Infants' Sensitivity to Lexical Tone and Word Stress in their First Year: A Thai and English Cross-Language Study

Marina Kalashnikova[1,2], Chutamanee Onsuwan[3], Denis Burnham[4]

[1]Basque Center on Cognition, Brain and Language

[2]Ikerbasque, Basque Foundation for Science

[3]MARCS-CILS NokHook BabyLab, Faculty of Liberal Arts, Thammasat University

[4]MARCS Institute for Brain, Behaviour and Development, Western Sydney University

Non-tone language infants' native language recognition is based first on supra-segmental then segmental cues, but this trajectory is unknown for tone-language infants. This study investigated non-tone (English) and tone (Thai) language 6- to 10-month-old infants' preference for English vs. Thai one-syllable words (containing segmental and tone cues) and two-syllable words (additionally containing stress cues). A preference for their native one-syllable words was observed in each of the two groups of infants, but this was not the case for two-syllable words where Thai-learning infants showed no native-language preference. These findings indicate that as early as six months of age, infants acquiring tone- and non-tone languages identify their native language by relying solely on lexical tone cues, but tone language infants no longer show successful identification of their native language when two pitch-based cues co-occur in the signal.

Keywords: Lexical tone; lexical stress; prosody; speech perception; infancy

In the first months of life, infants develop proficiency with the particular properties of their native language: its prosody and stress patterns, its phonological repertoire, and its syllabic structure and phonotactic rules, which then result in attunement to their native language. The developmental trajectory for this process has been investigated extensively in infants acquiring a variety of Germanic and Romance languages. This research has shown that infants rely on a combination of suprasegmental cues (e.g., word stress) that characterise the language, and segmental cues (e.g., consonant and vowel categories) that differentiate word meaning in the language, to identify and differentiate words that belong to their native language (e.g., Jusczyk, Cutler, & Redanz, 1993; Nazzi, Jusczyk, & Johnson, 2000; Nazzi & Ramus, 2003).

Here we extend this research to investigate the role of two pitch-based cues: lexical tone, a pitch-based cue that operates at the segmental (within-syllable) level and marks phonemic contrasts (Burnham & Mattock, 2007), and word stress, a pitch-based cue that operates at the suprasegmental (between-syllable) level, which is characteristic of the language but that does not mark phonemic contrasts. To address this issue and delineate more clearly the influence of native language on early speech perception, this study compares the development of native language preference in infants acquiring English, a non-tone language, which has a characteristic non-phonemic strong-weak stress pattern in disyllabic words, versus Thai, a lexical tone language with five lexical tones whose pitch patterns differentiate meaning at both the single-syllable and multi-syllabic level, and which has a characteristic (non-phonemic) weak-strong stress pattern in disyllabic words.

Infants learning non-tone languages differentiate language rhythm classes at birth (Nazzi, Bertoncini, & Mehler, 1998), then by around five months this ability becomes more finely-tuned to encompass within rhythm class differentiation (Bosch & Sebastian-Gallés, 1997; Molnar, Gervain, & Carreiras, 2014; Nazzi et al., 2000; Nazzi & Ramus, 2003; Ramus,

2002). In addition to this early sensitivity to prosodic information at the phrasal level, infants also attend to word-level prosodic cues, as evidenced by their preference for the dominant stress pattern in multisyllabic words in their native language. Infants between four and six months of age exhibit differential processing of two-syllable words that do and do not exhibit the dominant stress pattern of their language (e.g., for German-language infants strong-weak in German vs. weak-strong in French) (Höhle, et al., 2009; Friederici, Friedrich, & Christophe, 2007; Weber, Hahne, Friedrich, & Friederici, 2004). This sensitivity to stress plays a fundamental role in the segmentation of the speech stream into word units, as it provides infants with language-specific cues to identify word onsets in continuous speech (Jusczyk, 1997; Jusczyk, Cutler, et al., 1993; Mattys, Jusczyk, Luce, & Morgan, 1999).

Later, around nine months of age, infants begin to show preferences based on the phonemic and phonotactic qualities of their native and a non-native language (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk & Luce, 1994). For instance, Jusczyk and colleagues (Jusczyk, Friederici, et al., 1993) showed that English-learning six-month-olds differentiate unfamiliar words in English and Norwegian, two languages that differ in their dominant stress pattern (strong-weak in English and weak-strong in Norwegian), while only nine-month-olds differentiate between English and Dutch, two languages that do not differ in their stress pattern but do differ in their phonemic inventories and phonotactic structure. The emergence of the ability to attend to language-specific segmental information in this process assists infants in the challenging task of identifying word forms in continuous speech and linking these forms to meaning (Mattys et al., 1999).

