerges over a more complex and extended schedule than previously thought: first for lexical stress and tone (<5 months), then vowels (6-8 months), consonants (8.5-12 months), and finally phoneme duration (18 months). Acoustic salience likely plays an important role in determining the timing of phonetic development.

*Keywords:* tone, infancy, speech perception, English, Chinese, Cantonese, Mandarin

## Introduction

It is widely acknowledged that infants begin perceiving phonetic contrasts in language-specific ways from 6 to 12 months of age. A commonly reported pattern is one of maintenance and decline, where young infants initially perceiving many native and non-native phonetic contrasts, but maintain sensitivity only to native contrasts as the perception of non-native contrasts declines (Anderson, Morgan, & White, 2003; Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Bosch & Sebastián-Gallés, 2003; Cheour et al., 1998; Pegg & Werker, 1997; Polka & Werker, 1994; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005; Werker & Lalonde, 1988; Werker & Tees, 1984a, 1984b). Of course, many other developmental patterns have now been reported. For example, if a particular consonant contrast is difficult to assimilate into the native language, like Zulu clicks into English, then infants as well as adults continue perceiving it well (Best & McRoberts, 2003; Best, McRoberts, & Sithole, 1988). Some types of vowel contrasts also remain acoustically salient throughout development (Polka & Bohn, 1996), due to language-universal biases in perception (Polka & Bohn, 2011). Other consonant contrasts are difficult to perceive in infancy, requiring language-specific input to learn (Narayan, Werker, & Beddor, 2010). Still more work shows that native language input does not just maintain perceptual sensitivity for some native phonetic contrasts, but also enhances it (Kuhl et al., 2006; Polka, Colantonio, & Sundara, 2001; Sundara, Polka, & Genesee, 2006; Tsao, Liu, & Kuhl, 2006).

This work has built a rich understanding of how phonetic development unfolds, but remains limited in that it draws almost exclusively from the empirical study of phonetic segments (i.e., vowels and consonants). Less is known about how infants learn to perceive native prosodic contrasts—phonetic distinctions related to pitch (i.e., fundamental frequency [f0]), duration, and/or amplitude—likely because the use of these cues is relatively limited in Indo-

European languages. One perspective is that prosodic contrasts develop on a similar schedule as most vowel and consonant contrasts. Language-specific perceptual patterns for many prosodic contrasts seem to emerge between 6 to 12 months of age, including lexical stress (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Jusczyk, Cutler, & Redanz, 1993; Skoruppa et al., 2009; Skoruppa, Cristia, Peperkamp, & Seidl, 2011) as well as pitch accent (Sato, Sogabe, & Mazuka, 2009; see also Nazzi, Floccia, & Bertoncini, 1998). Nevertheless, a direct comparison with phonetic segments is difficult, as vowels and consonants occur within individual syllables, while lexical stress and pitch accent are defined across multiple syllables. A comparison between segments and those prosodic cues defined along similar timescales (i.e., within syllables) would provide a more closely matched comparison.

One such prosodic cue is phoneme duration (i.e., contrasts between short versus long vowels or single versus geminate consonants). Duration is used in many languages to distinguish words (e.g., Japanese, Arabic, Dutch, Berber), and developmental studies of its perception have shown very different developmental trajectories compared to those of vowels and consonants. For example, infants learning Japanese are only able to discriminate duration contrasts by 9-10 months of age, and have much more difficulty doing so at earlier ages (Sato, Kato, & Mazuka, 2012; Sato, Sogabe, & Mazuka, 2010). Moreover, cues to vowel duration are not used in language-specific ways until at least 18 months of age (Dietrich, Swingley, & Werker, 2007; Minagawa-Kawai, Mori, Naoi, & Kojima, 2007; Mugitani et al., 2009), which suggests instead a markedly protracted developmental trajectory for prosodic cues relative to vowels and consonants.

Another such prosodic cue is lexical tone. Tones are primarily defined by f0 variations within single syllables, and are found in most of the languages native to the Americas, Sub-

Saharan Africa, as well as East and Southeast Asia (Yip, 2002). For example, Mandarin uses four tones, each of which identifies a different word when spoken on the same syllable. In Mandarin, "ma" can mean *mother* (i.e., a high level tone), *hemp* (i.e., a rising tone), *horse* (i.e., a low dipping tone), or *to scold* (i.e., a falling tone), and thus tone constitutes an important factor in lexical retrieval for adults (Cutler & Chen, 1997).

Here we ask how the development of lexical tone perception unfolds in infancy. First we review tone perception in adults and in infants, which raises two important issues: how do infants converge on their native (versus non-native) tone system, and how does the developmental trajectory of tone perception compare to that of other phonetic units? Then we describe two experiments investigating these questions. Finally, we conclude by discussing the implications of this work for phonetic development more broadly.

**Lexical tone and its perception in adults**

Secondary acoustic cues to tone may include vowel duration (e.g., Blicher, Diehl, & Cohen, 1990; Gandour & Harshman, 1978), or voice quality such as creaky voice (e.g., Gottfried & Suiter, 1997), but it is widely agreed that the two primary acoustic cues to tone are the variations in f0 level (e.g., high, middle, low), and/or f0 contour (e.g., steady, rising, falling) that occur within single syllables (Gandour, 1981; Gandour & Harshman, 1978; Khouw & Ciocca, 2007; Vance, 1976). Speakers of tone languages must negotiate the linguistic function of tones at segmental timescales with both foot- and phrase-level prosodic cues that have both grammatical (e.g., Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Snedeker & Trueswell, 2002), and pragmatic functions in adults (e.g., Bock & Mazella, 1983; Dahan, Tanenhaus, & Chambers, 2002; Gussenhoven, 2004; Welby, 2003). Indeed, phonetic studies show that acoustic instantiations of tone strongly interact with the wider acoustic context, particularly carry-over from the preceding syllable and place within an utterance (see Xu, 1999 for review). However, adult listeners are still adept at extracting identifiable f0 patterns in the face of this variability, particularly from the latter portion of the vowel, where canonical tone patterns are particularly robust (Khouw & Ciocca, 2007; Xu, 1997).

Learning to perceive non-native tone contrasts can be difficult for adult speakers of non-tone languages (Wang, Spence, Jongman, & Sereno, 1999), just as adults have trouble learning non-native vowel and consonant contrasts (Flege, Bohn, & Jang, 1997; Iverson & Evans, 2007; Lively, 1993; Pallier, Bosch, & Sebastián-Gallés, 1997). Initial reports further suggested that speakers of a tone language are better at learning, identifying, and remembering all tone contrasts--even if the tones come from an unfamiliar language--than speakers of any non-tone language (Lee, Vakoch, & Wurm, 1996; Wayland & Guion, 2004). Later studies argued that a strict dichotomy between speakers of tone and non-tone languages is too simplistic, since non-native tone perception interacts in nuanced ways with the use of f0 in any language. For example, certain tone contrasts that are acoustically similar to phrase-level f0 distinctions can be quite easy for speakers of even non-tonal languages to distinguish.

There nevertheless remain important differences between the perception of f0 cues that signify *post-lexical* information such as grammatical distinctions, and the *lexical* information such as word identity (Braun & Johnson, 2011). Speakers of lexical tone languages show greater degrees of categorical perception for f0 cues compared to non-tone language speakers, and this happens in both speech and non-speech contexts (Francis, Ciocca, Wong, Leung, & Chu, 2006; Hallé, Chang, & Best, 2004; Peng et al., 2010; Xu, Gandour, & Francis, 2006). Moreover, several ERP studies have suggested that the brain signatures of tone processing are speeded and are asymmetric (towards the left-hemisphere) for tone language speakers compared to non-tone language speakers, supporting the idea that exposure to a lexical system that uses tone will change the way that the perceptual system encodes f0 cues (Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Krishnan, & Gandour, 2007; Kaan, Barkley, Bao, & Wayland, 2008; Kaan, Wayland, Bao, & Barkley, 2007; Luo et al., 2006; Xi, Zhang, Shu, Zhang, & Li, 2010). This suggests, even though speakers of all languages must pay some attention to f0, that important differences at the level of f0 perception can be found between speakers of tone and non-tone languages.

**Lexical tone and its perception in infants**

For infants it seems especially challenging to identify linguistically important f0 variation when these same cues also mark affective and communicative functions in infant-directed speech (Fernald, 1989; Papoušek, Bornstein, Nuzzo, Papoušek, & Symmes, 1990; Spence & Moore, 2003; Stern, Spieker, & MacKain, 1982). Only a few papers have begun to ask how infants begin to successfully unscramble and identify f0 cues in word segmentation and word learning tasks (Bortfeld & Morgan, 2010; Quam & Swingley, 2010, 2012; Singh & Foong, 2012; Singh, Morgan, & Best, 2002; Singh, Morgan, & White, 2004; Singh, White, & Morgan, 2008), and even fewer reports (reviewed below) have examined the phonetic perception of tone in infancy.

Tsao (2008) showed that the perception of Mandarin tone contrasts in 10- to 12-month-olds was easiest for more acoustically distinct tones, and harder for those that were acoustically more similar, echoing previous studies with adults (e.g., So & Best, 2010). Only three other published studies have further compared across different ages or across language groups (Harrison, 2000; Mattock & Burnham, 2006; Mattock, Molnar, Polka, & Burnham, 2008). These are discussed below in light of two central questions about the development of tone perception.

First, how do infants begin learning native versus non-native tone systems? Second, is the trajectory of perceptual development similar for tones, vowels, and consonants?

**Native versus non-native tone systems.** A seminal study by Mattock and Burnham (2006) showed the typical pattern of maintenance and decline in speech perception: English-learning infants discriminated a Thai tone contrast at 6 but not at 9 months of age, while Chinese-learning infants (i.e., a mixed group of Cantonese and Mandarin learners) discriminated the tone contrast at both ages. These authors concluded that Chinese learners maintain perceptual sensitivity for tones from 6 to 9 months of age, even if those tones do not come from the native system. In contrast, English learners show a decline in attention to tones within this same period of development.

Recall, however, that adult researchers have recently argued that a simple dichotomy between speakers of tone versus non-tone languages is too simplistic. So and Best (2010) suggested that Cantonese speakers' perception of Mandarin tones is at least partly influenced by patterns of assimilation into native tone categories (see also Best, 1995). Francis et al. (2008) suggested that tone perception is determined by language-specific acoustic weightings, invoking Gandour and colleagues' (Gandour, 1983; Gandour & Harshman, 1978) identification of distinct perceptual dimensions for tone. Languages like Yorùbá make tonal distinctions based on the perceived height of the tone (i.e., *level*), while speakers of Mandarin mainly use the direction(s) of pitch change (i.e., *direction*). Speakers of other languages, like Thai or Cantonese, use a combination of both cues to distinguish tones (see also Vance, 1976). Francis et al. (2008) used multidimensional scaling techniques to show that English speakers weighted f0 level more heavily than f0 direction when making identifications of Cantonese tones, but that Mandarin speakers use the inverse weighting. Interestingly, both groups added weight to the other, unattended dimension after several training sessions.

