Published in final edited form as:

Cognition. 2012 August ; 124(2): 128–142. doi:10.1016/j.cognition.2012.05.008.

Influences of Lexical Tone and Pitch on Word Recognition in Bilingual Infants

Leher Singh and Joanne Foong
National University of Singapore

Abstract

Infants' abilities to discriminate native and non-native phonemes have been extensively investigated in monolingual learners, demonstrating a transition from language-general to language-specific sensitivities over the first year after birth. However, these studies have mostly been limited to the study of vowels and consonants in monolingual learners. There is relatively little research on other types of phonetic segments, such as lexical tone, even though tone languages are very well represented across languages of the world. The goal of the present study is to investigate how Mandarin Chinese-English bilingual learners contend with non-phonemic pitch variation in English spoken word recognition. This is contrasted with their treatment of phonemic changes in lexical tone in Mandarin spoken word recognition. The experimental design was cross-sectional and three age-groups were sampled (7.5 months, 9 months and 11 months). Results demonstrated limited generalization abilities at 7.5 months, where infants only recognized words in English when matched in pitch and words in Mandarin that were matched in tone. At 9 months, infants recognized words in Mandarin Chinese that matched in tone, but also falsely recognized words that contrasted in tone. At this age, infants also recognized English words whether they were matched or mismatched in pitch. By 11 months, infants correctly recognized pitch-matched and - mismatched words in English but only recognized tonal matches in Mandarin Chinese.

From birth to about six months, infants demonstrate language-general discrimination abilities such that they are sensitive to a wide range of speech contrasts, both native and foreign to their ears (Best, McRoberts, & Sithole, 1988; Lasky, Syrdal-Lasky, & Klein, 1975; Werker, Gilbert, Humphrey, & Tees, 1981). However, over the following months, infants' phonetic sensitivities undergo a process of perceptual reorganization, gradually aligning with the properties of their native language. This results in a widely documented developmental transition when discrimination of most non-native sounds declines and discrimination of native sounds continues to improve (e.g. Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker et al., 1981; Werker & Tees, 1984; but see Best et al., 1988). For English-learning monolingual infants, discrimination of non-native vowel contrasts starts to attenuate from around 6 months of age on average (Kuhl et al., 1992; Polka & Werker, 1994) while for consonantal contrasts, this shift occurs later between 10-12 months of age (Werker & Tees, 1984).

Recent research has revealed a more nuanced developmental profile of phonetic discrimination abilities. Rather than venturing a 'wholesale' decline in sensitivity to non-

© 2012 Elsevier B.V. All rights reserved.

Address for correspondence: Leher Singh, National University of Singapore, Dept of Psychology, 9 Arts Link, Singapore 117570.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

native consonants, it seems the age at which this decline is observed depends on stimulus properties. For example, particular non-native click consonants remain discriminable to infants even at 12 to 14 months of age (Best et al., 1988) perhaps due to their unique acoustic form that sets them apart from more prototypical consonants. Even very young infants have difficulty discriminating consonant pairs that are acoustically similar (e.g. /d-/ versus /ð-/) and are therefore perceptually confusable (Polka, Colantonio, & Sundara, 2001). Furthermore, within a language, the frequency of consonant classes (e.g. coronal versus dorsal consonants) impacts the age at which discrimination of non-native consonants belonging to these classes declines (Anderson, Morgan, & White, 2003). Most recently, Narayan, Werker, and Beddor (2010) have demonstrated that the separability of phonetic contrasts in acoustic space influences the ease with which infants will discriminate those contrasts. In combination, these findings underscore the fact that early discrimination abilities are highly vulnerable to the distributional and acoustic properties of stimuli employed and that infants do not show indiscriminate shifts in perception as they attune to their native language.

Much of the prior research on early phonetic development focuses on particular language typologies and populations. Specifically, the majority of studies in this area have focused almost exclusively on vowels and consonants from Romance and Germanic languages. By contrast, there has been little investigation of the how infants treat sounds of other languages, such as tone languages, in infancy. However, around 60 to 70 % of the world's languages exploit tone contrasts to distinguish meaning (Yip, 2002). Moreover, at least half of the world's population speaks a tone language (Fromkin, 1978) suggesting that phonetic discrimination of sounds that feature quite prominently in both learners and languages remain largely unexplored. A second consideration is that English-learning monolingual participants are represented far in excess of non-English learning infants or bilingual learners in developmental psycholinguistic research. However, monolingual learners are not statistically representative of the majority of learners. Multilingual learners far outnumber monolingual learners (Grosjean, 1982) with an estimated two thirds of children raised in multilingual environments (Crystal, 1997). Moreover, the most frequently spoken language globally is not English but Mandarin Chinese, spoken by almost three times the number of people as English. These lacunae are important to consider given that developmental theories that describe the course of native and non-native phonetic discrimination are not normally confined to a particular language or type of learner, but rather are presumed to apply universally. Therefore, the large body of evidence that has amassed in the area of phonetic structure acquisition has been conducted primarily on languages that represent a typological minority and on learners that constitute a statistical minority. It remains an open question as to whether these findings generalize to other language and learners and furthermore, how the majority of learners progress in this area.

Recently however, there have been a few studies investigating phonetic discrimination in bilingual learners, yielding a somewhat incongruent pattern of results. The first study to investigate this issue explored vowel discrimination in Spanish-Catalan bilinguals, Spanish monolinguals and Catalan monolinguals on a Catalan vowel contrast (that was not contrastive in Spanish). At 4 months, all three groups were able to discriminate the contrast. However, at 8 months, both groups of monolingual infants showed language-specific responses, while bilingual infants failed to discriminate the contrast. This was a surprising finding given that the contrast was phonemic in one of their languages. By 12 months, the ability to discriminate this contrast re-emerged in the bilingual sample (Bosch & Sebastián-Gallés, 2003). Similar evidence of a U-shaped function in discrimination abilities over the first year after birth was found with a different vowel contrast (Sebastián-Gallés & Bosch, 2009), suggesting that the developmental course observed was not specific to a single contrast. These findings point to a distinct course of development for monolinguals and

bilinguals, reflected in a two-stage process for monolinguals and a three-stage process for bilinguals. However, in a recent investigation, Albareda-Castellot, Pons, and Sebastián-Gallés (2011) have reported a comparable developmental trajectory between monolinguals and bilinguals using a more sensitive experimental paradigm that tracks anticipatory eye movements rather than the standard familiarization preference procedure.

Consonant discrimination has also been investigated in bilingual infants. Burns, Yoshida, Hill, and Werker (2007) investigated a bilabial consonant distinction in French and English learning infants that was phonemic in both languages. However, the category boundary between the two consonants was different for each language. Results demonstrated that French-English bilinguals appeared quite similar to monolinguals, discriminating contrasts straddling a category boundary in both languages between 10 and 12 months. Similarly, Sundara, Polka, and Molnar (2008) found a similar pattern of results when French-English bilinguals were tested on a dental-alveolar contrast: Monolingual French 10-12 month olds did not discriminate an English-relevant contrast while bilingual French-English learning infants and monolingual English learners did.

Taken together, these findings point to similarities and distinctions in phonetic discrimination abilities between bilingual and monolingual learners. Investigations of consonant discrimination reveal a comparable developmental profile for monolingual and bilinguals. However, studies using vowel contrasts reveal a different pattern of results, unless highly sensitive experimental methods are employed, suggesting that changes in vowel perception in bilinguals may prove less tractable compared with those observed in consonant perception.

While most studies have focused on consonant and vowel discrimination, there has been much less research conducted on lexical tone in spite of tonal languages being well represented globally. Lexical tone is predominantly made up of variation in the mean fundamental frequency values (f_0) and in the f_0 contour, the primary perceptual correlate of which is pitch (Burnham & Mattock, 2007). Perceptually, tones are identified primarily by their f_0 contour (Liu & Samuel, 2004), although native speakers are able to identify tones using very brief acoustic input (i.e. consonantal segments plus six glottal cycles) above chance levels (Lee, 2009). Secondary determinants of tone include duration, amplitude and voice quality. While tone languages therefore make use of pitch variation to draw lexical distinctions, pitch changes are used to communicate meaning in all the languages in the world, for example, to signal changes in communicative intent or emotional prosody at the level of the intonational phrase. However, for speakers of tonal languages, pitch variation can drive changes in lexical identity as well as prosody. Therefore, for such speakers, tone drives segmental changes (i.e. changes that can alter lexical identity) and are phonemic as well as suprasegmental changes that are non-phonemic. This has led to a dual classification of tone as a phonetic segment as well as a suprasegmental feature depending on whether it is defined in structural or functional terms (Burnham & Mattock, 2007). Even for infants learning non-tonal languages such as English, it is communicatively relevant to integrate certain changes in pitch contours as they signal changes in prosody, a tier of the language code to which English learning infants and adults are highly sensitive (Birch & Clifton, 2002; Hirschberg & Ward, 1992; Seidl & Cristià, 2008).