The evidence so far suggests that infants learn to rely on suprasegmental and segmental cues to identify their native language. Specifically, stronger reliance on suprasegmental cues (prosodic and word stress cues) is observed early in the first semester of life, whereas stronger reliance on segmental (phonotactic and phonetic) cues is observed later

around nine months of age. The developmental locus of the emerging ability to attend to phonetic detail in this process is not surprising as it coincides with the development of perceptual attunement, the transition from language-general to language-specific perception, which is manifested in maintained or increased sensitivity to native and decreased sensitivity to most non-native speech contrasts (see Kuhl, 2004; Werker, 2018; Werker & Hensch, 2015; Werker & Tees, 2005 for comprehensive reviews).

While phonemic contrasts are marked by consonant and vowel segments in all languages of the world, the majority of the world's languages are lexical tone languages, which additionally use pitch-based lexical tones to mark phonemic contrasts. Lexical tone languages comprise approximately 60-70% of languages in the world (Yip, 2002) and are spoken by over 50% of the world's population (Fromkin, 1978). Lexical tones are based primarily on modulations related to F0 height and contour, but there are also other less prominent acoustic correlates such as duration, amplitude, and voice register (Gandour, 1983; Whalen & Xu, 1992). Given these acoustic properties and the fact that tones are carried on vowels and also adjacent consonants (Hombert, Ohala, & Ewan, 1979), lexical tones can be and have been classified as supra-segmental. Nevertheless, lexical tones operate in exactly the same way as consonants and vowels – each of the three differentiate meaning, and each does so independently of the other two. Thus, at the functional level, lexical tones can be considered segments. However, as they depend upon temporal modulations related to pitch just like stress and prosody, at the structural level lexical tones can be considered suprasegmental (Burnham & Mattock, 2007; Burnham et al., 2011). While the experiment to be conducted here does not depend upon resolution of this structural-functional debate, the issues it raises are of interest and will be revisited in the discussion of the results.

In this study we investigate the role of two pitch-based cues in infants' native language preference: (i) the presence (Thai) or absence (English) of syllable-level pitch-based

lexical tones that are phonemic, and (ii) the pattern of word-stress, the characteristic distribution pattern of pitch across syllables, Weak-Strong (Thai) and Strong-Weak (English) that is non-phonemic. As Thai incorporates lexical tone, a pitch-based cue to meaning not used in English, the results may shed light on the role that lexical tones could play in shaping early sensitivity to the specific pitch-based and segmental/supra-segmental information in infants' native languages.

Turning to developmental studies of lexical tones and in particular to perceptual attunement, there is support for the view that lexical tones function in a similar manner to segments. Mattock and Burnham (2006) demonstrated that Mandarin and Cantonese-learning infants showed a stable ability to discriminate Thai tones from six to nine months of age, whereas for English-learning infants, discrimination ability for Thai tones decreased from six to nine months (see also Mattock, Molnar, Polka, & Burnham, 2008; and for similar findings with other non-tone language infants, Cabrera et al., 2015; Shi, Santos, Gao, & Li, 2017). This time frame is similar to that for infants' perceptual attunement to consonants (around eight to 12 months; Best, McRoberts, LaFleur, & Silver-Isenstadt,1995; Werker, Gilber, Humphrey, & Tees, 1981) and vowels (around six to eight months; Kuhl, Williams, et al., 1992; Polka & Werker, 1994), suggesting similarity in the development of perceptual attunement for phonemically-relevant consonants, vowels, and lexical tones. On the other hand, Yeung, Chen, and Werker (2013), testing discrimination of Cantonese tones by four-month-olds, showed a language-specific advantage for native tone language Cantonese-learning infants not only over non-native non-tone language English-learning infants, but also over non-native tone-language Mandarin-learning infants. The results of this study by Yeung et al. highlight both similarities and differences between tone versus consonant/vowel perceptual attunement: they are similar because, as for consonants and vowels, perceptual attunement for tones is finely tuned to subtle phonetic differences between languages, in this

case the phonetic (tonetic) differences between Cantonese and Mandarin tones; but they are different because perceptual attunement for tones may begin earlier than for consonants and vowels (four months versus around six to eight months respectively).