Consider again the results of Mattock and Burnham (2006), who failed to find any differences between sub-groups of Cantonese- and Mandarin-learners in their infant sample. Two reasons may underlie this null result: first, their study was not designed to examine differences between Chinese sub-groups, and thus there may have been a lack of power when dividing the sample between Cantonese- and Mandarin-learners. A second possibility is that both groups of Chinese-learning infants may still *discriminate* many non-native tones by 9 months of age, but show *preferences* influenced by the native language, which could have been hard to detect using their methodology (i.e., a conditioned head-turn procedure). The first goal of the present study was to address these issues, using a more sensitive testing procedure to ask how the development of Cantonese tone perception unfolds among infants hearing tones *de novo* (English-learners), hearing non-native tones (Mandarin-learners), and hearing native tones (Cantonese-learners).

**The trajectory of perceptual development.** Previous work on infant tone perception has also raised the question of *when* language-specific influences emerge with respect to vowels and consonants. For example, language-specific perceptual patterns for vowels are reported as early as 6 months of age (Cheour et al., 1998; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994), and continue to develop until at least 8-12 months of age (Bosch & Sebastián-Gallés, 2003; Polka & Bohn, 1996), while language-specific consonant perception is first seen by 8.5 months of age (Anderson et al., 2003), and continues emerging from 10-12 months of age (see Werker & Gervain, in press, for review). This raises the question of whether tone perception better aligned with vowels or with consonants in their developmental trajectory.

This was the topic of a recent report testing French- and English-learning infants at 4, 6, and 9 months of age on the same Thai tones used by Mattock and Burnham (2006) (Mattock et al., 2008). Crucially, both 4- and 6-month-old French- and English-learning infants were equivalently successful in discriminating tones, but 9-month-olds from either language group were not. This suggests that language-specific perception of tone does not emerge until at least 6 months of age, which is more similar to the development of consonant perception than to vowel perception. This is somewhat perplexing, given that tones are instantiated on the vocalic parts of syllables. Yet, as Mattock et al. (2008) point out, this finding is not surprising given other considerations. First, tones are classified differently from vowels in linguistic analysis (Yip, 1995, 2002), independently motivating developmental differences between vowels and tones. Second, the perception of other prosodic units (i.e., lexical stress) also seems to change in this age range: 6-month-olds show no behavioural preferences for language-dominant stress patterns, but 9- or 10-month-olds do (Höhle et al., 2009; Jusczyk et al., 1993; Turk, Jusczyk, & Gerken, 1995; although see Friederici, Friedrich, & Christophe, 2007 for reports of language-specific stress perception by 4 months of age when measuring electrophysiological responses).

There are nevertheless some reasons to suspect that language-specific tone perception first emerges earlier than previously thought, as no study has yet compared learners of non-tone and tone languages when vowel perception is still considered language-independent (i.e., at 4 months of age). Harrison (2000), for example, reported a small study (*n* =12) that suggests there may indeed already be cross-linguistic differences. Here, 6- to 8-month-old Yorùbá- and English-learning infants' perception of f0 level was assessed, as this is the primary acoustic cue distinguishing tones from Yorùbá. By this age, Yorùbá-learning infants were already performing slightly better at f0 discrimination compared to English-learning infants. Moreover, previous work on tone perception has measured discrimination, while studies on the perception of other prosodic units, like lexical stress, often report infants' *preferences* in addition to (or instead of) discrimination (Echols, Crowhurst, & Childers, 1997; Höhle et al., 2009; Jusczyk et al., 1993; Pons & Bosch, 2010; Turk et al., 1995). Although a preference for one language pattern over another is sufficient to imply discrimination, discrimination by itself does not necessarily imply a preference. Correspondingly, several reports show language-specific differences emerging in infancy that manifest as preferences (Höhle et al., 2009) or asymmetric discrimination patterns (Friederici et al., 2007; Kuhl et al. 1992), even as infants of all language backgrounds remain at least partially capable of discriminating these contrasts.

Here we asked whether language-specific patterns of tone perception emerge earlier that previously reported, using a testing procedure assessing both preference *and* discrimination in 4- and 9-month-olds learning either non-tone or tone languages. Mattock et al.'s (2008) previous procedure was adapted, where infants were given 'alternating' trials containing two unique tone types, as well as 'non-alternating' trials containing only one tone type. Differences in looking between these trials types implies discrimination of the stimuli (Best & Jones, 1998), and Mattock et al.'s (2008) implementation of this procedure suggested that infants would prefer looking at the alternating trial type when discriminating the tone contrast. Here we modified the procedure by giving each infant alternating trials containing the same syllables with either Tone X or Tone Y, non-alternating trials containing one tone (Tone X), and non-alternating trials containing the other tone (Tone Y). With this procedure we could simultaneously measure both discrimination (by observing any differences in looking across the three trial types), and preference (by measuring the pattern of preferences for alternating, non-alternating Tone X, and non-alternating Tone Y trials).

**Overview of the two experiments**

Here we tested infants learning English, Mandarin, and Cantonese. The latter two are officially considered dialects of Chinese, but it remains important to note that they are often considered different languages altogether (like Spanish and Portuguese) due to substantial differences at morphosyntactic, lexical, as well as phonological levels (including their tonal systems) that render them mutually unintelligible.

Tones are described here by an impressionistic notational system (Chao, 1968), where ‘1’ designates the lowest level of a speaker’s f0 range, and ‘5’ the highest. In this system, the first number denotes a starting f0 level and subsequent numbers denote inflection points in the f0 contour, or the ending f0 level. For example, Mandarin tones would be notated as a ‘high level’ tone 55, a ‘rising’ tone 25, a ‘falling’ tone 51, and a ‘low dipping’ tone 214. Cantonese, on the other hand, has six contrastive tones (Bauer & Benedict, 1997). Three of these are level tones, where f0 dips only slightly from beginning to end: a ‘high level’ tone 55, a ‘mid level’ tone 33, and a ‘low level’ tone 22. The other three Cantonese tones are contours: a ‘high rising’ tone 25, a ‘low rising’ tone 23, and a ‘low falling’ tone 21. Although citation forms of Cantonese and Mandarin tones differ dramatically in precise f0 range, length, and endpoints, there is also quite a bit of overlap between these two systems. Figure 1 illustrates the four Mandarin tones and six Cantonese tones.

[Figure 1 about here]

In Experiment 1 we tested 4- and 9-month-old English-learning infants' perception of a Cantonese tone contrast. In Experiment 2, we ran an identical procedure on two groups of Chinese-learning infants: Mandarin-learners, for whom the tone contrast is non- native, and Cantonese-learners, for whom the contrast is native. Among the set of possible Cantonese contrasts, we chose the perception of the high rising tone ('Tone 25') versus the mid level tone ('Tone 33') for several reasons. First, this contrast would be the most similar to the Thai contrast used by Mattock and colleagues. Second, we could compare perceptual preferences in the Mandarin groups between one tone type easily assimilated to the native language (i.e., Tone 25 is very similar to the Mandarin rising tone) and one tone type that is not (i.e., Tone 33 is dissimilar to any Mandarin tone). Third, we could examine, within each group, infants' relative preference for contour (i.e., Tone 25) versus level (i.e., Tone 33) tones.

## Experiment 1

### Method

**Stimuli.** Cantonese tones were instantiated on a CV syllable, pronounced "chee" (/t͡ɕʰi/), written as *qi* following commonly used Cantonese and Mandarin romanizations[1]. These speech tokens were recorded in a sound-attenuated booth from an adult female native speaker of Cantonese, who produced sentences in an adult-directed register that included the target syllable with either Tone 25 (此 'this; thus'; 始 'start') or Tone 33 (次 'next'; 刺 'thorn'). Four tokens of each tone type were isolated, and all tokens were normalized for amplitude. Acoustic measures showed that the vowels of Tone 25 and Tone 33 stimuli overlapped in both duration (Tone 25: *M* = 439 ms, *R* = 422 - 466 ms; Tone 33: *M* = 462 ms, *R* = 454 - 469 ms), as well as in formant values and formant trajectories (Table 1). Figure 2 further plots the pitch tracks measured from each token.

[Table 1 and Figure 2 about here]

**Procedure.** The experiment took place in a sound-attenuated and dimly lit room. The infant sat on the parent's lap approximately 36 inches away from a 27-inch TV screen, which was placed in an opening in the middle of a black curtain. Auditory attention was measured by recording looking time towards a static visual stimulus (a black and red checkerboard) as infants were simultaneously presented with auditory tokens. The parent, who was listening to music over headphones throughout the study, was instructed not to speak and not to point at the screen. A video camera was hidden under the TV screen, and an experimenter could observe the infant's eye gaze direction from a small monitor in another room, where stimuli presentation using Habit 2002 (Cohen, Atkinson, & Chaput, 2000) was controlled. Sound was presented to infants at a level of about 65 dB. The experimenter was blind to the sound presented, and recorded infant looking times by pressing a button. In all trials, auditory and visual presentation continued until an infant looked away for 2 s, at which point the visual stimulus disappeared and a silent animation appeared to attract the attention of the infant. Once the infant looked again, the next trial began. Each trial had a maximum looking time of 30 s.

Infants accumulated 30 s of looking time to one tone type in a *familiarization phase*. Two separate strings were created for this (i.e., infants heard either Tone 25 or Tone 33), and each contained a pseudo-random order of all four tokens. In a subsequent *test phase* of eight trials, discrimination was assumed if infants looked longer at 'alternating' trials (i.e., Alt trials), which contained two stimuli types, compared to either type of 'non-alternating' trials (i.e., Non-Alt trials), which contained only one type. Both Alt and Non-Alt trials each contained four tokens.

Two Alt trials began with Tone 25, and the other two trials began with the Tone 33. Pseudo-random orders of Non-Alt trials were also created (i.e., two pseudo-random orders for Tone 25, and two orders for Tone 33). A 1 s inter-stimulus interval separated individual tokens in all trials.

In the test phase Alt and Non-Alt test trials were presented in rotating order (i.e., N-A-N-A-N-A-N-A or A-N-A-N-A-N-A-N). Each infant heard two Non-Alt trials containing Tone 25 (i.e., 'Tone 25 trials'), and two Non-Alt trials containing Tone 33 (i.e., 'Tone 33 trials'), and each kind of trial was heard at least once in each half of the test phase. This allowed us to calculate looking for Tone 25 trials and or Tone 33 trials separately. Across all infants, the tone type heard during familiarization as well as the order of Alt, Tone 25, and Tone 33 trials were maximally counterbalanced. Table 2 illustrates one experimental order for illustrative purposes. Trained observers who were blind to the auditory stimuli coded infant looking behaviour offline.