There have been a few recent studies investigating developmental change in tone perception in young infants. In one such study, Mattock and Burnham (2006) examined 6- and 9-month-old English and Chinese-learning infants' discrimination of Thai lexical tone contrasts. In addition to a set of tone stimuli, non-speech f_0 -equivalent synthetic violin contrasts were included to determine whether any evidence of attenuation in discrimination was specific to language. Using a conditioned headturn paradigm, Mattock and Burnham

demonstrated that English-learning infants showed a decline from 6 to 9 months of age in tone discrimination, whereas Chinese learning infants did not show an age-based decline. Crucially, neither group showed evidence of attenuation of discrimination abilities for non-speech stimuli (i.e. violin contrasts). The authors concluded that tone perception is pruned in response to native language experience and is subject to the kind of attenuation in non-tonal language learners formerly only reported for consonant and vowels.

Mattock, Molnar, Polka, and Burnham (2008) extended these findings by examining whether infants' attenuation for tone begins prior to 6 months of age, given the psychophysical similarity between vowels (for which reduced discrimination of non-native contrasts is observed as early as 6 months) and lexical tones. The authors tested French and English-learning infants, providing interesting points of comparison given that English is a stress-timed language and French a syllable-timed language. This comparison was drawn on the grounds that different types of experience with stress may lead to differences in tone perception. Specifically, the authors raised the possibility that the high degree of pitch variation inherent in a stress-timed language may enhance sensitivity to pitch-driven contrasts more generally, including lexical tone. However, the authors observed a similar pattern of results across French and English-learning infants although there were age-related changes in discrimination evidenced in both groups. Specifically, in both groups, 4- and 6-month-old English and French infants discriminated Thai tone contrasts, but 9-month-olds failed to do so. The results accord with previous research demonstrating a declining sensitivity to tone in the first 9 months after birth independent of native language intonational experience. Investigations of tone discrimination constitute an exception in the expansive field of infant phonetic perception. However, the studies completed to date suggest that the timeline according to which tone-language learners attune to tone overlaps with the timeline according to which non-tone language learners attune to vowels (Kuhl et al., 1992; Mattock & Burnham, 2006). Both types of changes seem to take place between 6 and 9 months for monolingual and bilingual learners, while consonant attunement appears later for both monolingual and bilingual infants (Burns et al., 2007; Werker & Tees, 1984).

The majority of prior research on the emergence of native language influences on infant perception, whether it involves vowels, consonants, or tones, has been confined to discrimination and habituation paradigms which essentially assess sensitivity to sound contrasts. However, developing an awareness of native phonology is presumably primarily useful for the purposes of establishing a well-articulated lexicon (Maye, Werker, & Gerken, 2002), yet there have been very few investigations of when native phonological properties become integrated into emergent word knowledge. The distinction between segment discrimination and word knowledge has proven to be quite important. Specifically, previous studies comparing phonetic discrimination and integration of phonological knowledge into word-based processing tasks have revealed seemingly contradictory findings (e.g. Stager & Werker, 1997; but see Fennell & Waxman (2010)). The demands of processing at a word level can compromise sensitivity to phonemic contrasts, a constraint that seems to endure longer in bilingual children compared with monolingual children (Fennell, Byers-Heinlein, & Werker, 2007; but see Mattock, Polka, Rvachew, & Krehm, 2010). Therefore, the age at which tone-language learning infants selectively represent tone in memories for word-based tasks, in contrast to speech discrimination tasks, remains unclear, yet represents a necessary step towards mastering a tonal language.

Early knowledge of wordforms can be evinced in young infants using a familiarization-recognition paradigm. Using this paradigm, it has been shown that monolingual English-learning infants can segment and recognize words in fluent speech by 8 months, even though wordform knowledge at this stage is not likely referential (Jusczyk & Aslin, 1995). However, their abilities are circumscribed by surface changes, such as those associated with

fundamental frequency shifts across encounters of words. For example, 7.5 month old infants are not able to equate words across encounters that vary in pitch (Singh, White, & Morgan, 2008), talker gender (Houston & Jusczyk, 2000), vocal emotion (Singh, Morgan, & White, 2004) or focusing stress (Bortfeld & Morgan, 2010). By contrast, by 9 months, infants are able to generalize across variations in pitch and by 10.5 months, they are able to generalize across changes in emotion (Singh et al., 2004; Singh et al., 2008). Therefore, by the end of the first year, monolingual English learning infants successfully disregard pitch as a source of non-phonemic variation in their language and in doing so, represent words in accordance with the acoustic-phonemic correspondences of their language.

Learners attempting to master a tonal and non-tonal language face a complex sequence of demands when faced with tone and pitch variation in word recognition paradigms. First, they have to identify the language, as language identity is presumably not explicitly or comprehensibly specified by talkers. Second, if the target language is tonal, they have to distinguish lexical tone from intonational variation when equating repeated encounters of a word, distinguishing words that are tonal contrasts but lexically equating words that are associated with other pitch contrasts. If the target language is non-tonal, learners have to disregard pitch changes when equating repetitions of the same word. These abilities are essential for the formation of language-specific phonological inventories and in turn, are prerequisites for the construction and expansion native lexica. Therefore, establishing phonological constancy for such learners is a complex task requiring language-specific distinctions to be drawn based on the linguistic functions of pitch both within and across languages.

The purpose of this study is to integrate the fields of bilingual speech perception in infancy, early word-form recognition and tone perception in order to explore developmental changes in Mandarin-English bilingual learners' abilities to incorporate phonemic and non-phonemic pitch changes into emergent word knowledge in each of their languages.

Overview of the Current Study

The goal of the current pair of experiments was to investigate when bilingual infants learning Mandarin and English incorporate tone into Mandarin word recognition and disregard pitch variation in English word recognition. Experiment 1 is focused on Mandarin-English bilingual infants' capacities for spoken word recognition in English when words match or contrast in terms of pitch. Experiment 2 asks a similar question with regards to Mandarin, yet target words are matched or mis-matched in lexical tone. Therefore, Experiment 1 addresses the issue of how infants respond to non-phonemic pitch variation in English and Experiment 2 addresses the issue of how the same sample of infants respond to phonemic pitch variation in Mandarin. To chronicle changes that occur from 7.5 to 11 months in each language, separate groups of infants were tested at 7.5 months, 9 months and 11 months.

EXPERIMENT 1

In this study, infants' abilities to recognize words amidst pitch changes were investigated in Mandarin English bilingual infants. All stimuli in this experiment were presented in English. Different samples of infants were tested across three age groups. Participants were familiarized with two words and then tested on recognition of those words. One word was matched in pitch across familiarization and recognition and one word was mis-matched in pitch across these two phases. It was hypothesized that like monolingual English infants, 7.5 month old infants would only recognize words that were matched in pitch across both phases of the experiment. However, by 9 months, it was hypothesized that infants would be able to generalize across pitch changes, recognizing both matched and mis-matched words

as observed in monolingual learners (Singh et al., 2008). This capacity for generalization across pitch contrasts is expected to be sustained at 11 months.

Participants

Forty-eight bilingual Mandarin-English learners (27 females and 21 males) participated in this study. Age groups tested were 7.5 months (mean age: 221 days, range 210-237 days; 7 males and 9 females), 9 months (mean age: 263 days, range 255-274 days; 9 females and 7 males), 11 months (mean age: 332 days, range 325-348 days; 8 females and 8 males). Sixteen participants were tested at each age-group. The experimental design was cross-sectional with no shared participants between age-groups. Data from three infants were not included for inattention and fussiness. Participants were defined as bilingual if they had exposure to Mandarin Chinese and English only. All participants were addressed in both languages from adult family members living in the same residence as the child. For 46 participants, exposure was provided by a parent living in the child's primary residence and for 2 participants, exposure was provided by a live-in grandparent. Bilingualism was defined as having more than 40% exposure to Mandarin Chinese but no more than 70% exposure to Mandarin Chinese. The average amount of exposure was 43% to Mandarin Chinese and 57% to English. Participants all resided in the Greater Boston area and were recruited through advertisements, birth records, and word-of-mouth. All participants were born in the United States and had no significant change in the amount of exposure to either of their languages over the course of their lives. As such, the sample consisted of simultaneous bilinguals.

Thirty-nine of forty-eight infants were tested on the Experiments 1 and 2 in direct succession. Of these infants, 23 infants received the Mandarin task (Experiment 1) first and 16 infants received the English task (Experiment 2) first. The remaining 9 infants were brought back for testing at a later date (within 5 days of initial testing) due to fussiness or inattention after the first session. Two of these infants received the Mandarin Task first and 7 received the English task first.