There are further differences; more recent research has indicated that the developmental trajectory of tone perception is more complex than that of consonants and vowels (see Singh & Fu, 2016 for a review), calling into question the parallel between tones and consonants and vowels. For instance, for Mandarin tone language infants there is a gradual improvement in native tone perception across the second semester of life (Singh et al., 2018; Singh, Poh, & Fu, 2016; Tsao, 2017). Similarly, non-tone language infants have been shown to maintain or improve tone perception across their first 18 months of life (Chen & Kager, 2016; Chen, Stevens & Kager, 2017), in some cases following a U-shaped pattern of initial decrease in sensitivity followed by a resurgence (Götz et al., 2018; Liu & Kager, 2014; 2018), thus maintaining sensitivity to non-native tone categories well beyond the age at which perceptual attunement is proposed to have been completed (Hay, Cannistraci, & Zhao, 2019; Hay, Graf-Estes, Wang, & Saffran, 2015; Singh, Tam, Chan, & Golinkoff, 2014).

Based on the similarities between perceptual attunement for tones and for consonants and vowels, native language preference in tone language infants would be expected to follow the same developmental trajectory as in non-tone language infants, i.e., an early reliance on non-phonemic suprasegmental cues, such as word stress, followed by later inclusion of phonemic segmental cues – consonants and vowels for non-tone language infants, and consonants, vowels, and lexical tones for tone language infants. That is, tone language infants' speech perception would be expected to be shaped by native language experience early in life in the same manner as is the speech perception of non-tone language infants.

However, the differences between perceptual attunement for tones, and for consonants and vowels would militate against such a conclusion. Compared to perceptual

attunement for consonants and vowels, perceptual attunement for tones starts earlier (Yeung et al., 2013), has a non-linear trajectory (Liu & Kager, 2014; 2018; Götz et al., 2018), and has less sharply defined differences between trajectories for tone and non-tone language infants (Chen & Kager, 2016; Chen et al., 2017). In light of these differences and given that native language preferences and perceptual attunement are related, there is reason to expect differences in the developmental trajectory for native language preference in non-tone and tone language infants. That is, it is possible that an earlier reliance on word stress followed by a later reliance on consonants, vowels, and lexical tones would not be observed in infants acquiring a tone language.

In addition to the two alternative possibilities presented above, the question may be raised as to whether infants employ lexical tone information at all to differentiate their native from a non-native language. To date, just one study has investigated this question, but only in infants learning a non-tone language. Burnham, Kitamura, and Lancuba (1999) tested six-, seven-, and nine-month-old English-learning infants' differentiation of unfamiliar one- and two-syllable words in English and in Thai in a listening preference paradigm. Aside from its use of lexical tones, as noted above, Thai also differs from English in its dominant stress pattern (weak-strong in Thai compared with strong-weak in English). In their task, over and above different phonological inventories of consonants and vowels, infants had access to both word stress and lexical tone cues in the two-syllable words, but only to lexical tone cues in the one-syllable words. When data for the three age groups were collapsed, infants showed a significant native language preference in the two-syllable but not the one-syllable condition. Therefore, in two-syllable words, English-learning infants used language-specific information, the dominant word stress pattern, but in one-syllable words they did not – the absence of lexical tone in English one-syllable words compared with its presence in Thai one-syllable words did not elicit a native language preference. Furthermore, even in the two-

syllable condition, the native language preference was not robust. Separate analyses for each age group revealed that only the seven-month-old group, and neither the six- or nine-month-old groups, showed a statistically significant preference for English over Thai.

The Burnham et al. (1999) findings in the one-syllable condition indicate that infants who do not encounter lexical tones in their language background, do not rely on this cue to differentiate a non-native language from their native language. However, as this study only involved non-tone language infants, it leaves open the question about whether and to what extent infants from tone language backgrounds rely on lexical tones in the process of language differentiation. To investigate this issue, we assessed the ability to use Thai lexical tones to differentiate lists of words in two groups of six- to ten-month-old infants, those acquiring a non-tonal language (English) and those acquiring a tonal language (Thai). Following Burnham et al., we manipulated infants' access to word stress information by using two-syllable and one-syllable words. Two-syllable words provide phonetic segmental information, and two forms of pitch information – lexical tone information (present or absent), and word stress information (native or non-native). In one-syllable words, there is no word-stress information, so only phonetic segmental and lexical tone information (native or non-native) is available.