[Table 2 about here]

**Participants.** Two groups of English-learning infants participated. The first group consisted of 24 English-learning 4-month-old infants (12 boys, *range:* 3 mo 20 d – 4 mo 15 d). Eight additional infants were not included in the analysis for the following reasons: occasionally hearing a tone language in their home environment (n = 1); excessive fussiness (n = 4); experimenter / equipment error (n = 2). The second group consisted of 24 English-learning 9-month-old infants (12 boys, *range:* 8 mo 15 d – 9 mo 17 d). Seven additional infants were not included in the analysis for the following reasons: excessive fussiness (n = 4); experimenter / equipment error (n = 3). All infants were exposed to English at least 90% of the time (by parental report) without hearing any language containing tone contrasts.

**Results**

Looking times were analysed in an omnibus ANOVA with a within-subjects factor of test trial TYPE with three levels (Tone 25, Tone 33, Alt), and between-subjects factors of GENDER (male, female), FAMILIARIZATION (Tone 25, Tone 33), and AGE (4 months, 9 months). A significant interaction of TYPE x AGE was detected, $F(2, 80) = 3.49$, $p = .035$, $\eta^2 = .080$, indicating that looking among test trials differed between the younger and older age group. No other interactions or main effects reached significance, indicating that neither gender nor the type of tone used for familiarization affected discrimination performance (*alpha* = .05). One-way repeated-measures ANOVAs with a within-subjects factor of TYPE were conducted at each age. Results showed a significant main effect of TYPE among 4-month-olds, $F(2, 46) = 3.47$, $p = .040$, $\eta^2 = .13$, but not among 9-month-olds, $F(2, 46) = 1.16$, $p = .32$. As illustrated in Figure 3, this showed that 4-month-olds discriminated tones, but that English-learning 9-month-olds did not discriminate tones, looking equivalently at all trial types.

As we were interested in examining 4-month-olds' preferences among the different test trial types, looking times to Tone 25 trials, $M = 6.30$, $SD = 5.31$, Tone 33 trials, $M = 7.63$, $SD = 5.92$, and Alt trials, $M = 8.28$, $SD = 5.45$ were analysed in a pairwise fashion using Fisher-Hayter's correction for multiple comparisons (Hayter, 1981; Seaman, Levin, & Serlin, 1991), as was done in all such comparisons reported presently. Results showed that looking to Tone 25 trials was marginally less than to Tone 33 trials, although this difference did not reach statistical significance, $M_{difference} = 1.34$, *95% CI* [-.22, 2.89], $p < .10$. When comparing Alt trials to each type Non-Alt trial, looking was longer to Alt than to Tone 25 trials, $M_{difference} = 1.98$, *95% CI* [.28, 3.54], $p < .05$, but was equivalent between Alt and Tone 33 trials, $M_{difference} = .65$, *95% CI* [-.91, 2.20], $p > .10$.

[Figure 3 about here]

**Discussion**

Experiment 1 replicated Mattock et al. (2008) by showing that English-learning 4-month-olds successfully discriminated a high-rising versus mid-level tone contrast, while 9-month-olds showed no discrimination. However, our modified procedure also revealed a more nuanced pattern of looking within the 4-month-old group. Looking at the Alt trials was *not* longer than *both* types of Non-Alt trials. The observed pattern of longer looking at the Alt trials compared to Tone 25 trials was expected, but 4-month-olds did not look longer at the Alt trials compared to the Tone 33 trials.

There are two general explanations for this result. One possibility, which is consistent with claims made previously in the literature, is that listening preferences for tones at 4 months of age are independent of language experience, driven only by acoustic factors. On this account, infants looked longer at Alt trials because they preferred listening to two tone types within a trial than to just one tone type, but at the same time also had greater interest in Tone 33 trials due to other kinds of acoustic factors. For example, perhaps unchanging f0 contours are easier to process, which increased overall looking to Tone 33 trial types, yielding the observed pattern.

A second possibility is that 4-month-olds' preferences reflect developing sensitivity to English prosody. For example, Tone 25 trials may have been regarded as particularly anomalous, where a series of syllables with marked f0 changes of any kind rarely occurs in English[2]. Tone 33 trials might have been regarded as less anomalous, as f0 changes in English usually vary over much larger prosodic units than the syllable, and a series of relatively constant f0 sequences is perhaps less marked (Pierrehumbert, 1980). On this account, the looking patterns observed in Experiment 1 are similarly explained by two tendencies. Infants preferred Alt trials to Non-Alt trials, but their experience with English also lead them to prefer Tone 33 trials because the relatively flat prosody was more familiar compared to the f0 contours in Tone 25 trials.

To distinguish between these explanations, a cross-linguistic comparison is needed. In Experiment 2, we tested several groups of infants in the same procedure at both 4 and 9 months of age who were learning either Mandarin or Cantonese.

## Experiment 2

The procedure and design from Experiment 1 was repeated with separate groups of 4- and 9-month-olds learning both Mandarin and Cantonese. In our non-native Mandarin group, infants heard a tone contrast that included one tone that could be assimilated to the native inventory (i.e., Tone 25), but another tone that was a relatively poorer match (i.e., Tone 33, which is lower than the only other level tone [55] in Mandarin).

### Method

**Stimuli.** The same stimuli from Experiment 1 were used. Single syllables in either Cantonese or Mandarin are often polysemous given the limited phonological inventories in these languages, and so “qi” is a fairly common syllable—either in isolation or in lexical compounds—when examining Cantonese (i.e., Tone 25: 此 ‘this; thus’ or 始 ‘start’; Tone 33: 次 ‘next’ or 刺 ‘thorn’) and Mandarin (i.e., the rising tone: 其 [possessive particle] or 奇 ‘strange’) (Da, 2004). It is thus very likely that infants in both language groups had encountered this syllable with at least one of these tones before, but it is unlikely that infants in any experimental group treated our stimuli as lexical entries (Tardif, Fletcher, Zhang, & Liang, 2008). This assumption is reinforced by the fact that our stimuli were not presented in any kind of referential context (Fennell & Waxman, 2010; Stager & Werker, 1997).

**Participants.** Two groups of Mandarin-learning infants participated. The first group consisted of 16 Mandarin-learning 4-month-old infants (8 boys, *range:* 3 mo 15 d – 4 mo 15 d). One additional infant was tested but not included due to experimenter / equipment error (n = 1). The second group consisted of 24 Mandarin-learning 9-month-old infants (12 boys, *range:* 8 mo 17 d – 9 mo 26 d). Eight additional infants were tested but not included in the analysis for the following reasons: excessive fussiness (n = 6); experimenter / equipment error (n = 2).

Two groups of Cantonese-learning infants also participated. The first group consisted of 16 Cantonese-learning 4-month-old infants (9 boys, *range:* 3 mo 17 d – 4 mo 19 d). Three additional infants were tested but not included in the analysis for the following reason: excessive fussiness (n = 2); experimenter / equipment error (n = 1). The second group consisted of 24 Cantonese-learning 9-month-old infants (13 boys, *range:* 8 mo 16 d – 9 mo 26 d). Eight additional infants were tested but not included in the analysis for the following reasons: excessive fussiness (n = 6); experimenter / equipment error (n = 2).

All infants were raised in metropolitan Vancouver, British Columbia, and were likely exposed to some English in their daily lives. Nevertheless, most infants lived in communities where either Mandarin or Cantonese are widely spoken, and none were spending significant amounts of time outside of the home with English-speaking caregivers. A few families reported that their infants heard other regional dialects of Chinese from friends or family members (n = 5 in our total sample of 80 infants), but with these few exceptions, infants were reported by parents to be hearing either exclusively Mandarin or exclusively Cantonese at least 90% of the time, hearing mostly English in the remaining 10%.

**Results**

Two main questions motivated the current experiment. How results bear on each of these questions is described separately below.

**Native and non-native tone systems.** One goal of the study was to ask whether Cantonese- and Mandarin-learning infants' begin perceiving tones as either *native* or *non-native* from early in development, and if not, then when language-specific perception emerges. To examine this question, looking time from Experiment 2 was analyzed in a mixed-effects omnibus ANOVA with a within-subjects factor of test trial TYPE (Tone 25, Tone 33, or Alt), and between-subjects factors[3] of FAMILIARIZATION (Tone 25, Tone 33), AGE (4 months, 9 months), and LANGUAGE (Cantonese, Mandarin). There was a main effect of TYPE, $F(2, 144) = 9.12$, $p < .01$, $\eta^2 = .11$, indicating that Chinese infants could discriminate the tones as a group. In addition, there was a significant 3-way interaction between the factors of TYPE, FAMILIARIZATION, and LANGUAGE, $F(2, 144) = 3.97$, $p = .021$, $\eta^2 = .052$, but no other significant interactions or main effects (*alpha* = .05).

Results differed from Experiment 1 in that there was no effect of AGE, showing that Chinese-learning infants show similar patterns across both age groups. However, the 3-way interaction indicated that looking to test trials differed as a function of whether infants were familiarized with Tone 25 or Tone 33, and whether the infants were learning Cantonese or Mandarin. To explore how these factors affected tone perception, both the effects of familiarization and of language exposure were explored separately below, collapsing across age.

*Effect of familiarization* (*Figure 4)*. Follow-up analyses were conducted within each familiarization sub-group. In the Tone 25 familiarization group ($n$ = 40), there was both an interaction of TYPE x LANGUAGE, $F(2, 76) = 4.70$, $p = .012$, $\eta^2 = .11$, and main effect of TYPE, $F(2, 76) = 7.76$, $p = .001$, $\eta^2 = .17$ (*alpha* = .05). Thus, Chinese-learning infants successfully discriminated the tones when familiarized with Tone 25, and further showed differences between Cantonese and Mandarin sub-groups (described below).

In contrast, infants in the Tone 33 familiarization group ($n$ = 40) showed no significant effects: neither a significant main effect of TYPE, $F(2, 76) = 1.45$, $p = .24$, nor an interaction of TYPE x LANGUAGE, $F(2, 76) = .268$, $p = .77$ (*alpha* = .05). Infants familiarized with Tone 33 thus showed no evidence of tone discrimination, indicating that any tone discrimination in the Experiment 2 was carried entirely by those infants familiarized with Tone 25.

[Figure 4 about here]

*Effect of hearing Mandarin or Cantonese (Figure 5)*. Within the Tone 25 familiarization group, the TYPE x LANGUAGE interaction suggested that looking times were influenced by whether infants were learning Mandarin or Cantonese. Follow-up analyses were conducted within Mandarin- and Cantonese-learning sub-groups of those infants hearing Tone 25 during the familiarization phase.

In the Mandarin-learning group ($n$ = 20), a main effect of TYPE was observed, $F(2, 38) = 9.87$, $p < .001$, $\eta^2 = .34$. Here we examined Mandarin-learners' pattern of preferences to see how they differed from Cantonese-learners. Looking times to Tone 25 trials, $M = 7.86$, $SD = 3.53$, Tone 33 trials, $M = 5.19$, $SD = 3.22$, and Alt trials, $M = 8.23$, $SD = 3.91$ were analysed in a pairwise fashion. Results showed that looking to Tone 25 trials was significantly longer than to Tone 33 trials, $M_{difference} = 3.04$, *95% CI* [1.48, 4.60], $p < .01$. When comparing Alt trials to each type of Non-Alt trial, looking was equivalent to Alt and Tone 25 trials, $M_{difference} = .37$, *95% CI* [-1.87, 1.93], $p > .10$, but was longer to Alt than to Tone 33 trials, $M_{difference} = 3.04$, *95% CI* [1.48, 4.60], $p < .01$.