Stimuli

Stimuli for Experiment 1 consisted of four monosyllabic words (*bike*, *hat*, *tree* and *pear*) and four six-sentence passages recorded by a mother addressing her infant. When recording the stimuli, the speaker was asked to use infant-directed speech for all tokens and to communicate positive affect.

Target words consisted of 15 tokens of four different words (*bike*, *hat*, *tree* and *pear*). During familiarization, infants were trained on these target words. Recognition stimuli consisted of sentences containing the target words and consisted of 24 sentences containing the target words. Each target word appeared in a 6-sentence passage, once per sentence. Within each passage, the target word appeared twice each in initial, medial and final sentence positions (see Appendix A for all stimuli). All stimuli were modified to create two new stimulus sets, each involving a uniform transformation of the fundamental frequency. One set of stimuli (henceforth the High Pitch set) was created by raising the fundamental frequency of all words and passages by $\frac{1}{4}$ octave (3 semitones). This was done by applying a uniform translation of all pitch points up by $\frac{1}{4}$ octave. A second set of stimuli (henceforth the Low Pitch set) was created by decreasing the fundamental frequency of all words by the same amount ($\frac{1}{4}$ octave). Therefore, the difference between the two sets of stimuli was half an octave. Both sets of stimuli involved pitch manipulations, so that infants' preferences would not be affected by the naturalness of the stimuli. All acoustic transformations and measurements were completed using a PRAAT script (Boersma & Weenink, 1996).

Amplitude and duration measures were identical across the two stimulus sets. Mean pitch values and standard deviations for all stimuli are displayed in Table 1.

During familiarization, infants heard citation form tokens of two words. Half of the infants heard the words *bike* and *hat* while the other half heard *tree* and *pear*. For each infant, one word was heard in High Pitch and the other in Low Pitch. Half of the infants then heard all recognition passages in a High Pitch and half the infants heard all passages in a Low Pitch. During recognition testing, infants heard 6 sentences that made up a series of passages containing all four words. Listening times to passages containing familiarized words were divided into listening times to passages where familiarization-recognition stimuli were matched in pitch, those were stimuli were mismatched in pitch, and both results were compared with listening times to passages containing novel (unfamiliar) words.

Procedure

Infants were tested using the Headturn Preference Procedure (HPP) (Kemler Nelson et al., 1995). The infant was seated on the parent's lap facing the center light. The parent listened to instrumental music over headphones to mask the stimuli. Each trial began with the center light flashing until the experimenter judged that the infant fixated on the flashing light. At that point, this light was turned off, and one of the side lights began to flash to attract the infant's attention to the side. Side of presentation was randomized across trials, so that all stimuli occurred on both sides. After the infant turned to look at the flashing side light, the speech stimuli for that trial began to play. The sound continued to play and the side light remained on for the duration of the infant's fixation on the light. Each trial continued until the infant looked away for two seconds, or until 20 seconds of looking time had been accumulated during that trial. If the infant looked away, but then looked back within two seconds, the trial continued. If the infant's looking time was below 2 seconds, the trial was repeated with a new randomization of the trial stimuli; otherwise, the procedure advanced to the next trial.

Several measures were taken to avoid bias. The experimenter was located in an observation room which constituted a sound-proofed acoustic chamber. The infant and parent were situated in an adjacent sound-proofed acoustic chamber. Sound from the infant booth was not transmitted to the observation room. However, the experimenter wore a pair of Bose noise-cancelling headphones as an extra precaution against bias. The experimenter could see the infant and parent through a CCTV system in order to observe the infants' direction of gaze in order to start and end trials. The computer program used to administer the experiment did not display the stimuli or trial type but rather the trial number and the phase of experiment (training or test).

Familiarization began with trials alternating between the two target words. Once the infant had exceeded 30 seconds of looking time with one word, all subsequent familiarization trials presented the alternate word. This modification of the HPP was instituted to ensure that differences in looking times during recognition testing could not be due to different amounts of familiarization with the two target words. When the infant reached 30 seconds of looking time with the second word, the test phase began. The words used as familiarization stimuli and the assignment of amplitude to words were counterbalanced across subjects. As a result of this design, across subjects each item served every possible role (matched familiarization word/mismatched familiarization word/unfamiliar word).

Recognition testing consisted of four blocks of trials, each block containing one trial with each of the four passages. The order of passages within each block was randomized for each infant. In addition, the order of sentences within passages was also randomized on each trial. The test procedure was similar to the familiarization procedure, except that the side light

continued to flash while infants were fixated on the light. As in the familiarization phase, if the infant continued to look at the light for 20 seconds, the trial ended automatically and the next trial began. Similarly, if the infant failed to look at the side light for at least 2 seconds, the trial was automatically repeated. A minimum criterion of 2 seconds was necessary to allow the infant to hear at least one token of the target word in a sentence.

Results

In this experiment, infants were exposed to two words. One word was presented at a pitch that matched the recognition stimuli and one word was presented at a pitch that did not match the recognition stimuli. Therefore, there was one dependent variable of interest: looking times in association with recognition stimuli. The independent variable (sentence type) had three levels: looking times to the sentences containing the pitch-matched word, looking times to the sentences containing the pitch-mismatched word, and looking times to the sentences containing unfamiliar words (averaged across both types of unfamiliar words). At 7.5 months, infants listened for an average of 9690 msec (SD: 2027) to matched words, for an average of 7371 msec (SD: 3068) to mismatched words and for an average of 7861 msec (SD: 2876) to unfamiliar words. At 9 months, infants listened for an average of 9032 msec (SD: 2795) to matched words, for an average of 8335 msec (SD: 2091) to mismatched words and for an average of 6737 msec (SD: 2150) to unfamiliar words. At 11 months, infants listened for an average of 10133 msec (SD: 3213) to matched words, for an average of 9503 msec (SD: 3955) to mismatched words and for an average of 7759 msec (SD: 2693) to unfamiliar words. A 3×3 repeated-measures ANOVA was conducted with age (7.5, 9, 11 months) as a between-subjects factor and stimulus type (pitch-matched, pitch-mismatched, unfamiliar words) as a within-subjects factor. Results indicated a main effect of stimulus type, $F(2, 90) = 12.67$, $p < .0001$, no main effect of age, $F(2, 90) = 1.04$, NS and no interaction of stimulus type and age, $F(2, 90) = 1.53$, NS. A test of simple main effects of stimulus type revealed a significant difference in recognition scores for matched items over unfamiliar items, $F(1, 45) = 22.19$, $p < .0001$ as well as between recognition scores for mismatched items over unfamiliar items, $F(1, 45) = 6.2$, $p < .05$.

Post-hoc analyses were computed using Tukey's Honestly Significant Difference (HSD) adopting a significance criterion of $p < .05$. Three pairs of contrasts (matched vs. unfamiliar; mismatched vs. unfamiliar) for each age group (7.5, 9 and 10.5 months) were computed. As word recognition is typically measured by comparing looking times to passages containing familiarized words with those for passages containing unfamiliar words, each type of familiarization item (matched/mismatched) were compared with unfamiliar passages, but not with each other. At 7.5 months, infants showed significantly higher listening times for pitch-matched words compared with those containing unfamiliar words. For mismatched words, infants did not show higher listening times compared with those containing unfamiliar words. At 9 months, infants showed significantly higher listening times for pitch-matched words compared with those containing unfamiliar words. At this age group, infants also showed significantly higher listening times for pitch-mismatched words compared with unfamiliar words (see Figure 1b). At 11 months, infants showed significantly higher listening times for pitch-matched words compared with those containing unfamiliar words. As with the 9 month old sample, 11 month old infants also showed significantly higher listening times for pitch-mismatched words compared with unfamiliar words (see Figure 1c).

The results of Experiment 1 demonstrate consistency in bilingual and monolingual learners' treatment of pitch variation in spoken word recognition, with both groups demonstrating a comparable developmental progression from fragile to generalizable memories for familiarized words. These findings suggest that in this particular domain of language, monolingual and bilingual learners proceed in lock step with no observable time cost borne

by bilingual learners in this specific task. However, pitch variation serves a dual function for this sample of learners, being associated with non-phonemic changes in English and signalling lexical distinctions in Mandarin Chinese. Experiment 2 was designed to investigate how the same sample of infants develops in their capacity to incorporate phonemic pitch variation (i.e. lexical tone) into Mandarin Chinese spoken word recognition.