We employed a preferential looking paradigm in which infants' longer looking time to native compared with non-native words is used to index a native language preference (Colombo & Bundy, 1983). The design incorporated two between-subject factors, Group (English infants vs. Thai infants) and Condition (one-syllable words vs. two-syllable words), and one within-subjects factor, Stimulus Language (English words vs. Thai words). Age was treated as a continuous variable in order to capture the developmental trajectory in infants' performance rather than differences between discrete time points and to maximise the power of our statistical models. We expected to find an overall native-language preference in each

group of infants (Group × Stimulus Language interaction). In addition, we expected this preference to be modulated by infants' native language, age, and the availability of suprasegmental word stress cues (present in two- but not one-syllable words). First, if it is the case that native phonemic information alone (consonants, vowels, and lexical tones) is not sufficient for successful language differentiation as suggested by Burnham et al. (1999), then we predicted that regardless of language background, infants would only show a preference for their native language in the two-syllable and not in the one-syllable condition (Group × Stimulus Language × Condition interaction). Alternatively, we considered the possibility that infants' language background would modulate their reliance on lexical tones, and that the developmental trajectory for the emergence of this sensitivity would be related to the developmental process of perceptual attunement. In this case, the prediction for the two-syllable condition was as stated above, but for the one-syllable condition, we expected to observe an increase in native language preference across our age range, but only in the group of infants acquiring a tone language (Group × Stimulus Language × Condition × Age interaction).

Method

Participants

Sixty-nine infants between six and ten months of age participated in the study. Thirty-five infants were growing up in Sydney, Australia, learning English as their native language (*M* age = 34.61 weeks, *SD* = 5.49; 15 male). Thirty-four infants were growing up in Bangkok, Thailand, learning Thai as their native language (*M* age = 35.74, *SD* = 5.69; 26 male). Age did not differ between the two groups, $t(67) = .843$, $p = .402$, $d = .206$. Half of the infants in each language group were randomly assigned to the one-syllable condition (17 English, 17 Thai) and half to the two-syllable condition (18 English, 17 Thai). Eight additional infants participated but were excluded (3 English and 5 Thai) due to experimenter

error (2), failure to complete the experiment due to fussiness or crying (4), failure to look at the two screens in the familiarisation trials (1), and exposure to the non-native test language at home (1).

All infants were born full-term, were typically developing, had no exposure to any additional language, and were recruited via each laboratory's database of families who expressed interest in taking part in university infancy research. It was ensured that identical procedures were followed in Australia and in Thailand for recruitment, obtaining consent, and laboratory visits.

Stimuli and Apparatus

Four lists of 20 words each were used as stimuli in this study: English one-syllable words, English two-syllable words, Thai one-syllable words, and Thai two-syllable words. A balanced bilingual Thai-English speaker who is also a trained phonetician (second author) was recorded producing all the words in adult-directed speech. While infants show greater preference to listening to infant-directed speech (Dunst, Gorman, & Hamby, 2012), adult-directed speech was used to avoid possible interactions between the exaggerated F0 contours in this register and those corresponding to the lexical tones. All words were low frequency abstract words, so they were unlikely to be familiar to the infants. Several precautions were taken to ensure that the words accurately represented the lexical inventories of English and Thai. First, 50% of words in each list contained phonetic segments that were non-phonemic or that violated the phonotactic constraints in the other language. For example, in English the phoneme /s/ occurs in both word-initial and word-final positions, but only in word-initial position in Thai, and the English phoneme /v/ is not part of the Thai phonemic inventory; and the vowel /ɯ/ is phonemic in Thai but not phonemic in English. Second, 90% of words in English followed the dominant strong-weak stress pattern, and 10% followed the non-dominant, but permissible weak-strong or strong-strong patterns (Cutler & Carter, 1987). The

reverse was the case for Thai with 90% of the two-syllable words following the dominant weak-strong stress pattern, and 10% of two-syllable words following the strong-weak or strong-strong patterns (Potisuk, Gandour, & Harper, 1996). Finally, in both the one- and two-syllable Thai word lists, the proportion of words carrying each of the five tones was as follows: 40% with mid-tone, 20% low, 20% falling, 10% rising and 10% high, which accords with their relative occurrence in the language at large (Gandour & Gandour, 1982). The complete lists of English and Thai words are available in the Online Supplementary Materials.