In the Cantonese-learning group ($n = 20$), a main effect of TYPE was also observed, $F(2, 36) = 3.95$, $p = .028$, $\eta^2 = .17$. Looking times to Tone 25 trials, $M = 5.45$, $SD = 2.72$, Tone 33 trials, $M = 6.43$, $SD = 3.98$, and Alt trials, $M = 8.07$, $SD = 4.63$ were again analysed in a pairwise fashion, which showed differences from the Mandarin group. Specifically, Tone 25 trials were not significantly different than Tone 33 trials, $M_{difference} = .39$, *95% CI* [-1.08, 1.85], $p > .10$. Moreover, when comparing Alt trials to each type of Non-Alt trial, longer looking was observed to Alt than to Tone 25 trials, $M_{difference} = 2.63$, *95% CI* [.67, 4.59], $p < .01$. Similar to Mandarin-learning infants, however, looking was slightly longer to Alt trials than to Tone 33 trials, but this difference was only marginally significant, $M_{difference} = 1.64$, *95% CI* [-.33, 3.61], $p < .10$.

Overall, these results show that both Mandarin- and Cantonese-learning infants (at least when familiarized with Tone 25) show language-specific patterns of tone preference that remain stable from 4 to 9 months of age. More precisely, Mandarin-learning infants preferred Alt and Tone 25 trials to Tone 33 trials. Unlike Mandarin-learners, Cantonese-learners preferred Alt trials to Tone 25 trials, and did not show significant differences in looking between Tone 25 and Tone 33 trials.

[Figure 5 about here]

**The trajectory of perceptual development (Figure 6).** The second goal of the study was to ask whether differences between language groups in tone perception can be detected at 4 months of age, when tones are supposed to be perceived on a "language-universal" basis (Mattock et al., 2008). The present results have already indicated that infants in Experiment 2 have language-specific preferences without any interaction of age, so to more fully assess this possibility, we asked how 4-month-old English-learning infants from Experiment 1 would compare to both Mandarin- and Cantonese-learning 4-month-olds.

An ANOVA was conducted with a within-subjects factor of test trial TYPE (Tone 25, Tone 33, Alt), and between-subjects factors of FAMILIARIZATION (Tone 25, Tone 33) and LANGUAGE (English, Cantonese, Mandarin). Results show that there was a main effect of TYPE, $F(2, 100) = 8.66$, $p < .001$, $\eta^2 = .15$, showing again that 4-month-olds looked differently at the test trial types. Results also revealed a marginal interaction of TYPE x LANGUAGE, $F(4, 100) = 2.03$, $p = .096$, $\eta^2 = .075$, but no other interactions or main effects were significant (*alpha* = .10). This result, in conjunction with Experiment 1 and the analysis so far in Experiment 2, suggested that there were differences among 4-month-olds by language group. Four-month-olds are analyzed separately by language below.

*English-learning 4-month-olds*. Recall from Experiment 1 that infants looked significantly longer at Alt trials than at Tone 25 trials.

*Mandarin-learning 4-month-olds*. A pairwise analysis was performed on looking to Tone 25 trials, $M = 7.75$, $SD = 4.70$, Tone 33 trials, $M = 6.09$, $SD = 3.80$, and Alt trials, $M = 8.63$, $SD = 4.73$. Results showed that looking to Tone 25 trials was longer than to Tone 33 trials, but this difference did not reach significance, $M_{difference}$ = 1.65, *95% CI* [-.04, 3.35], $p < .10$. When comparing Alt trials to each type of Non-Alt trial, infants looked equivalently at Alt trials compared to Tone 25 trials, $M_{difference}$ = .88, *95% CI* [-.81, 2.58], $p > .10$, but look significantly longer at Alt trials than at Tone 33 trials, $M_{difference}$ = 2.53, *95% CI* [.84, 4.23], $p < .01$. Results thus showed a very different pattern of looking from English-learning infants.

*Cantonese-learning 4-month-olds*. A pairwise analysis was performed on looking to Tone 25 trials, $M$ = 6.59, $SD$ = 3.64, Tone 33 trials, $M$ = 6.30, $SD$ = 5.31, and Alt trials, $M$ = 8.36, $SD$ = 4.27. Results showed that looking to Tone 25 trials was equivalent to Tone 33 trials, $M_{difference}$ = .39, *95% CI* [-1.08, 1.85], $p > .10$. When comparing Alt trials to each type of Non-Alt trial, infants looked significantly longer to Alt trials than to Tone 25 trials, $M_{difference}$ = 1.77, *95% CI* [.31, 3.23], $p < .05$, and also looked significantly longer to Alt trials than to Tone 33 trials, $M$ = 6.30, $SD$ = 5.31, $M_{difference}$ = 2.54, *95% CI* [.84, 4.23], $p < .01$. Results thus differed from both English- and Mandarin-learning groups in showing a clear pattern of discrimination, where looking at Alt trials was longer than to either type of Non-Alt trial.

[Figure 6 about here]

**Discussion**

Our results replicated previous work on tone perception in Chinese-learning infants, showing stable discrimination abilities from early in the first-year of life (i.e., 6 months in Mattock & Burnham, 2006; 4 months in the present study) until 9 months of age. However, our results also differed in three important ways from previous work.

The first major finding (Figure 4) reveals that Chinese-learning infants perceived tones asymmetrically, looking differently between the test trial types when familiarized with Tone 25, but not when familiarized with Tone 33. Perceptual asymmetries are commonly reported in vowel perception, where infants (and adults) often have difficulty perceiving Vowel A in the context of Vowel B relative to when Vowel B is perceived in the context of Vowel A (Cowan & Morse, 1986; Kuhl, 1991; Kuhl et al., 1992; Polka & Bohn, 1996, 2003, 2011; Polka & Werker, 1994; Repp, Healy, & Crowder, 1979). One account of these asymmetries appeals to *perceptual magnets*. If Vowel B is more prototypical or frequent, then contexts which have many tokens of Vowel B "capture" less prototypical instances of Vowel A, making them less discriminable (Kuhl, 1991; Kuhl et al., 1992). In our tone example, Tone 25 is more likely the prototypical or salient tone, since Tone 33 is non-native for half of our sample (i.e., Mandarin-learners). Moreover, young infants pay special attention to f0 contours in infant-directed speech (Cooper, 1997; Fernald, 1985; Fernald & Kuhl, 1987; M. Papoušek, H. Papoušek, & Symmes, 1991), and the proportion of Mandarin rising tones increases in samples of infant-directed speech relative to adult-directed speech (Papoušek & Hwang, 1991). However, since our results show that Chinese-learning infants were *better* at tone discrimination when familiarized with Tone 25, we observed an effect in the opposite direction of a perceptual magnet.

An alternative account suggests that vowel contrasts are more discriminable when Vowel A is more "extreme" (i.e., found further towards the "periphery" of acoustic/articulatory space) than Vowel B (Polka & Bohn, 1996; 2003; 2011). While it is unknown why perceptual asymmetries run this way, one possibility is that extreme vowels have spectral properties that make them particularly salient, and infants thus take these vowels to be "referent vowels" when building a perceptual representation of acoustic space (Polka & Bohn, 2011). A related possibility is that peripheral vowels may help to anchor perceiver's normalization of articulatory/acoustic space, thus improving the discrimination of vowels in this context (Polka & Bohn, 2003; see also Neary, 1977). Indeed, the study of tone perception in adults has also emphasized the importance of f0 normalization (Huang & Holt, 2009; Leather, 1983; Moore & Jongman, 1997), particularly for level tones, which can be perceived as either high or low depending on the preceding f0 context (Francis et al., 2006; Wong & Diehl, 2003). For Chinese-learning infants in the present experiment, the familiarization provided by Tone 33 (a series of relatively invariable f0 tracks) may have made it more difficult for infants to achieve normalization, since level tones provide comparatively less information about f0 range in the local acoustic context than contour tones.

Future work will likely establish that there are many more factors contributing to asymmetries in tone perception than simple f0 normalization. For example, Tsao (2008) reported that 10- to 12-month-old Mandarin learners noticed when a dipping tone (Mandarin tone 214) was embedded in a train of high-level tones (Mandarin tone 55), but not vice versa, which is not the prediction made by an f0 normalization hypothesis. Additionally, Francis and Ciocca (2003) reported an asymmetry in tone perception such that high-low pairs of level tones were easier to discriminate than acoustically equivalent low-high pairs for Cantonese speakers, but not for English speakers. This suggests that some aspects of perception contributing to asymmetrical patterns are language-specific, perhaps changing with experience. Indeed, one aspect of these results supports this idea that the asymmetry in our results develops over age. As illustrated in Appendix A, Chinese-learning 4-month-olds did show a small familiarization effect (i.e., larger mean differences between test trial types in the group familiarized with Tone 25 compared to Tone 33), but the corresponding familiarization effect in 9-month-olds was far larger. Statistical analysis of these differences was not significant, but future research using experiments specifically designed to test this possibility may show that asymmetries in tone perception can result from increasing exposure to a tone language. Indeed, this would be a result echoing previous reports that have examined both the perception of vowels (Polka & Bohn, 2011; Pons, Albareda-Castellot, & Sebastián-Gallés, 2012), and vowel duration (Mugitani et al., 2009).

The second major finding (Figure 5) reveals consistent differences between the perception of native and non-native tones in Mandarin- versus Cantonese-learning infants. The specific patterns reported here can be interpreted in two ways. The first way requires consideration of which tone type(s) can be assimilated as a "native" tone (see So & Best, 2010). For the Mandarin-learners, Tone 25 is very similar to the native rising tone, and adult Mandarin speakers rarely misidentify Tone 25 when perceiving Cantonese tones (Francis et al., 2008). Thus, Mandarin-learning infants may have preferred Tone 25 to Tone 33 because of the relative familiarity of the former, while Cantonese infants may not shown any clear preferences for either, as they were both native. Another possibility is that Mandarin-learning infants differ from Cantonese infants in their overall preferences for contour versus level tones more generally. For example, Francis et al. (2008) suggested adult Mandarin speakers more heavily weight f0 direction when identifying tones (i.e., whether, and in which direction f0 is changing) compared to f0 height (i.e., the part of the f0 range where a particular tone is instantiated). Adult Cantonese speakers, on the other hand, must weight both kinds of information, since their tone inventory requires both dimensions to identify native categories. Thus, Mandarin-learning infants may similarly prefer Tone 25 to Tone 33 because infants are trying to identify f0 direction, while Cantonese-learning infants would find it equivalently effortful to identify both tones.