EXPERIMENT 2

Results from Experiment 1 demonstrate considerable similarity between monolingual English speakers and bilingual Chinese English learners when exposed to a word recognition task in English. Specifically, learners fail to generalize across pitch changes at 7.5 months, but are able to generalize across pitch changes at 9 months. Although monolingual learners have not been tested on this task at 11 months, it is to be expected that the ability to generalize across pitch changes is maintained at 11 months in preparation for early lexical development. Therefore, even though this cohort of learners was mastering another language where pitch variation can drive lexical identity, they were able to disregard pitch changes when listening to English. A direct question to follow from this is how such learners treat tonal changes that distinguish words in Mandarin Chinese. The timeline according to which tone language learners incorporate tone into lexical representations remains uncharted. The results of Experiment 1 motivate our predictions for Experiment 2. In Experiment 1, infants were able to generalize based on non-phonemic pitch variation at 9 and 11 months, but not at 7.5 months. This invites the hypothesis that task performance after 9 months is influenced by acoustic-phonemic correspondences of the test language. Based on this, it is predicted that the current sample of infants would recognize tone-matched words at 7.5 months (but not tone mis-matched words) as observed in the case of pitch, affect, stress and talker gender. This prediction is motivated by infants' over-reliance on surface form at this age independent of its phonemic or communicative relevance, suggesting a focus on prosody as well as phonology to match words. However, at 9 months and 11 months, it is predicted that infants would continue to recognize tone-matched words but not recognize tone-mismatched words in accordance with Mandarin phonology. This prediction is motivated by the emergence of language-specific influences on word recognition observed in Experiment 1 at 9 and 11 months.

Participants

Participants consisted of the same sample of infants as those that participated in Experiment 1. As with Experiment 1, the experimental design was cross-sectional with no shared participants across age-groups.

Stimuli

Mandarin Chinese stimuli were adapted from Tsay and Jusczyk (2003) and are listed in Appendix B. Familiarization stimuli consisted of two words, one spoken in a tone that matched the recognition stimuli and one in a tone that did not match the recognition stimuli. All stimuli were recorded by a native female speaker of Mandarin Chinese and were recorded in infant-directed speech. Fifteen familiarization tokens were selected per syllable-tone combination. Although Mandarin Chinese has four lexical tones, only Tones 1, 2 and 4 were used in the current study. Tone 3, which was not used, is arguably a more complex form (consisting of a falling contour followed by a rising contour) and is therefore characterized by an inflection point. Tone 3 is also highly confusable for Tone 2 for native adult (Shen & Lin, 1991) and child speakers (Li & Thompson, 1977). Perhaps most crucially, Tone 3 is the least stable tone in a continuous speech context as it is most vulnerable to effects of tonal sandhi undergoing a tonal alternation in certain contexts (Wang

& Li, 1967). For these reasons, Tones 1, 2, and 4 were selected for use by Tsay and Jusczyk (2003) and in the present study.

Recognition passages contained target words 'bei' (Tone 1) meaning 'cup', 'tou' (Tone 2) meaning 'head', 'dan' (Tone 4) meaning 'egg', and 'tian' (Tone 1) meaning 'sky'. Recognition passages were the same for all infants. However, there were four within-subject conditions based on familiarization stimuli. For each pair of familiarization words, one word matched in tone with the target word during the recognition phase and one word mismatched in tone. In the first condition, familiarization words were either 'bei' Tone 1 (match) and 'tou' Tone 1 (mismatch). In the second condition, infants were familiarized with 'tou' Tone 2 (match) and 'bei' Tone 2 (mismatch). In a third condition, infants were familiarized with 'tian' Tone 1 (match) and 'dan' Tone 1 (mismatch), or 'tian' Tone 4 (mismatch) and 'dan' Tone 4 (match). The assignment of matched versus mis-matched tones to words was counterbalanced across participants. The experimental design is displayed in Table 2.

Acoustic analyses of tones focused on defining the tonal contour, which proves to be the most influential factor in native tone identification (Gandour, 1984). Tonal contour is optimally characterized by the starting and ending fundamental frequency values for each syllable (Jongman, Wang, Moore, & Sereno, 2006). Tone 1 is associated with a high, level contour, Tone 2 with a rising contour, and Tone 4 with a falling contour. Examples of stimuli for each tone type are displayed in Figures 2a-c. A secondary cue for tone identification is tone duration, with tones 2 and 3 usually being the longest and Tone 4 being the shortest (Blicher, Diehl, & Cohen, 1990). Acoustic analyses for Mandarin stimuli are displayed in Table 3.

In addition to acoustic analyses, ten mothers of infants participants were asked to label familiarization stimuli as High (Tone 1), Rising (Tone 2), Falling-rising (Tone 3), and Falling (Tone 4). As there were a total of 120 familiarization tokens, 5 tokens per syllable-tone combination were randomly selected for this stimulus validation procedure. Mean accuracy rates were 96.75% (SD = 3.55, range = 90-100%).

Procedure

The procedure was identical to that of Experiment 1.

Results

In this experiment, infants were exposed to two words in Mandarin Chinese. One word was presented at lexical tone that matched recognition stimuli and one word was presented at a tone that did not match the recognition stimuli. Again, the dependent variable was looking times to recognition sentences. There were three levels of the independent variable (sentence type): looking times to the sentences containing the tone-matched word, looking times to the sentences containing the tone-mismatched word, and looking times to the sentences containing unfamiliar words (averaged across both types of unfamiliar words). At 7.5 months, infants listened for an average of 12003 msec (SD: 4055) to matched words, for an average of 8749 msec to mismatched words (SD: 3499) and for an average of 8897 msec (SD: 3109) to unfamiliar words. At 9 months, infants listened for an average of 11513 msec (SD: 2704) to matched words, for an average of 10827 msec (SD: 3449) to mismatched words and for an average of 8803 msec (SD: 2451) to unfamiliar words. At 11 months, infants listened for an average of 10637 msec (SD: 4218) to matched words, for an average of 7778 msec (SD: 1438) to mismatched words and for an average of 7492 msec (SD: 2212) to unfamiliar words. A 3×3 repeated measures ANOVA was conducted with age (7.5, 9, 11 months) as a between subjects factor and stimulus type (tone-matched, tone-mismatched, unfamiliar words) as a within-subjects factor. Results indicate a main effect of stimulus type,

$F(2, 90) = 15.86, p < .0001$, no main effect of age, $F(2, 90) = 2.41, p = .10$ and no interaction of stimulus type and age, $F(2, 90) = 1.2, NS$. Tests for simple main effects of stimulus type revealed a significant difference in recognition scores for matched items over unfamiliar items, $F(1, 45) = 26.72, p < .0001$ but no significant differences in recognition scores for mismatched items over unfamiliar items, $F(1, 45) = 2.12, NS$.

Posthoc pairwise comparisons were computed using Tukey's HSD test with a significance criterion of $p < .05$. At 7.5 months, infants showed significantly higher listening times for tone-matched words compared with those containing unfamiliar words. For mismatched words, listening times were not significantly different to those for sentences containing unfamiliar words (see Figure 3a). At 9 months, infants showed significantly higher listening times for tone-matched words compared with those containing unfamiliar words. At this age group, infants also showed significantly higher listening times for tone-mismatched words compared with unfamiliar words (see Figure 3b). At 11 months, infants showed significantly higher listening times for tone-matched words compared with those containing unfamiliar words. At this age, infants did not show a significant difference in listening times for tone-mismatched words compared with sentences containing unfamiliar words (see Figure 3c).

These results reveal considerable similarity in how pitch changes are treated in English and tone changes are treated in Chinese in infants at 7.5 and 9 months of age. By 11 months however, it appears that infants deploy language specific rules in their treatment of pitch-driven contrasts in English and Mandarin. In English, pitch changes are successfully disregarded whereas in Chinese, those changes that drive tone shifts are successfully incorporated into lexical representations. An additional question is whether these abilities develop in tandem or independently. Previous studies investigating phoneme discrimination have demonstrated evidence of an inverse relationship between native language discrimination and non-native discrimination of vowel contrasts (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005). However, there have been no other reports of whether the degree of precocity in one language is associated with the degree of precocity in a second language in bilingual infants. To investigate this, correlation coefficients were computed between listening times to pitch mis-matched words and tone-mismatched words at 11 months when infants' responses begin to align themselves with native language properties. The correlation between these variables approached significance ($r(14) = .47, p = .07$), tentatively suggesting that infants' abilities to normalize for pitch changes in English may be associated with their abilities to assign relevance to tone in Mandarin at 11 months (see Figure 4).