To maintain infants' attention during the task, an image of a smiling female face and a blinking red light under the image were presented as visual stimuli on both the left and the right monitors. The decision to employ this type of visual stimulus was taken to ensure a close replication of the Burnham et al. (1999) procedure. Given that our study included infants from two different language and cultural backgrounds, two sets of visual stimuli were used: a Caucasian face was presented for English-language infants and a South-East Asian face for the Thai-language infants. Only static images were used. While the language of the spoken words differed according to presentation side, the face on the two screens was the same on both sides. The face presentation was identical across the familiarisation and test phases.

The visual stimuli were presented on three 20-inch computer monitors placed adjacent to each other to the center, right and left of the infant's sagittal plane, and auditory stimuli were presented via loudspeakers hidden behind the right and the left monitor. English stimuli were presented through one of the audio speakers, and Thai stimuli through the other, with the side of presentation for each language maintained throughout each infant's session but counterbalanced between infants. A centre monitor was used to present an attention getter (a white and blue circle expanding and contracting on the screen in silence) to gain and re-gain

infants' attention and fixation to the centre before the start of each trial. A custom-built Matlab-based experimental software was used for stimulus presentation. During each trial (in the familiarisation and test phases), the software randomly selected words from each list and presented them consecutively with an inter-stimulus-interval of 500ms until the trial ended. These word lists were automatically retrieved by the software prior to the start of each testing session, thus avoiding potential delays related to the retrieval of individual audio files. All infants completed two familiarisation trials and six test trials. The maximum duration of each trial was approximately 25 seconds (as words were selected individually by the software, the exact duration of each trial could vary slightly to avoid presentation of incomplete words).

Infants sat on their caregiver's lap facing the monitors, approximately 60 cm from the centre monitor. Caregivers listened to masking sounds presented over noise-cancelling headphones and were instructed not to speak to their infant and not to direct their attention to the screen by pointing. A CCTV camera was used to stream the live image of the infant's face to the experimenter who sat in an adjoining room and recorded the direction of the infant's gaze by pressing the arrow keys (left, right, and center) on a computer keyboard.

Procedure

A preferential listening procedure, incorporating a familiarisation and a test phase, was used. Stimuli used in each phase were identical and consisted of word lists in English and Thai described in Stimuli above. The familiarisation phase had a fixed number of two trials with fixed duration. Infants first listened to one language (Trial 1) and then to the other language (Trial 2) to ensure that they were familiarised with the location of the set up corresponding to each language. The order of native and non-native language presentation was counterbalanced between infants in each group. The test phase had a fixed number of six trials, and their duration was infant-controlled. On each trial, the infant's gaze to the left or right screen within trials determined the language being played.

Familiarisation phase: This phase served to familiarise infants with the side-language pairing, left-Thai/right-English or left-English/right-Thai, counterbalanced across infants. Infants were first presented with the attention getter on the centre screen. Once the infant had fixated the centre for two seconds, the visual stimulus (face) appeared on one of the side screens, and the audio presentation of the language assigned to that side began to play. The experimenter recorded infants' gaze duration to the screen, but presentation of auditory and visual stimuli continued without interruption for 25 seconds regardless of whether the infant looked or did not look towards the screen. When the familiarisation trial corresponding to the first language was completed, infants saw the attention getter on the centre monitor, and this procedure was repeated for the other language on the other side.

Test phase: After completing the two familiarisation trials, infants immediately proceeded to the test phase. At the start of each test trial, the identical visual stimulus (smiling face) appeared on both the right and left screens in silence. When the infant directed their gaze to either screen, the experimenter (blind to the side-language pairings), pressed the corresponding button on the keyboard, and the auditory stimuli assigned to that side began to play. If the infant switched their gaze to the other screen within the trial, the experimenter pressed the other button on the keyboard, triggering the presentation of the other language. The switch in the side of presentation followed the experimenter's button press immediately, with any minor delays being due to the time required to finish the presentation of the final word from the preceding word list. The trial ended when it reached the maximum duration of 25 seconds of stimuli presentation or when the infant did not look at either screen for more than two seconds. There were six test trials, each with a maximum duration of 25 seconds. Infants' looking duration to each side on each trial was recorded.