Future work may be able to tease apart these two hypotheses by testing the perception of Cantonese rising tones (25) against high-level tones (55) in both groups of Chinese-learning infants. Both are easily assimilated to an existing Mandarin tone, and if Mandarin-learning infants' preferences are driven simply by the native versus non-native status of specific tones, we would expect their looking to be similar to that of Cantonese infants. If looking patterns were instead driven by acoustic characteristics of level versus contour tones, we would expect to replicate the presently observed pattern.

The third major finding (Figure 6) suggests that language-specific influences act on tone perception as early as 4 months of age. This challenges the notion that tone perception is relatively "language-universal" at this age (Mattock et al., 2008). The specific looking patterns in Figure 6 may further reflect two general kinds of perceptual strategies. The first possibility is related to f0 patterns commonly occurring in the native language. Recall that for English-learning infants, situations where every syllable has rising f0 contours may seem strange given that f0 variation in English typically occurs over much larger prosodic units. Thus Tone 33 trials may have been relatively more familiar/native-like than Tone 25 trials. Similarly, Mandarin-learning infants may have preferred Tone 25 trials to Tone 33 trials because the former trial type contains an easily assimilated tone, and would thus be more familiar/native-like. Finally, Cantonese infants may have shown no preference for either Tone 25 or Tone 33 trials, precisely because both tone types are common in the native language.

The second possibility is that looking patterns reflect language-specific attentional weighting for various acoustic dimensions of f0: English speakers weight f0 height most heavily, Mandarin speakers give the most weight to f0 contour, and Cantonese speakers give more equivalent weighting both kinds of information. Hence English-learners in the present study would prefer Tone 33 to Tone 25 because they find identification of f0 height easier for the level tone (i.e., an unchanging, steady f0 height); Mandarin-learners would prefer Tone 25 to Tone 33 because they find identification of f0 direction easier for the contour tone (i.e., an obvious and marked f0 rise); while Cantonese-learning infants would show no difference between Tone 25 and Tone 33 because they are trying to extract both f0 height and direction (i.e., find this task equivalently difficult for both types).

In summary, several new patterns of developmental tone perception are reported here. Yet it seems unlikely that *all* developmental patterns in tone perception will echo the results presented here, likely mirroring the diversity of developmental trajectories described in the consonant and vowel literature. For example, adult Mandarin speakers correctly classifying the presently used Tone 25 - Tone 33 contrast in Cantonese (Francis et al., 2008), just as we report consistent discrimination of this non-native tone contrast by Mandarin-learning infants across both ages tested. Future studies may well find classical *declines* in discrimination for other tone contrasts, especially if two acoustically more *similar* Cantonese tones are tested, or if the tested tones are assimilated to the same category in Mandarin. The low rising (23) and high rising (25) tones in Cantonese constitute one pair that meets these criteria, as it is contrast quite difficult for adult Mandarin speakers to classify (Francis et al., 2008). Future research will need to examine a wider array of tone contrasts to map out how both acoustic and language-specific factors interact.

## General Discussion

Research on infant speech perception has long focused on the development of consonants and vowels with a relative dearth of studies on other lexically relevant prosodic cues (i.e., tone, stress, pitch accent, and phoneme duration). The current study investigated tone perception in infancy, identifying two main questions to which we return below.

**How do infants begin learning native versus non-native tone systems?**

Previous literature on infant tone perception had only compared non-tone language learners (i.e., English-learners) against tone language learners (i.e., a mixed group of Chinese-learners). Thus, it was unclear whether infants are initially sensitive only to the fact that their language contains tone (and then learn the particulars of their native tone system later), or whether there are differences between different types of tone systems from the very earliest stages of perceptual development. Our results disconfirm the first notion. Rather, it appears that Cantonese- and Mandarin-learning infants both show distinct, language-specific perceptual patterns from at least 4 months of age, which appear to carry over into adulthood (e.g., Francis et al., 2008; So & Best, 2010; Lee et al., 1996). This suggests that language-specific effects of hearing a tone language manifest from the beginning in ways that are characteristic of the native language, and not simply characteristic of hearing just any tone system.

Note again the important distinction between discrimination versus preference. Chinese infants in Experiment 2 were able to discriminate the present tone contrast as a group, no matter whether it was native or non-native for them. Conversely English-learning infants in Experiment 1 no longer discriminated the same tone contrast by 9 months of age. This does show some degree of perceptual flexibility in the Chinese group, as hearing some tones, even non-native ones, can help maintain tone discrimination ability throughout infancy. This replicates the results from Mattock and Burnham (2006), who showed that Chinese-learners could still discriminate a non-Chinese tone contrast that was acoustically similar to the presently used one (i.e., involving a level-contour contrast). Even so, this result does not imply that tone discrimination is uniformly easy across all possible contrasts. Recall that 10 to 12-month-old Mandarin-learning infants show variable discrimination abilities even for different sets of native tones (Tsao, 2008). This underscores the notion that *both* acoustic distinctiveness *and* language-specific influences likely interact in determining an infant's perceptual sensitivity for a particular phonetic contrast (e.g., Narayan et al., 2010).

**Is the trajectory of perceptual development similar for tones, vowels, and consonants?**

Previous research has suggested an earlier effect of native language input on the perception of vowels (i.e., at least from 6 months of age) than on the perception of consonants (i.e., at least from 8.5 months of age) (Anderson et al., 2003; Kuhl et al., 1992; Polka & Werker, 1994). According to the *periodicity bias* hypothesis, these developmental trajectories for vowels versus consonants are determined by acoustic factors (Cutler & Mehler, 1993). As infants are better able to attend to larger periodic units at earlier points in developmental time, according to this hypothesis, different types of language input have their effects at different ages. Specifically, the longer, more sustained, or *periodic* a particular unit in the speech signal is, the more salient this unit is for young infants. Evidence for Cutler and Mehler's hypothesis comes from studies showing that infants have very early language-specific preferences for speech melody at the level of clauses or phrases (see Nazzi & Ramus, 2003 for review), showing language-specific stress perception beginning from at least 4 months of age (Friederich et al., 2009), and then finally begin to show language-specific vowel and consonant perception from 6 and 8.5 months of age, respectively. At segmental timescales, more periodic units (e.g., vowels) are thus considered privileged over the less periodic ones (e.g., consonants), as the latter require finer temporal resolution to be identified.

An important element missing from the original articulation of the periodicity bias is the role of tones, which operate at roughly the same timescales as consonants and vowels, and are just as periodic as vowels. Contrary to the predictions from the periodicity hypothesis, our data show that tone perception is language-specific by at least 4 months of age, before either vowels or consonants show similar effects. This suggests a view of development similar to that suggested by Cutler and Mehler (1993), except that perceptual development is not guided simply by how periodic information in the speech signal is, but rather how *salient* this information is (see also Narayan et al., 2010 for a further example of differences in acoustic saliency affecting phonetic reorganization among different kinds of consonants).

Saliency is, of course, difficult to define, especially when considering the varying effects on perception that different acoustic cues may have. It is likely that saliency is a product of several interrelated factors, at least one of which (particularly when considering tones, vowels, and consonants) might simply be amount of experience that infants have processing these cues. For example, those acoustic properties that can reach the foetal auditory system, like f0 and amplitude, are likely to become perceptually salient from early on in development. This may be followed by vowels (for which partial spectral information in lower frequencies may be able to reach the womb), and then by consonants (for which relatively little acoustic information is likely available in the womb). Consider also that both neonates (Nazzi et al., 1998) and 2- to 3-month-olds (Karzon & Nicholas, 1989) robustly perceive both f0 level and direction across two syllables, while the spectral cues needed to identify other kinds of vocalic or consonantal contrasts may be difficult for infants at these ages, particularly if they are instantiated on multi-syllabic words (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Karzon, 1985). Moreover, neuroimaging evidence suggests that f0 changes in the speech signal may be processed differently than spectral information in vowels, becoming specialized to a specific hemisphere early in life due to differences in the relative physical complexity between f0 and vowel formant cues (see Minagawa-Kawai, Cristià, & Dupoux, 2011 for review).

The psychophysical properties of acoustic cues are not the only properties that likely contribute to perceptual salience. Phonetic contrasts are almost always cued by several acoustic properties, and the degree to which these cues are in concordance can often affect both infants' (e.g., Eimas, 1985; Sato et al., 2012) and adults' (see McMurray & Jongman, 2011 for review) abilities to perceive phonetic contrasts. Future work will need to ask how a system of co-varying cues to a particular phonetic contrast contributes to perceptual salience in development. Consider, for example, the interesting case of phoneme duration. Even though the duration of constriction distinguishing two consonant geminates (e.g., /pata/ versus /patta/, or /pada/ versus /padda/) are sometimes *longer* than the equivalent amount of constriction distinguishing two voicing categories (e.g., /pada/ and /pata/), voicing may be still be thought of as *more* salient for several possible reasons. One reason may be because categorizing phoneme duration depends first on the successful identification of the vocalic or consonantal pattern (i.e., as /d/ or /t/), and then a second step to categorize the duration of constriction as either short or long. Another reason may be because voicing distinctions can have as many as 16 co-varying acoustic cues (e.g., Lisker, 1986), while (consonant) duration distinctions have as little as 4 such cues (Idemaru & Guion, 2008; Kawahara, 2006), and the presence of extra cues may increase the perceptual salience of voicing. Such properties may explain why phoneme duration contrasts are difficult for infants to detect until at least 9.5 months of age (Sato et al., 2012, 2010), and why cross-linguistic differences in the perception of (vowel) duration has not been reported until at least 18 months of age (Mugitani et al., 2009).

One additional factor that may contribute to differences in the timing of language-specific perception for tones, vowels, consonants, and phoneme duration is the frequency with which phonetic contrasts of the critical type occur in speech input (e.g., Anderson et al., 2003; Narayan et al., 2010; Polka et al., 2001; Sato et al., 2012). Indeed, not just frequency, but specific *distributional* characteristics of phonetic information in a mother's speech can be predictive of her infants' discrimination abilities (Cristia, 2011; see Werker, Yeung, & Yoshida, 2012 review). More work will need to establish how the statistical properties of f0 variation manifest in tonal languages (e.g., Gauthier, Shi, & Xu, 2007a, 2007b), and how these input characteristics are related to the precocious emergence of language-specific tone perception.

**Conclusions**

Speech perception develops remarkably quickly in infancy, as infants become attuned to the properties of the native language within a very short amount of time. The present study contributes to our understanding of this process in two main ways. First, this study shows, in more detail, how the perception of lexical tone develops in infancy. We replicated previous reports showing language-specific differences infant tone perception (Mattock & Burnham, 2006; Mattock et al., 2008), and further showed that infants hearing native versus non-native tone systems perceived tones in language-specific ways from at least 4 months of age. Additionally, perceptual asymmetries are reported among Chinese-learning infants, which may be related to their abilities to normalize some of the acoustic variability of f0 within a speaker. This finding is relevant to a broader discussion in the literature about the character f0 variability in infant-directed speech, and how that variability affects early word-learning (Bortfeld & Morgan, 2010; Singh et al., 2004, 2008).