Discussion

The objective of the present study was to explore the phenomenon of native phoneme attunement in infants who are in the process of learning two distinct language systems: English, in which pitch changes are not lexically paramount to word identity versus Mandarin, in which pitch changes are essential to lexical meaning. In the current sample of Mandarin English learners, early memories for words appeared highly specific and to over-preserve pitch characteristics whether they are phonemic or not. When tested with Chinese stimuli, the responses of 7.5 month old infants appear to conform to Chinese phonology in that words are distinguished based on tone. However, the same pattern of results was observed with English stimuli (recognition of pitch-matched forms and rejection of pitch-mismatched forms) and therefore is interpreted as being due to a heavy reliance on acoustic and phonetic matches between encounters of words at this age (Houston & Jusczyk, 2000; Singh et al., 2004; Singh et al., 2008). At 9 months, pitch information appears to be downweighted both when tested on recognition of pitch transpositions of familiarization stimuli in English and on tone shifts in Mandarin Chinese. Disregarding the pitch changes introduced in each language at this age is conducive to word recognition in English but contravenes the phonological rules of Mandarin Chinese. At this age, infants appeared to

globally under-represent the importance of pitch in their lexical representations and did not selectively define words according to language-specific properties. Finally, at 11 months, still prior to the onset of a productive vocabulary for most children, bilingual infants showed evidence of linguistic differentiation. At this age, non-phonemic pitch changes in English were disregarded in word recognition tasks but tonal variants in Mandarin Chinese were rejected as repetitions of the same word. Finally, there was a suggestion of correlated abilities to disregard pitch variation in English with the ability to incorporate tone changes into memories for words in Chinese at 11 months. There is prior evidence that children learning two languages demonstrate linguistic interdependence in that in various domains of language, proficiency is correlated across both languages in bilingual children (Cummins, 1979; Ordóñez, Carlo, Snow, & McLaughlin, 2002; Sheng, McGregor, & Marian, 2006). However, this is the first suggestion of potential interdependence in bilingual infants. Whether any association at this stage is a direct one or mediated by information processing abilities more generally cannot be ascertained by the current study.

In previous studies with monolingual infants, learners were seen to transition from an over-reliance on suprasegmental detail at 7.5 months to generalization across mismatched forms at 9 months for pitch changes (Singh et al., 2008) and at 10.5 months for more complex spectral transformations such as those introduced by variation in talker gender, focusing stress, and emotion changes (Bortfeld & Morgan, 2010; Houston & Jusczyk, 2000; Singh et al., 2004). In the current study (Experiment 1), bilingual learners seemed to demonstrate a similar developmental sequence in English, beginning with over-representation of suprasegmental detail in both languages to generalizing across pitch changes by 9 months. However, in Experiment 2, the course of change in Mandarin word recognition entailed a three-stage sequence, a pattern not uncommon in child language acquisition (e.g. Bosch & Sebastián-Gallés, 2003; Ferguson & Farwell, 1975; Vihman & Croft, 2007). Such sequences have been observed in bilingual learners' phonetic sensitivities but also in morphological and phonological development in toddlers. This pattern is commonly attributed to an initial state where the units of representation are unanalyzed linguistic objects, followed by a period of fine-grained analysis which may result in the appearance of imprecision in language processing tasks (a putative consequence of a learner's closer inspection of linguistic structure). This analysis then culminates in a more sophisticated ability to coordinate and simultaneously process multiple sources of information in the input and to appropriately weight phonetic details based on their linguistic function. In the current study, the presence of an intermediate state differentiates bilingual learners' performance in English and Mandarin word recognition. However, by virtue of our design, it remains to be seen whether this developmental sequence is typical for Mandarin monolingual learners or whether it is typical for bilingual learners of English and Mandarin. The inclusion of a Mandarin-only group would help to clarify this issue.

The appearance of a three-stage sequence in Mandarin Chinese is consistent with previous findings on infant vowel discrimination in bilingual learners (e.g. Sebastián-Gallés & Bosch, 2009), demonstrating a middle stage when perceptual sensitivities are aligned with the properties of one language prior to language-specific responses to each language. Mattock et al. (2008) note the connection between vowels and lexical tone, which are similar in that both types of segments are important carriers of fundamental frequency variation. While cues to lexical tone span the entire syllable (Lee, 2009), tone cannot be communicated independently of vowels. Due to the acoustic similarity of tones and vowels, it is possible that changes in tone perception are associated with a developmental trajectory that closely mirrors that observed in investigations of vowel discrimination (e.g. Bosch & Sebastián-Gallés, 2003; Sebastián-Gallés & Bosch, 2009).

There are at least two possible explanations for concordant findings for vowels and tones, which are not mutually exclusive: a periodicity bias and prosodic factors. First, vowels and tones share acoustic features. Both types of sounds are long-term acoustic events associated with steady-state components of the waveform (Crystal & House, 1988). It is possible that acoustic events such as vowels and tones are processed earlier in development than more transient units such as consonants because they are preferentially attended to early in development. This idea is known as a 'periodicity bias', originally proposed by Cutler and Mehler (1993), which predicts that segments such as vowel and lexical tone will be favored by younger infants and therefore, perhaps analyzed earlier in development by virtue of their attentional appeal. This may be particularly relevant to the early phonological lexicon, which is defined by the sound structure of words rather than their relationship to meaning. It is plausible that structures that are analyzed earlier in development are done so when the architecture of the phonological lexicon is less mature, hence the appearance of an intermediate 'regressive' stage in language-specific responsiveness to vowels and lexical tones. Towards the end of infancy, when linguistic analysis is more sophisticated, phonological processing associated with this age (i.e. consonant discrimination) can perhaps develop with greater efficiency for both monolinguals and bilinguals. Given the similarities between vowels and tones, it is perhaps to be expected that vowels and tones are subject to perceptual reorganization between 6 and 9 months (Kuhl et al., 1992; Mattock & Burnham, 2006; Polka & Werker, 1994) whereas consonant categories are pruned later between 10 and 12 months (Werker & Tees, 1984). Therefore, the process of linguistic differentiation at the phonological level may be a two-stage process for consonants and a three-stage process for vowels and tones as a result of the age at which these two types of information are incorporated into a learner's phonological inventory¹. Alternatively, it is possible that the appearance of an intermediate 'regressive' state is primarily a product of methodological differences, and that when task demands are relatively simple, infants' phonological development advances in a more linear fashion (see Fennell & Waxman, 2010).

The notion that structural similarities of vowels and tones may lead them to be similarly treated in language processing as compared with consonants could be empirically tested in a number of possible ways. First, it is possible to experimentally investigate infants' sensitivities to vowel, tonal and consonantal mispronunciations in word learning tasks. There is a growing literature on infants' abilities to detect vowel and consonants mispronunciations both in learning novel words (Mani & Plunkett, 2007) and in recognizing known words (Mani & Plunkett, 2007; Nazzi, 2005; Swingley & Aslin, 2000; 2002) although there have been no studies to date comparing vowel and consonant mispronunciations with tone mispronunciations. Such a study would potentially enable a ranking of different types of phonetic segments in terms of how strongly encoded they are into memories for words by virtue of how sensitive infants are to substitutions. Secondly, to date, studies focusing on the phonological attunement of tones (Mattock & Burnham, 2006; Mattock et al., 2008) have focused on discrimination of non-native tones. To date, there have not been any studies on native tone discrimination in tone language learners. Such a study, if compared to vowel and consonant discrimination, would shed light on whether infants attune to vowels and tones according to a similar developmental timetable, as distinct from consonants. Therefore, a 'periodicity bias' account awaits further empirical investigation and should be judiciously interpreted as an explanation for the current study as the paradigms used in speech discrimination research are notably different to auditory word segmentation task. In particular, the incorporation of tone changes into lexical units likely may not exploit the

¹It should be noted that there is uncertainty as to when phonological processing becomes distinct from phonetic processing. For the purposes of this discussion, the term *phonological* processing is used to refer to perceptual sensitivities that are aligned with the properties of a learner's native language.

same resources as when phonetic segments are presented in isolation in discrimination or habituation paradigms and findings may not be directly comparable.

Secondly, vowels and tones are not just carriers of segmental information but also of suprasegmental information such as prosody. Prosody is particularly favoured by infants in attentional tasks (e.g. Fernald & Kuhl, 1987; Mehler et al., 1988). As a result, the singular attention-getting features of infant-directed prosody have been characterized as a possible point of entry or 'hook' into more sophisticated linguistic analyses (Jusczyk, 2001). Tonal languages are not exceptional in their prosodic exaggeration: maternal IDS by tone language speakers shares similar acoustic properties as that of non-tone languages speakers (Grieser & Kuhl, 1988). Therefore, there may be competition between low-level perceptual biases towards salient prosodic features conveyed by vowels/tones and a linguistic analysis of these units as phonetic segments. By contrast, the temporal dynamics of consonants restrict their functionality, such that they are not primary carriers of prosodic variation in any language. Again, the relative complexity of this task is compounded for bilingual learners learning one tone language and one non-tonal language as such learners must distinguish pitch changes that drive prosodic variation from those that drive tone alternations. Similarly in the case of vowel perception, learners must differentiate shifts in vowel quality associated with prosodic variation from those which are phonemic, even though the acoustic correlates of these changes can overlap, such as in the case focusing stress versus lexical stress (de Jong & Zawaydeh, 2002). The added burden of sorting segmental and suprasegmental information in the case of vowels and tones may be reflected in a more complex pathway for these types of phonemes compared with consonants. The role of an early periodicity bias and possible confusion of segmental and suprasegmental features are therefore two candidate explanations for the consistency between the current findings with vowel perception and not with consonant perception. However, further study is required in order to conjoin research from separate laboratories, populations and paradigms to form a more coherent account of phonetic category structure acquisition in bilingual learners.