Given that stimulus presentation was fully controlled by the infants, in principle, infants could have shown a predominantly absolute preference for one side on each test trial,

meaning that they would have little or no opportunity to hear the stimuli from the other side after the familiarisation phase. However, this was not the case. Out of a total of 414 trials completed across infants and conditions, only 21 trials had looks to a single location (5.07%). On average, infants' looking duration across locations was 17.15 seconds out of the possible 25, and the average looking durations were 7.82 seconds (*SD* = 4.75) to the screen corresponding to the native stimuli and 9.33 seconds (*SD* = 5.25) to the screen corresponding to the non-native stimuli. Therefore, while this task is not designed to maintain infants' identical exposure durations to each language, it was not the case that infants failed to receive sufficient exposure to one of the stimulus languages during the test phase in this task.

Results

Familiarisation Phase

Ahead of the analysis of test trials, looking times recorded during the familiarisation phase were compared across the four infant groups (2 Language groups × 2 Conditions). Mean looking duration to each side and comparisons between sides are presented in Table 1. As can be seen, infants attended to the stimuli during this phase. Importantly, looking times to the native and non-native languages did not differ for either group of infants indicating that they had an equal opportunity to learn the location for each language during familiarisation.

Table 1.

Mean (SD) looking duration (seconds) to the native and non-native side during familiarisation trials by English-learning and Thai-learning infants in the one- and two-syllable conditions, and results of paired-samples t-tests comparing looking duration to the native and non-native side.

		Native side	Non-Native side	*t* (*p*)
English-learning	One-syllable	18.62 (4.54)	19.37 (4.27)	-0.497 (.622)
	Two-syllable	19.45 (3.48)	18.21 (2.89)	1.167 (.251)

Thai-learning	One-syllable	19.29 (3.84)	17.12 (5.18)	1.390 (.172)
	Two-syllable	17.05 (5.03)	15.47 (5.27)	0.890 (.379)

Test Phase

Scatterplots and mean trajectories over age for English and for Thai words are shown in Figure 1 separately for the four groups (2 Language groups × 2 Conditions). Data were analyzed using Linear Mixed Effects (LME) models conducted using the lmer package in R (Bates, 2005). The lmerTest package was used to calculate *p*-values and to conduct pairwise comparisons as required (Kuznetsova, Brockhoff, & Christensen, 2015). The model was specified with infants' looking duration in seconds as the dependent variable, Language Group (English-learning, Thai-learning), Condition (one-syllable, two-syllable), Stimulus Language (English, Thai), and Age (months) as the predictor variables. The two continuous variables (Age and Looking Duration) were scaled and centered around zero to assist with model convergence. The initial model was specified with the maximum random effects structure (Barr, Levy, Scheepers, & Tilly, 2013), but random effects were pruned due to failure to converge until only random intercept for participants and random slopes for age were maintained. In addition to the fixed effects, the model also included the following interactions aimed to test our predictions: Group × Stimulus Language, Group × Condition × Stimulus Language, and Group × Condition × Stimulus Language × Age. A simulation-based power analysis conducted using the simr function in R (Green & Macleod, 2016) confirmed the suitability of our sample size for this model, yielding an observed power of 80.3% (CI: 77.7, 82.72%).