Second, this study shows that language-specific input reorganizes phonetic perception at different points in development for different kinds of speech units. Here we show here that language-specific perceptual patterns for tones are evident as early as 4 months of age, before either vowel or consonant perception is affected by language input. Future research must determine what precise acoustic factors determine these differences in the timing of language-specific perceptual development: first for “easier” prosodic cues (i.e., stress/pitch accent/tones), then vowels, then consonants, and finally for “harder” prosodic cues (i.e., phoneme duration). Future work may also explore the curious similarity between this hierarchy of development in *perception* with a seemingly parallel hierarchy of development in *production*, where infants first learn to modulate f0, then produce vocalizations with some characteristics of vowels, and finally begin producing consonant-like structures in babbling (Oller, 1980; Stark, 1980).

Any proposed hierarchy must be qualified, however, by the inherent complexity of perceptual behaviour. Indeed, the respective salience of stress/pitch accent/tones, vowels, consonants, and phoneme duration may not reflect performance in all experimental tasks. As has been shown with adults (e.g., New, Araújo, & Nazzi, 2008), as well as with infants and children (e.g., Havy & Nazzi, 2009; Mani & Plunkett, 2007; Nazzi, 2005), vowels are sometimes *more* confusable than consonants when accessing word forms. Additionally, an increasing number of studies have shown that vowels and consonants serve different kinds of functions in other kinds of language processing (Bonatti, Peña, Nespor, & Mehler, 2005; Nespor, Peña, & Mehler, 2003), and that even infants are sensitive to these differences (Pons & Toro, 2010). This serves as a potent reminder that the emergence of language-specific perception in infancy likely involves a complex interaction between the acoustic factors that denote different kinds of speech units, and the functional role that these units have in language processing.

**Acknowledgements**

This research was supported by a grant from the National Sciences and Engineering Research Council of Canada (81103) to J. F. W., a grant from the McDonnell Foundation (412783-001G) to Richard N. Aslin and J. F. W., and a fellowship from the Fondation Fyssen to

H. H. Y. We thank Valter Ciocca for his helpful knowledge and advice on tone contrasts and tone perception, as well as on stimuli construction. We also thank Jacqueline Chong for help preparing stimuli, and recruiting subjects and Lawrence Chen for help collecting and analyzing data, as well as for valuable discussion. Thanks also to several anonymous reviewers, Laurel Fais, Nurit Gurel, and Tianlin Wang for helpful comments and discussion.

## Notes

[1] As is commonly the case in cross-linguistic studies, this consonant is slightly different than the native English post-alveolar affricate /t͡ʃʰi/ (i.e., "chee"). However, adult English speakers easily assimilate the Cantonese consonant to that category, as has already happened diachronically in some Cantonese registers/dialects. Moreover, infants were tested at an age likely before this consonant category becomes strongly language-specific.

[2] An exception to this generalization is that when lists of items are read aloud in English, each item tends to have a rising f0 contour. Even then, however, the list is terminated by a falling contour, and would still make Tone 25 trials relatively more anomalous. We thank a reviewer for calling our attention to this observation.

[3] Gender was left out as a factor because it was impossible to balance across conditions in our hard-to-recruit sample of Chinese infants. Moreover, there are no previously reported gender effects in any previous study on tone perception, nor were there any effects from Experiment 1.

**References**

Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, *46*(2-3), 155–182.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Boston: Mouton de Gruyter.

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P. W., Kennedy, L. J., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, *117*(1), 21–33. doi:10.1037/0096-3445.117.1.21

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development*, *21*, 295.

Best, C. T., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and speech*, *46*(Pt 2-3), 183–216.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development*, *18*(3), 339–350. doi:10.1016/0163-6383(95)90022-5

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360.

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, *18*(1), 37–49.

Bock, K., & Mazella, J. (1983). Intonational marking of a given and new information: Some consequences for comprehension. *Memory & Cognition*, *11*(1), 64–76.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: the role of consonants and vowels in continuous speech processing. *Psychological science*, *16*(6), 451–9. doi:10.1111/j.0956-7976.2005.01556.x

Bortfeld, H., & Morgan, J. L. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology*, *60*(4), 241–66. doi:10.1016/j.cogpsych.2010.01.002

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *46*(2-3), 217–243. doi:10.1177/00238309030460020801

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, *39*(4), 585–594. doi:10.1016/j.wocn.2011.06.002

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience*, *25*(3-4), 195–210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research*, *1128*(1), 148–56. doi:10.1016/j.brainres.2006.10.064

Chao, Y. R. (1968). *A Grammar of spoken Chinese: Zhong guo hua de wen fa*. Berkeley, CA: Univ of California Press.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). Habit 2000: A new program for testing infant perception and cognition. *Computer software]. Austin, the University of Texas*.

Cooper, R. P. (1997). The Development of Infants ' Preference for Motherese. *Infant Behavior and Development*, *20*(4), 477–488.

Cowan, N., & Morse, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. *The Journal of the Acoustical Society of America*, *79*(2), 500–7.

Cristia, A. (2011). Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category. *Journal of the Acoustical Society of America*, *129*(5), 3271–3280. doi:10.1121/1.3562562

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Attention, Perception, & Psychophysics*, *59*(2), 165–179. doi:10.3758/BF03211886

Da, J. (2004). Chinese text computing. Retrieved July 26, 2012, from http://lingua.mtsu.edu/chinese-computing

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314. doi:10.1016/S0749-596X(02)00001-3

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(41), 16027–16031. doi:10.1073/pnas.0705270104

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, *225*(36), 202–225.

Eimas, P. D. (1985). The equivalence of cues in the perception of speech by infants. *Infant Behavior and Development*, *8*(2), 125–138. doi:10.1016/S0163-6383(85)80001-1

Fennell, C. T., & Waxman, S. R. (2010). What Paradox? Referential Cues Allow for Infant Use of Phonetic Detail in Word Learning. *Child Development*, *81*(5), 1376–1383. doi:10.1111/j.1467-8624.2010.01479.x

Fernald, A. (1985). Four-Month-Old Infants Prefer to Listen to Motherese ”. *Infant Behavior and Development*, *8*(2), 181–195.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, *60*(6), 1497–1510.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. doi:10.1016/0163-6383(87)90017-8

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Francis, A. L., & Ciocca, V. (2003). Stimulus presentation order and the perception of lexical tones in Cantonese. *The Journal of the Acoustical Society of America*, *114*(3), 1611. doi:10.1121/1.1603231

Francis, A. L., Ciocca, V., Wong, N. K.-Y., Leung, W. H.-Y., & Chu, P. C.-Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America*, *119*(3), 1712. doi:10.1121/1.2149768

Friederici, A. D., Friedrich, M., & Christophe, A. (2007). Brain responses in 4-month-old infants are already language specific. *Current Biology*, *17*(14), 1208–11. doi:10.1016/j.cub.2007.06.011

Gandour, J. T. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, *9*(1), 20–36.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*(2), 149–175.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, *21*(1), 1–33. doi:10.1177/002383097802100101

Gauthier, B., Shi, R., & Xu, Y. (2007a). Learning phonetic categories by tracking movements. *Cognition*, *103*(1), 80–106. doi:10.1016/j.cognition.2006.03.002

Gauthier, B., Shi, R., & Xu, Y. (2007b). Simulating the acquisition of lexical tones from continuous dynamic input. *Journal of the Acoustical Society of America*, *121*(5), EL190–EL195. doi:10.1121/1.2716160

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, *25*(2), 207–231. doi:10.1006/jpho.1997.0042

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge, UK: Cambridge University Press.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421. doi:10.1016/S0095-4470(03)00016-0

Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua*, *110*(8), 581–616. doi:10.1016/S0024-3841(00)00003-6

Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, *14*(4), 439–456. doi:10.1080/15250000902996532

Huang, J., & Holt, L. L. (2009). General perceptual contributions to lexical tone normalization. *Journal of the Acoustical Society of America*, *125*(6), 3983–3994. doi:10.1121/1.3125342

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior & Development*, *32*(3), 262–274. doi:10.1016/j.infbeh.2009.03.004

Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association*, *38*(02), 167–186.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842–54. doi:10.1121/1.2783198

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, *64*(3), 675–687. doi:10.1111/j.1467-8624.1993.tb02935.x

Kaan, E., Barkley, C., Bao, M., & Wayland, R. (2008). Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. *BMC neuroscience*, *9*(1), 53. doi:10.1186/1471-2202-9-53

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: an event-related potential study. *Brain research*, *1148*, 113–22. doi:10.1016/j.brainres.2007.02.019

Karzon, R. G. (1985). Discrimination of polysyllabic sequences by one- to four-month-old infants. *Journal of Experimental Child Psychology*, *39*(2), 326–342. doi:10.1016/0022-0965(85)90044-X

Karzon, R. G., & Nicholas, J. G. (1989). Syllabic pitch perception in 2- to 3-month-old infants. *Attention, Perception, & Psychophysics*, *45*(1), 10–4.

Kawahara, S. (2006). A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese. *Language*, *82*(3), 536–574.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*(1), 104–117. doi:10.1016/j.wocn.2005.10.003

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Attention, Perception, & Psychophysics*, *50*(2), 93–107.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21.

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*(5044), 606–608.

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics*, *11*(4), 373–382.

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527–542. doi:10.1007/BF01758181

Lisker, L. (1986). “Voicing” in English: A Catalogue of Acoustic Features Signaling /b/ Versus /p/ in Trochees. *Language and Speech*, *29*(1), 3–11. doi:10.1177/002383098602900102

Lively, S. E. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242–1255. doi:10.1121/1.408177

Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences*, *103*(51), 19558 –19563. doi:10.1073/pnas.0607065104

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272. doi:10.1016/j.jml.2007.03.005

Mattock, K., & Burnham, D. (2006). Chinese and English Infants ' Tone Perception : Evidence for Perceptual Reorganization. *Infancy*, *10*(3), 241–265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367–1381.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*(2), 219–46. doi:10.1037/a0022325

Minagawa-Kawai, Y., Cristia, A., & Dupoux, E. (2011). Cerebral lateralization and early speech acquisition: A developmental scenario. *Developmental Cognitive Neuroscience*, *1*(3), 217–232. doi:10.1016/j.dcn.2011.03.005

Minagawa-Kawai, Y., Mori, K., Naoi, N., & Kojima, S. (2007). Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. *Journal of Neuroscience*, *27*(2), 315–321. doi:10.1523/JNEUROSCI.1984-06.2007

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America*, *102*(3), 1864–1877. doi:10.1121/1.420092

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J. F., & Amano, S. (2009). Perception of vowel length by Japanese- and English-learning infants. *Developmental Psychology*, *45*(1), 236–247. doi:10.1037/a0014043

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407–420. doi:10.1111/j.1467-7687.2009.00898.x

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, *98*(1), 13–30. doi:10.1016/j.cognition.2004.10.005

Nazzi, T., Floccia, C., & Bertoncini, J. (1998). Discrimination of pitch contours by neonates. *Infant Behavior & Development*, *21*(4), 779–784. doi:10.1016/S0163-6383(98)90044-3

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*(1), 233–243. doi:10.1016/S0167-6393(02)00106-1

Neary, T. (1977). Phonetic feature for vowel systems. *Unpublished PhD Dissertation. University of Connecticut*.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, *2*(2), 203–230.