While infants' performance on Experiment 1 has been contrasted with their performance on Experiment 2 to form a coherent account of how pitch variation is treated in bilingual learners, the comparability of the paradigms involved should not be overstated. It should be noted that there are significant differences between Experiment 1 and Experiment 2 in terms of stimuli and task demands placed on the participant. First, in Experiment 1, there was an overall pitch transformation at the level of the sentence, whereas in Experiment 2, pitch contrasts are limited to target words. As a result, it is possible that Experiment 1 placed greater demands on participants than Experiment 2 in that learners encountered a different pitch level during the recognition phase that spanned entire phrases, rather than a change in pitch contour at the level of the word. The decision to avoid elevating or depressing sentential pitch levels in Experiment 2 was based on the finding that pitch variation within a tone can shift a tonal speaker's tone assignment under conditions of uncertainty (e.g. Leather, 1983). The alternative possibility of only elevating or depressing pitch levels of the target words within a sentence in Experiment 1 was discounted based on the reduction in stimulus naturalness observed at the onset and offset of the target word. Second, the pitch manipulation in Experiment 1 is a digital manipulation whereas tone variations in Experiment 2 reflect contrasts produced by the human vocal tract. Certainly, there are other changes that are naturally produced in speech which are primarily driven by shifts in pitch contours (e.g. questions/statements; affective prosody) that do not change the lexical identity of words that could have been employed in place of the pitch manipulation in Experiment 1. However, these reflect dimensions that do bear on semantic interpretation and also strongly recruit attention in infants (Best, Levitt, & McRoberts, 1991; Kitamura & Lam, 2009; Singh, Morgan, & Best, 2002). Therefore, these types of changes are communicatively relevant and may not be normalized for based on their attention-getting properties and relative salience in

the input to infants. Therefore, Experiments 1 and 2 involve different task demands owing to the structure of the experiments and to the specific stimulus manipulations involved (i.e. digitized pitch transformations across the phrase versus naturally produced tonal alternations of a target word). There are at two least possible influences of the latter distinction on our findings. First, it is possible that non-natural transformations are more easily disregarded than natural transformations. A second (and related) possibility is that simpler spectral changes are more easily normalized for than more complex changes. The sort of pitch variation employed in Experiment 1 is a simple transformation where one dimension of the familiarization tokens (fundamental frequency) is multiplied by a constant value to derive the mis-matched forms. This is arguably simpler than the range of acoustic parameters that drive changes in lexical tone. The principal feature in tone variation is considered to be fundamental frequency but there are concomitant changes in amplitude, duration and voice quality (Burnham & Mattock, 2007) resulting in a wider range of transformations to 'undo' when generalizing across forms. It is possible that the simplicity of changes affects the ease of generalization, an idea put forth by Sommers, Nygaard, and Pisoni (1994) with regards to adult speech processing. These possibilities could be further investigated by replicating the current design with naturally produced pitch variations that are equated for discriminability with the tonal variants employed herein.

While the current study informs us about when bilingual learners begin to appreciate the conflicting role of pitch in their languages, it leaves open a number of questions. First, it remains unclear when Chinese learners learn to disregard non-phonemic pitch or intonational changes in Chinese word recognition and whether infants' abilities in this regard develop concurrently in English and Chinese. Prosodic variation in tone languages is not well studied, although there is some evidence that emotional prosody at a syllable level may be tempered in tone languages due to the possibility of lexical ambiguity resulting from pitch modulation in such languages (Ross, Edmondson, & Seibert, 1986). It would be interesting to investigate when and how tone language learners differentiate suprasegmental prosody from lexical tone in a tonal language. In addition, further studies are underway to determine how changes in tone, vowel, and consonant identity influence performance in word to meaning mapping tasks. In infant research, unimodal (auditory-only) paradigms have occasionally been shown to yield different findings as compared with intermodal paradigms (e.g. Ostroff & Cooper, 2003; Stager & Werker, 1997 vs. Werker & Tees, 1984; Werker & Lalonde, 1988). For example, it is possible that auditory paradigms inflate the use of acoustic information given the absence of meaningful visual cues and the resultant low level of visual stimulation. Infants' tolerance of mispronunciations of tone versus consonants and vowels in an intermodal context is currently being investigated to establish the order of precedence of these three types of phonemes. There is developmental evidence that tone-learning monolingual children can successfully map even subtle tone contrasts onto novel objects and that by 10 years of age, their performance on this task is comparable to that of adults (Ciocca & Lui, 2003). In an eye-tracking study with monolingual English children, Quam and Swingley (2010) demonstrated that two-year old English-learning infants and adults were sensitive to changes in vowel quality when learning new words but less so to non-phonemic changes in pitch (although English learners' abilities to use pitch cues to ascribe emotions to others develops much later (see Quam & Swingley, in press)). However, there is currently no evidence investigating whether bilingual learners of a tone and non-tone language can differentially deploy phonological rules in word learning tasks. In other words, it remains to be seen whether such learners can incorporate tone variation into lexical representations in their tonal language but disregard intonational variation in both of their languages.

The primary goal of this study was to investigate the point at which tone/non-tonal bilingual learners master the conflicting roles of pitch variation across their two languages. Although

they represent a scarcely researched population, tone language learners face a unique set of challenges due to this conflict. Studying the acquisition of tone languages can elucidate the constraints on early phonological development and in particular, on how the input is partitioned into phonemic and non-phonemic information in learners for whom the valuable segmental/suprasegmental distinction is blurred. In broader terms, the overall under-representation of tone language learners in developmental psycholinguistics is incongruent with the vast representation of tone language learners across the world with the most widely spoken languages being tone languages from Mainland China. Likewise, it is important that the normative course of first language acquisition be defined with reference to both monolinguals and bilinguals, particularly given that it is the latter that represents the statistical norm (Grosjean, 1982; Romaine, 1995). There is increasing evidence of distinct pathways for monolinguals and bilinguals (Bialystok, 1988; Genesee & Nicoladis, 1995), highlighting a need for further basic research in this area. As astutely observed by Grosjean (1989), a bilingual is a different type of learner who “is not two monolinguals in one person” and as such, theories of monolingual first language acquisition cannot readily be extended to bilinguals. For example, in clinical settings, simply being a bilingual learner in a predominantly monolingual society increases risk for diagnosis of language impairment (Crutchley, Conti-Ramsden, & Botting, 1997; Gutierrez-Clellen, 1996) which has been attributed to longstanding inappropriate use of monolingual norms for this population (Genesee & Nicoladis, 1995; Patterson, 1998; Pearson, Fernandez, & Oller, 1993). This exemplifies the fact that the typical language learning journey of a bilingual child remains somewhat elusive and is much less well understood than that of monolinguals. Over the past few years, increasing emphasis has been placed on understanding the earliest phase of this journey, demonstrating an amalgam of precocities and protractions as compared with monolinguals (Bosch & Sebastián-Gallés, 2003; Byers-Heinlein, Burns, & Werker, 2010; Fennell et al., 2007; Kovacs & Mehler, 2009a; 2009b). However, much of the evidentiary record has yet to be filled in. The current study provides evidence that in one domain of language processing in infancy - early spoken word recognition - at 7.5 and 9 months, bilingual learners appear to treat pitch variation in English and tone variation in Chinese in similar ways in spite of their distinct linguistic functions. However, Chinese English learners and English monolinguals demonstrate quite sophisticated mastery of language-specific phonological rules by 11 months for their native language(s) with no evidence of a delay in English as compared with English monolinguals.

Acknowledgments

This project was supported by the National Institutes of Health (NIH 1 R03 HD046676-01A1) to L.S. We thank Sarah Nestor, Chandni Parikh and Ashley Yull for assistance with recruitment and testing and Dr. Wei Quin Yow for assistance with statistical analyses.

Appendix A

Words and passages used in Experiment 1

Bike

His **bike** had big black wheels

The girl rode her big **bike**

Her **bike** could go very fast

The bell on the **bike** was really loud

The boy had a new red **bike**

Your **bike** always stays in the garage.

Hat

She put on her **hat** to play in the snow.

The **hat** was soft and warm

Her brother had knitted the **hat**

The **hat** was blue and white

She liked how the **hat** covered her ears

Her friends also liked her **hat**.