New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. *Psychological science*, *19*(12), 1223–1227. doi:10.1111/j.1467-9280.2008.02228.x

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komishan, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology, Vol. 1: Production* (pp. 92–112). New York: Academic Press.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9–B17. doi:10.1016/S0010-0277(97)00030-9

Papoušek, M., Bornstein, M. H., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant Responses to Prototypical Melodic Contours in Parental Speech. *Infant Behavior and Development*, *13*, 539–545.

Papoušek, M., & Hwang, S.-F. C. (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*, *12*(04), 481–504.

Papoušek, M., Papoušek, H., & Symmes, D. (1991). The meanings of melodies in motherese in tone and stress languages. *Infant Behavior and Development*, *14*(4), 415–440.

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *The Journal of the Acoustical Society of America*, *102*(6), 3742–3753.

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624. doi:10.1016/j.wocn.2010.09.003

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. M. I. T., Dept. of Linguistics and Philosophy.

Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *The Journal of the Acoustical Society of America*, *100*(1), 577–592. doi:10.1121/1.415884

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221–231.

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467–478. doi:10.1016/j.wocn.2010.08.007

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d /–/ð / perception: Evidence for a new developmental pattern. *The Journal of the Acoustical Society of America*, *109*(5), 2190. doi:10.1121/1.1362689

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435.

Pons, F., Albareda-Castellot, B., & Sebastián-Gallés, N. (2012). The Interplay Between Input and Initial Biases: Asymmetries in Vowel Perception During the First Year of Life. *Child Development*, *83*(3), 965–976. doi:10.1111/j.1467-8624.2012.01740.x

Pons, F., & Bosch, L. (2010). Stress Pattern Preference in Spanish-Learning Infants: The Role of Syllable Weight. *Infancy*, *15*(3), 223–245. doi:10.1111/j.1532-7078.2009.00016.x

Pons, F., & Toro, J. M. (2010). Structural generalizations over consonants and vowels in 11-month-old infants. *Cognition*, *116*(3), 361–7. doi:10.1016/j.cognition.2010.05.013

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*(6), 2956–2970.

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*(2), 135–150. doi:10.1016/j.jml.2009.09.003

Quam, C., & Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236–250. doi:10.1111/j.1467-8624.2011.01700.x

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(1), 129–145.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental science*, *8*(2), 162–72. doi:10.1111/j.1467-7687.2005.00403.x

Sato, Y., Kato, M., & Mazuka, R. (2012). Development of single/geminate obstruent discrimination by Japanese infants: early integration of durational and nondurational cues. *Developmental Psychology*, *48*(1), 18–34. doi:10.1037/a0025528

Sato, Y., Sogabe, Y., & Mazuka, R. (2009). Developmen of Hemispheric Specialization for Lexical Pitch-Accent in Japanese Infants. *Journal of Cognitive Neuroscience*, *22*(11), 2503–2513.

Sato, Y., Sogabe, Y., & Mazuka, R. (2010). Discrimination of phonemic vowel length by Japanese infants. *Developmental Psychology*, *46*(1), 106–119. doi:10.1037/a0016718

Singh, L., & Foong, J. (2012). Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition*, *124*(2), 128–142. doi:10.1016/j.cognition.2012.05.008

Singh, L., Morgan, J. L., & Best, C. T. (2002). Infants' Listening Preferences: Baby Talk or Happy Talk? *Infancy*, *3*(3), 365. doi:10.1207/S15327078IN0303_5

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*(2), 173–189. doi:10.1016/j.jml.2004.04.004

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, *4*(2), 157–178.

Skoruppa, K., Cristia, A., Peperkamp, S., & Seidl, A. (2011). English-learning infants' perception of word stress patterns. *The Journal of the Acoustical Society of America*, *130*(1), EL50–5. doi:10.1121/1.3590169

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., Limissuri, R. A., et al. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science*, *12*(6), 914–919. doi:10.1111/j.1467-7687.2009.00835.x

Snedeker, J., & Trueswell, J. (2002). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*, 103–130.

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*(2), 273–293. doi:10.1177/0023830909357156

Spence, M. J., & Moore, D. S. (2003). Categorization of infant-directed speech: development from 4 to 6 months. *Developmental Psychobiology*, *42*(1), 97–109. doi:10.1002/dev.10093

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381–382.

Stark, R. E. (1980). Stages of speech development in the first year of life. In G. H. Yeni-Komishan, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child Phonology, Vol. 1: Production* (Vol. 1, pp. 73–92). New York: Academic Press.

Stern, D. N., Spieker, S., & MacKain, K. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology*, *18*(5), 727–735. doi:10.1037//0012-1649.18.5.727

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of /d-th/ in monolingual and bilingual acquisition of English. *Cognition*, *100*(2), 369–88. doi:10.1016/j.cognition.2005.04.007

Tardif, T., Fletcher, P., Zhang, Z. X., & Liang, W. L. (2008). *The Chinese Communicative Development Inventory (Putonghua and Cantonese versions): Manual, Forms, and Norms.* (Peking Uni.). Beijing.

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

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, *120*(4), 2285–2294. doi:10.1121/1.2338290

Turk, A. E., Jusczyk, P. W., & Gerken, L. (1995). Do English-Learning Infants use Syllable Weight to Determine Stress? *Language and Speech*, *38*(2), 143–158. doi:10.1177/002383099503800202

Vance, T. J. (1976). An experimental investigation of tone and i intonation in Cantonese. *Phonetica*, *33*(5), 368–392.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658. doi:10.1121/1.428217

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese Listeners to Perceive Thai Tones: A Preliminary Report. *Language Learning*, *54*(4), 681–712. doi:10.1111/j.1467-9922.2004.00283.x

Welby, P. (2003). Effects of Pitch Accent Position, Type, and Status on Focus Projection. *Language and Speech*, *46*(1), 53–81. doi:10.1177/00238309030460010401

Werker, J. F., & Gervain, J. (n.d.). Speech perception and phonological development. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology*. Oxford, UK: Oxford University Press.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental psychology*, *24*(5), 672–683.

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*(6), 1866–1878.

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception? *Current Directions in Psychological Science*, *21*(4), 221–226. doi:10.1177/0963721412449459

Wong, P. C.-M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–421. doi:10.1044/1092-4388(2003/034)

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in Chinese revealed by mismatch negativity. *Neuroscience*, *170*(1), 223–231. doi:10.1016/j.neuroscience.2010.06.077

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61–83. doi:10.1006/jpho.1996.0034

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*(1), 55–105. doi:10.1006/jpho.1999.0086

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America*, *120*(2), 1063. doi:10.1121/1.2213572

Yip, M. J. W. (1995). Tone in east Asian languages. In J. H. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 476–494). Cambridge, UK: Blackwell.

Yip, M. J. W. (2002). *Tone* (p. 341). Cambridge, UK: Cambridge University Press.

**Figures**

Figure 1.

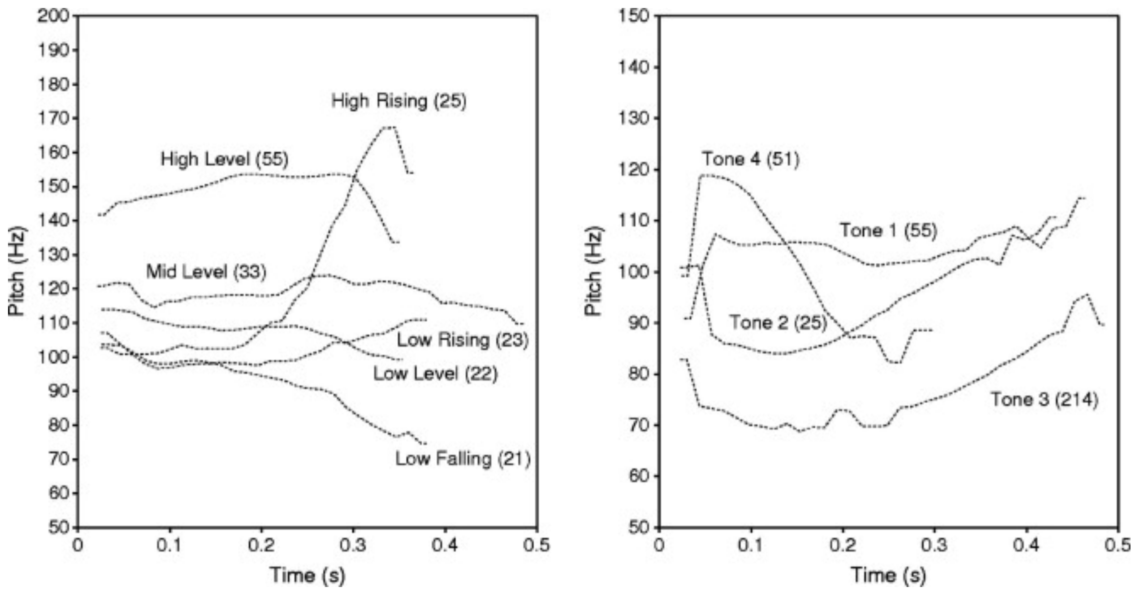

Figure 2.

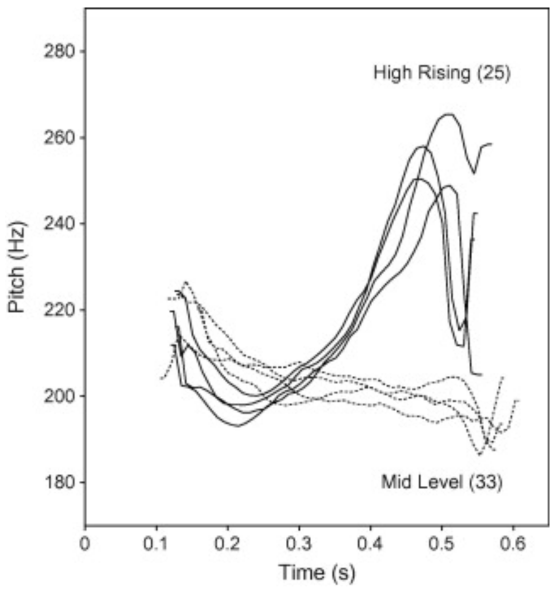
High Rising (25)
Mid Level (33)
Pitch (Hz)
Time (s)
280
260
240
220
200
180
0
0.1
0.2
0.3
0.4
0.5
0.6

Figure 3.