Tree

The **tree** was a hundred years old

The **tree** grew in the man's back yard

He liked to look outside at the **tree**

Hanging from the **tree** was a swing

The man's grandchild played in the **tree**

The leaves on the **tree** were yellow

Pear

The juicy, green **pear** came from the basket

The **pear** is her favorite fruit.

She wanted to eat the biggest **pear**.

The **pear** in the basket looked very good.

Next to the **pear** was an apple.

She ate the whole **pear**.

Appendix B

Words and passages used in Experiment 2

Tou (Tone 2)“head”

Meige ren de **tou** dou bu yiyang. “Everybody's head is different.”

Tou keyi cangzai maozi dixia. “A head can be hidden inside a hat.”

Wo de hao pengyou shi Da-**tou**. “My best friend is Big Head.”

Shizi de **tou** hen qiguai. “The head of a lion is very weird.”

Tou buyao shenchu che wai. “Don’t stick your head outside of the car.”

Linju de nanhai jiao Xiao-**tou**. “The neighbor’s boy is called Small Head.”

Dan (Tone 4) “egg”

Dan fangzai bingxiang li. “Eggs are put in the refrigerator.”

Wo mai da **dan**-gao gei ni. “I bought a big egg-cake for you.”

Muji xiale haoduo **dan**. “The hen laid many eggs.”

Youxie ren bu ai chi **dan**. “Some people don’t like eating eggs.”

Xiaohua ba **dan** dapole. “Xiaohua broke the eggs.”

Dan gundao zhuo-zi dixia qu le. “Eggs rolled down under the table.”

Tian (Tone 1) “sky, day”

Tian shang you ji duo baiyun. “There are some clouds in the sky.”

Ta tangzai dishang kan lan-**tian**. “He is lying on the ground looking at the sky.”

Women xingqi- **tian** qu kao rou. “We had a barbecue on Sunday (lit. week-day).”

Tian kong feichang de yin’an. “The sky is very dark.”

Dajia dou xihuan qing **tian**. “Everybody likes clear days.”

Xiayu **tian** bu neng qu dongwu yuan. “One can’t go to the zoo in raining days.”

Bei (Tone 1) “cup”

Milaoshu zai zhao cha-**bei**. “Mickey Mouse is looking for a teacup.”

Zhe ge xiao **bei**-zi hen piaoliang. “This little cup is very pretty.”

Bei-zi li you guozhi. “There is juice in the cup.”

Bei di zuozhe yi zhi qingwa. “There is a frog sitting in the bottom of the cup.”

Ta ba yi da **bei** niunai heguangle. “He drank up a big cup of milk.”

Wo song ni yi ge boli **bei**. “I’ll give you a glass cup.”

REFERENCES

- Albareda-Castellot B, Pons F, Sebastián-Gallés N. The acquisition of phonetic categories in bilingual infants: New data from an anticipatory eye movement paradigm. *Developmental Science*. 2011; 14(2):395–401. [PubMed: 22213908]
- Anderson JL, Morgan JL, White KS. A statistical basis for speech sound discrimination. *Language and Speech*. 2003; 46(2-3):155–182. [PubMed: 14748443]
- Best, CT.; Levitt, A.; McRoberts, G. Examination of language-specific influences in infants’ discrimination of prosodic categories; Proceedings of the 12th International Congress of Phonetic Sciences; Aix-en-Provence, France. 1991; p. 162-165.

- Best CT, McRoberts GW, Sithole NM. 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*. 1988; 14(3):345–360. [PubMed: 2971765]
- Bialystok E. Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology*. 1988; 24(4):560–567.
- Birch S, Clifton C. Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *Journal of Memory and Language*. 2002; 47(4):571–588.
- Blicher DL, Diehl RL, Cohen LB. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*. 1990; 18(1):37–49.
- Boersma P, Weenink D. PRAAT: A system for doing phonetics by computer. Report of the Institute for Phonetic Sciences of the University of Amsterdam. 1996; 132:1996. <http://www.praat.org>.
- Bortfeld H, Morgan JL. Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology*. 2010; 60(4):241–266. [PubMed: 20159653]
- Bosch L, Sebastián-Gallés N. Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*. 2003; 46(2-3):217–243. [PubMed: 14748445]
- Burnham, D.; Mattock, K. The perception of tones and phones. In: Bohn, O-S.; Munro, MJ., editors. *Language experience in second language speech learning: In honor of James Emil Flege*. John Benjamins Publishing Company; Amsterdam: 2007. p. 259-280.
- Burns TC, Yoshida KA, Hill K, Werker JF. The development of phonetic representation in bilingual and monolingual infants. *Applied Psycholinguistics*. 2007; 28(3):455–474.
- Byers-Heinlein K, Burns TF, Werker JF. The roots of bilingualism in newborns. *Psychological Science*. 2010; 21(3):343–348. [PubMed: 20424066]
- Ciocca V, Lui JYK. The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders*. 2003; 1(2):141–147.
- Crutchley A, Conti-Ramsden G, Botting N. Bilingual children with SLI and standardized assessments: Preliminary findings from a study of children in language units. *International Journal of Bilingualism*. 1997; 1(2):117–134.
- Crystal, D. *English as a global language*. Cambridge University Press; Cambridge: 1997.
- Crystal TH, House AS. Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*. 1988; 83(4):1553–1573.
- Cummins J. Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*. 1979; 49(2):222–251.
- Cutler A, Mehler J. The periodicity bias. *Journal of Phonetics*. 1993; 21(1-2):103–108.
- de Jong K, Zawaydeh B. Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*. 2002; 30(1):53–75.
- Fennell CT, Byers-Heinlein K, Werker JF. Using speech sounds to guide word learning: The case of bilingual infants. *Child Development*. 2007; 78(5):1510–1525. [PubMed: 17883445]
- Fennell CT, Waxman SR. What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*. 2010; 81(5):1376–1383. [PubMed: 20840228]
- Ferguson CA, Farwell CB. Words and sounds in early language acquisition. *Language*. 1975; 51(2):419–439.
- Fernald A, Kuhl PK. Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development*. 1987; 10(3):279–293.
- Fromkin, V. *Tone: A linguistic survey*. Academic Press; New York: 1978.
- Gandour J. Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*. 1984; 12(2):235–261.
- Genesee, F.; Nicoladis, E. Language development in bilingual preschool children. In: Garcia, E.; McLaughlin, B., editors. *Meeting the challenge of linguistic and cultural diversity in early childhood education*; 1995; New York: Teachers College Press; p. 18-33.
- Grieser DL, Kuhl PK. Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*. 1988; 24(1):14–20.

- Grosjean, F. Life with two languages: An introduction to bilingualism. Harvard University Press; Cambridge: 1982.
- Grosjean F. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*. 1989; 36(1):3, 15. [PubMed: 2465057]
- Gutierrez-Clellen, VF. Language diversity: Implications for assessment. In: Cole, KN.; Dale, PS.; Thal, DJ., editors. *Assessment of communication and language*. Vol. 6. Paul H. Brookes; Baltimore: 1996. p. 29-56.
- Hirschberg J, Ward G. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics*. 1992; 20(2): 241–251.
- Houston DM, Jusczyk PW. The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26(5):1570–1582. [PubMed: 11039485]
- Jongman, A.; Wang, Y.; Moore, CB.; Sereno, JA. Perception and production of Mandarin tones. In: Ping, L.; Li, HT.; Bates, E.; Tzeng, OJL., editors. *The handbook of East Asian psycholinguistics Vol I: Chinese*. Cambridge University Press; New York: 2006. p. 209-218.
- Jusczyk, PW. Bootstrapping from the signal: Some further directions. In: Hohle, B.; Weissenborn, J., editors. *Approaches to Bootstrapping: Phonological, Lexical, Syntactic and Neurophysiological Aspects of Early Language*. John Benjamins; Amsterdam: 2001.
- Jusczyk PW, Aslin RN. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*. 1995; 29(1):1–23. [PubMed: 7641524]
- Kemler Nelson DG, Jusczyk PW, Mandel DR, Myers J, Turk A, Gerken L. The head-turn preference procedure for testing auditory perception. *Infant Behavior & Development*. 1995; 18(1):111–116.
- Kitamura C, Lam C. Age-specific preferences for infant-directed affective intent. *Infancy*. 2009; 14:1–24.
- Kovacs AM, Mehler J. Flexible learning of multiple speech structures in bilingual infants. *Science*. 2009a; 325:611–612. [PubMed: 19589962]
- Kovacs AM, Mehler J. Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*. 2009b; 106:6556–6560.
- Kuhl PK, Conboy BT, Padden D, Nelson T, Pruitt J. Early speech perception and later language development: Implications for the critical period. *Language Learning and Development*. 2005; 1(3-4):237–264.
- Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*. 1992; 255(5044):606–608. [PubMed: 1736364]
- Lasky RE, Syrdal-Lasky A, Klein RE. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*. 1975; 20(2):215–225. [PubMed: 1194843]
- Leather J. Speaker normalization in perception of lexical tone. *Journal of Phonetics*. 1983; 11(4):373–382.
- Lee C. Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study. *The Journal of the Acoustical Society of America*. 2009; 125(2):1125–1137. [PubMed: 19206887]
- Li CN, Thompson SA. The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*. 1977; 4(2):185–199.
- Liu S, Samuel AG. Perception of Mandarin lexical tones when f_0 information is neutralized. *Language and Speech*. 2004; 47(2):109–138. [PubMed: 15581188]
- Mani N, Plunkett K. Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*. 2007; 57(2):252–272.
- Mattock K, Burnham D. Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy*. 2006; 10(3):241–265.
- Mattock K, Molnar M, Polka L, Burnham D. The developmental course of lexical tone perception in the first year of life. *Cognition*. 2008; 106(3):1367–1381. [PubMed: 17707789]