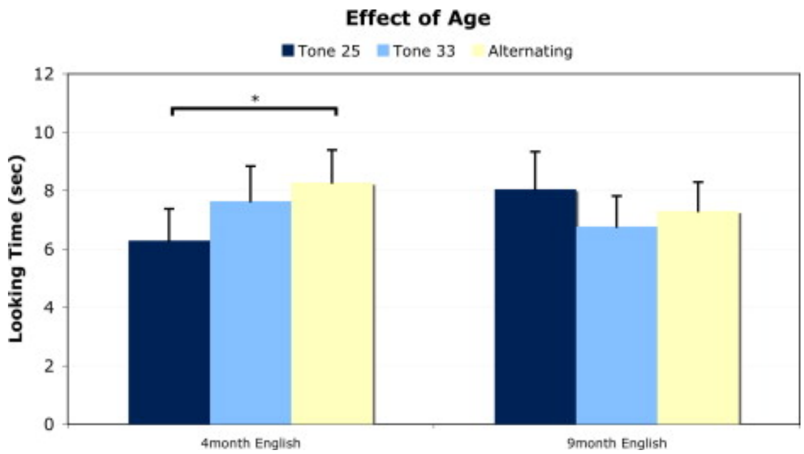
Effect of Age
Tone 25
Tone 33
Alternating
Looking Time (sec)
12
10
8
6
4
2
0
*
4month English
9month English

Figure 4.

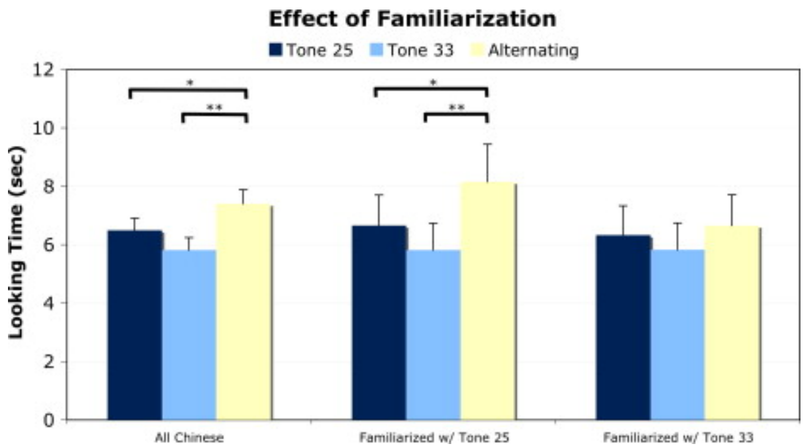
Effect of Familiarization
Tone 25
Tone 33
Alternating
Looking Time (sec)
12
10
8
6
4
2
0
*
**
*
**
All Chinese
Familiarized w/ Tone 25
Familiarized w/ Tone 33

Figure 5.

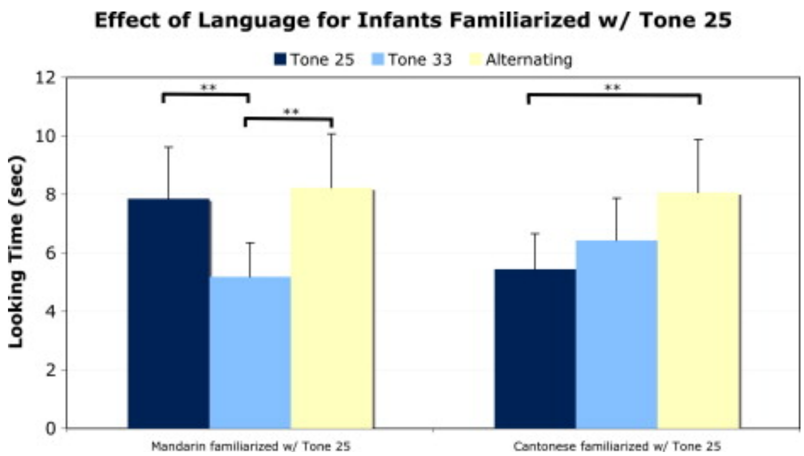
Effect of Language for Infants Familiarized w/ Tone 25
Tone 25
Tone 33
Alternating
Looking Time (sec)
0
2
4
6
8
10
12
**
**
**
Mandarin familiarized w/ Tone 25
Cantonese familiarized w/ Tone 25

Figure 6.

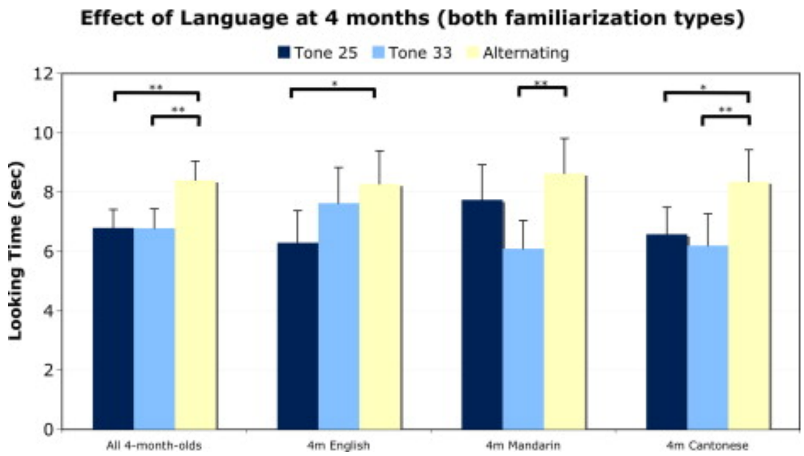
Effect of Language at 4 months (both familiarization types)
Tone 25
Tone 33
Alternating
Looking Time (sec)
0
2
4
6
8
10
12
All 4-month-olds
4m English
4m Mandarin
4m Cantonese

**Figure Captions**

*Figure 1.* F0 contours for tones in Mandarin (left, produced by a male speaker on the syllable [da]) and in Cantonese (right, produced by a male speaker on the syllable [ji]). Note that, because these Cantonese and Mandarin tones come from different speakers, absolute f0 levels cannot be compared across languages, only relative levels are comparable. Images come from a previously published article (Francis, A. L., Ciocca, V., Ma, L., & Fenn, K., 2008), reproduced here with the authors' permission.

*Figure 2.* F0 contours for all tokens used in the present experiments. Solid lines indicate Tone 25 and dashed lines indicate Tone 33. Note that absolute f0 values are higher overall than in Fig. 1 because these stimuli come from a female speaker.

*Figure 3.* Results from Experiment 1 divided by age group. Looking to test trials was equivalent in the 9-month age group, but different in the 4-month age group, indicating tone discrimination in the younger age. Error bars = std. errors, double asterisks = $p < .01$, and single asterisks = $p < .05$.

*Figure 4.* Results from Chinese infants at both ages in Experiment 2, illustrating the effect of familiarization. Looking to test trials differed for these infants overall, but this effect was carried by those in the group familiarized with Tone 25. Pairwise comparisons were conducted in the same manner as other comparisons in the text (although not reported there). Error bars = std. errors, double asterisks = $p < .01$, and single asterisks = $p < .05$.

*Figure 5.* Results from Chinese infants at both ages in Experiment 2, who were familiarized with Tone 25, and divided by language. Each group showed distinct looking patterns overall. Error bars = std. errors, double asterisks = $p < .01$, and single asterisks = $p < .05$.

*Figure 6.* Results from 4-month-olds from both Experiments 1 and 2, collapsing across familiarization group. Each group showed distinct looking patterns overall. Error bars = std. errors, double asterisks = $p < .01$, and single asterisks = $p < .05$.

## Tables

Table 1. Acoustic characteristics of the vowels used in the tone stimuli.

| Position in vowel | Statistic | F0 (Hz) | 1st Formant (Hz) | 2nd Formant (Hz) | 3rd Formant (Hz) |
|---|---|---|---|---|---|
| *Tone 25* | | | | | |
| Initial | *Mean* | 199 | 349 | 2487 | 3039 |
| | *Range* | 194–204 | 337–364 | 2455–2501 | 3012–3081 |
| Middle | *Mean* | 211 | 344 | 2512 | 2982 |
| | *Range* | 210–213 | 331–360 | 2502–2530 | 2958–3002 |
| Final | *Mean* | 253 | 382 | 2545 | 2928 |
| | *Range* | 244–264 | 346–417 | 2516–2568 | 2883–2979 |
| *Tone 33* | | | | | |
| Initial | *Mean* | 209 | 334 | 2497 | 3091 |
| | *Range* | 207–213 | 311–369 | 2468–2542 | 3029–3173 |
| Middle | *Mean* | 202 | 342 | 2582 | 3091 |
| | *Range* | 198–205 | 324–356 | 2555–2609 | 3065–3155 |
| Final | *Mean* | 198 | 352 | 2574 | 3038 |
| | *Range* | 194–204 | 348–355 | 2561–2625 | 2952–3100 |

*Note.* Measurements taken at initial position were from 15% into the vowel, at middle position were from 50% into the vowel, and at final position were from 85% into the vowel. These positions were selected to ensure that measurements at initial position did not reflect formant transitions from the preceding consonant, and that measurements at final position did not reflect variations in voice quality (i.e., creakiness) that commonly occur at the end of a vowel.

Table 2. A sample of trials heard in one experimental order.

| **Trial type** | **List used**[a] | **Tokens heard in trial** | | | |
|---|---|---|---|---|---|
| Familiarization (Tone 25) | Familiarization List | Tone $25_1$ | Tone $25_2$ | Tone $25_3$ | Tone $25_4$ |
| Tone 25 | Non-Alt List A | Tone $25_1$ | Tone $25_2$ | Tone $25_3$ | Tone $25_4$ |
| Alt | Alt List A | Tone $25_1$ | Tone $25_2$ | Tone $33_1$ | Tone $33_2$ |
| Tone 33 | Non-Alt List A | Tone $33_1$ | Tone $33_2$ | Tone $33_3$ | Tone $33_4$ |
| Alt | Alt List B | Tone $25_1$ | Tone $25_2$ | Tone $33_1$ | Tone $33_2$ |
| Tone 25 | Non-Alt List B | Tone $25_1$ | Tone $25_2$ | Tone $25_3$ | Tone $25_4$ |
| Alt | Alt List A | Tone $25_3$ | Tone $25_4$ | Tone $33_3$ | Tone $33_4$ |
| Tone 33 | Non-Alt List B | Tone $33_1$ | Tone $33_2$ | Tone $33_3$ | Tone $33_4$ |
| Alt | Alt List B | Tone $25_3$ | Tone $25_4$ | Tone $33_3$ | Tone $33_4$ |

*Note*. Subscripts indicate tokens #1–#4. Whether infants heard a familiarization phase with Tone 25 or Tone 33, whether the first test trial was an Alt or Non-Alt trial, and whether the first (and third) Non-Alt trial was Tone 25 or Tone 33 were maximally counterbalanced across infants.

[a] Different lists indicate different pseudo-random orders of tokens. In Alt List A, tokens of Tone 25 and Tone 33 alternated for the first 4 tokens beginning with Tone 25, and were then random. In Alt List B, the first four tokens were similarly alternated, but began with Tone 33.