- Mattock K, Polka L, Rvachew S, Krehm M. The first steps in word learning are easier when the shoes fit: Comparing monolingual and bilingual infants. *Developmental Science*. 2010; 13(1):229–243. [PubMed: 20121879]
- Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*. 2002; 82(3):B101–B111. [PubMed: 11747867]
- Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoni J, Amiel-Tison C. A precursor of language acquisition in young infants. *Cognition*. 1988; 29(2):143–178. [PubMed: 3168420]
- Narayan CR, Werker JF, Beddor PS. The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*. 2010; 13(3):407–420. [PubMed: 20443962]
- Nazzi T. Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*. 2005; 98(1):13–30. [PubMed: 16297674]
- Ordóñez CL, Carlo MS, Snow CE, McLaughlin B. Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology*. 2002; 94:719–728.
- Ostroff WL, Cooper RP. Female voice facilitates 11-month-old infants' nonnative phoneme discrimination. 2003 Unpublished manuscript.
- Patterson JL. Expressive vocabulary development and word combinations of Spanish-English bilingual toddlers. *American Journal of Speech and Language Pathology*. 1998; 7(4):46–56.
- Pearson BZ, Fernandez SC, Oller DK. Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*. 1993; 43(1):93–120.
- Polka L, Colantonio C, Sundara M. A cross-language comparison of /d/-/delta / perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*. 2001; 109(5): 2190–2201. [PubMed: 11386570]
- Polka L, Werker JF. Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*. 1994; 20(2):421–435. [PubMed: 8189202]
- Quam C, Swingle D. Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*. 2010; 62:135–150. [PubMed: 20161601]
- Quam C, Swingle D. Development in children's interpretation of pitch cues to emotions. *Child Development*. in press.
- Romaine, S. *Bilingualism*. 2nd ed.. Blackwell; Oxford, UK: 1995.
- Ross ED, Edmondson JA, Seibert GB. The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice. *Journal of Phonetics*. 1986; 14(2):283–302.
- Sebastián-Gallés N, Bosch L. Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental Science*. 2009; 12(6):874–887. [PubMed: 19840043]
- Seidl A, Cristià A. Developmental changes in the weighting of prosodic cues. *Developmental Science*. 2008; 11(4):596–606. [PubMed: 18576967]
- Shen XS, Lin M. A perceptual study of Mandarin Tones 2 and 3. *Language and Speech*. 1991; 34(2): 145–156.
- Sheng L, McGregor K, Marian V. Lexical-semantic organization in bilingual children: Evidence from a repeated word association task. *Journal of Speech, Language, and Hearing Research*. 2006; 49:572–587.
- Singh L, Morgan JL, Best CT. Infants' listening preferences: Baby talk or happy talk? *Infancy*. 2002; 3:365–394.
- Singh L, Morgan JL, White KS. Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*. 2004; 51(2):173–189.
- Singh L, White KS, Morgan JL. 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*. 2008; 4(2):157–178.

- Sommers MS, Nygaard LC, Pisoni DB. Stimulus variability and spoken word recognition: I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America*. 1994; 96:1314–1324. [PubMed: 7962998]
- Stager LC, Werker JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*. 1997; 388:381–382. [PubMed: 9237755]
- Sundara M, Polka L, Molnar M. Development of coronal stop perception: Bilingual infants keep pace with their monolingual peers. *Cognition*. 2008; 108(1):232–242. [PubMed: 18281027]
- Swingle D, Aslin RN. Spoken word recognition and lexical representation in very young children. *Cognition*. 2000; 76(2):147–166. [PubMed: 10856741]
- Swingle D, Aslin RN. Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*. 2002; 13(5):480–484. [PubMed: 12219818]
- Tsay, J.; Jusczyk, PW. Detection of words in fluent Chinese by English-acquiring and Chinese-acquiring infants. In: Houston, D.; Seidl, A.; Hollich, G.; Johnson, E.; Jusczyk, A., editors. *Jusczyk Lab Final Report*. 2003.
- Vihman M, Croft W. Phonological development: Toward a “radical” templatic phonology. *Linguistics*. 2007; 45(4):683–725.
- Wang WS, Li K-P. Tone 3 in Pekinese. *Journal of Speech & Hearing Research*. 1967; 10(3):629–636. [PubMed: 6081948]
- Werker JF, Gilbert JH, Humphrey K, Tees RC. Developmental aspects of cross-language speech perception. *Child Development*. 1981; 52(1):349–355. [PubMed: 7238150]
- Werker JF, Lalonde CE. Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*. 1988; 24(5):672–683.
- Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*. 1984; 7(1):49–63.
- Yip, M. *Tone*. Cambridge University Press; Cambridge, UK: 2002.

Highlights

- We investigated sensitivity to pitch and tone variation in bilingual infants.
- Like monolinguals, younger infants defined words by pitch and tone characteristics.
- By 11 months, infants demonstrated language specific sensitivities.
- Sensitivity to tone contrasts in word recognition emerges by 11 months.

a

b

c

Figure 1.

a: Listening Times to Matched, Mis-matched and Unfamiliar English Passages in 7.5 month infants. b: Listening Times to Matched, Mis-matched and Unfamiliar English Passages in 9 month infants. c: Listening Times to Matched, Mis-matched and Unfamiliar English Passages in 11 month infants.

Figure 2.
a: Sample Pitch Contours of Tone 1 Familiarization Stimuli ('bei') b: Sample Pitch Contours of Tone 2 Familiarization Stimuli ('tou') c: Sample Pitch Contours of Tone 4 Familiarization Stimuli ('dan')

Figure 3.

a: Listening Times to Matched, Mis-matched and Unfamiliar Mandarin Chinese Passages in 7.5 month infants. b: Listening Times to Matched, Mis-matched and Unfamiliar Mandarin Chinese Passages in 9 month infants. c: Listening Times to Matched, Mis-matched and Unfamiliar Mandarin Chinese Passages in 11 month infants.

Figure 4.
: Listening Times to Tonal Mismatches (Chinese) in relation to Pitch Mismatches (English)

Table 1

Acoustic Analyses of Words and Sentences (means and SD)

	Mean fundamental frequency (Hz)	
	Words	Target Word in Sentences
Low Pitch	269.38 (43.91)	255.78 (23.53)
High Pitch	383.72 (61.47)	367.44 (32.64)

Table 2

Experiment 2 Design

Conditions	Training Stimuli		Target Words in Sentences		
	Matched Familiarization Item	Mismatched Familiarization Item	Sentence containing Tone Match	Sentence containing Tone Mismatch	Control Sentences
n=4 at each age group	bei [Tone 1]	tou [Tone 1]	bei [Tone 1]	tou [Tone 2]	tian [Tone 1] dan [Tone 4]
n=4 at each age group	tou [Tone 2]	bei [Tone 2]	tou [Tone 2]	bei [Tone 1]	tian [Tone 1] dan [Tone 4]
n=4 at each age group	tian [Tone 1]	dan [Tone 1]	tian [Tone 1]	dan [Tone 4]	bei [Tone 1] tou [Tone 2]
n=4 at each age group	dan [Tone 4]	tian [Tone 4]	dan [Tone 4]	tian [Tone 1]	bei [Tone 1] tou [Tone 2]

Table 3

Acoustic Analyses of Mandarin Chinese Stimuli

Tone	Mean fundamental frequency at onset and offset of Word (Hz)			Duration of word (msec)	
	Contour	Words	Target Word in Sentences	Words	Target Word in Sentences
Tone 1 Stimuli	High	354-348	366-378	698.24	442.54
Tone 2 Stimuli	Rising	223-425	213-439	708.15	552.24
Tone 4 Stimuli	Falling	431-126	432-164	523.91	381.12