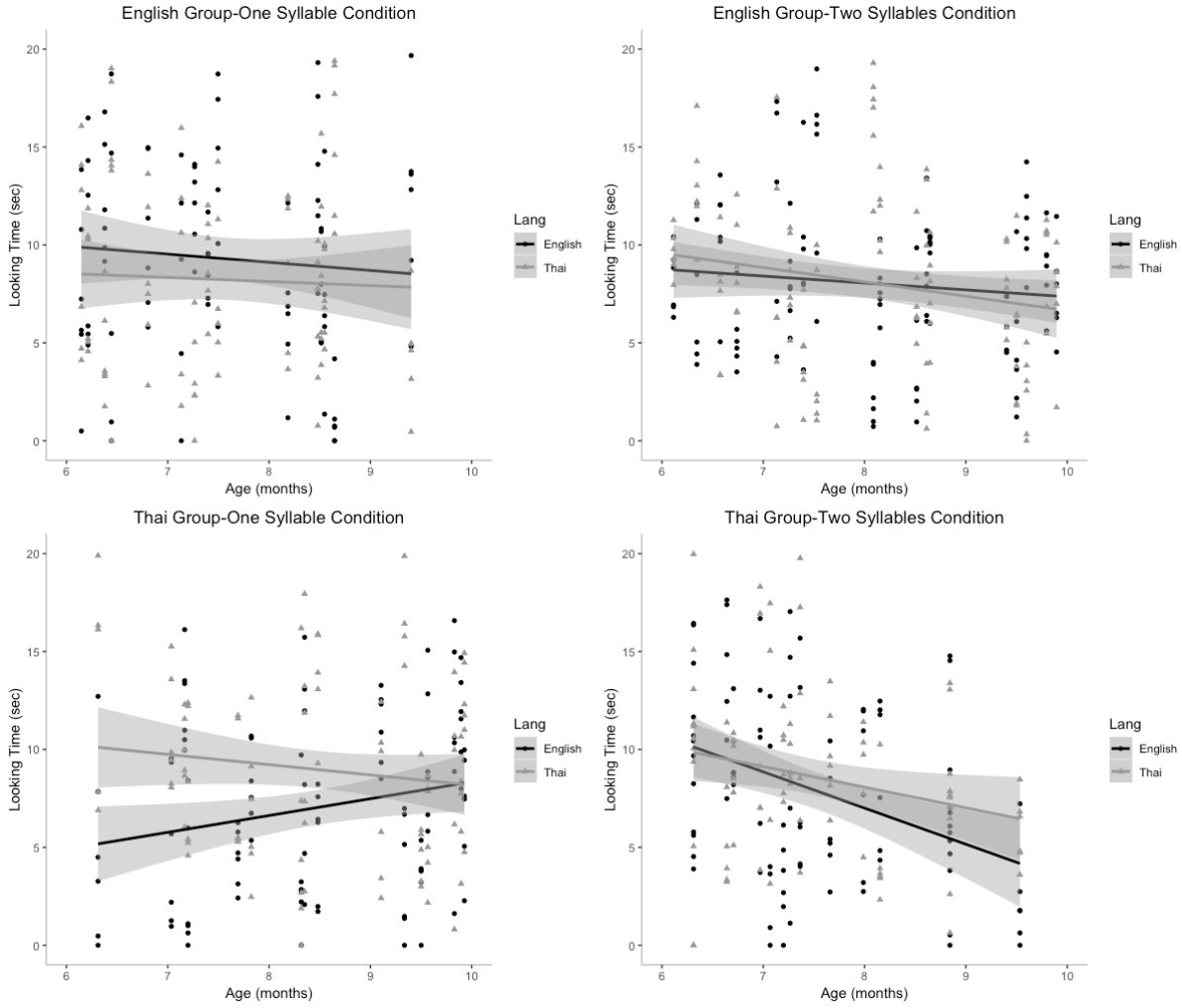

Figure 1. English and Thai infants' performance in the one- and two-syllable conditions of the visual preference task (lines represent the modeled regression line, and shaded areas represent the 95% Confidence Intervals).

The results of the ANOVA on the model are presented in Table 2 and the detailed model output is presented in Table A1 in the Appendix. As can be seen in Table 2, the model yielded no significant main effects, but two significant interactions: Group × Stimulus Language and Group × Condition × Stimulus Language. As can be seen in Figure 1, infants produced longer looking times to their native-language side in the one-syllable condition but looking times to the two languages were more similar in the two-syllable condition. In order

to investigate the source of these interactions two additional LME models were constructed assessing performance separately in the English- and Thai-learning groups.

Table 2.

ANOVA results for the Linear Mixed Effects model assessing the effects of Group, Language, Condition, and Age on English and Thai infants' performance in the visual preference task.

	F	*df(res)*	*p*
Group	0.074	57.51	.787
Condition	0.117	57.56	.733
Language	2.117	751	.146
Age	2.839	32.32	.102
Group × Language	**20.822**	**751**	**<.001**
Group × Condition	2.292	58.09	.135
Condition × Language	0.011	751	.917
Group × Age	0.076	32.22	.784
Condition × Age	1.492	32.28	.231
Language × Age	2.643	751	.104
Group × Condition × Language	**7.576**	**751**	**.006**
Group × Condition × Age	0.101	32.25	.753
Group × Language × Age	0.046	751	.831
Condition × Language × Age	0.604	751	.437
Group × Condition × Language × Age	0.173	751	.678

These models were identical to the initial model with the exception that Group was no longer specified as an independent variable. The results of the ANOVAs on the models are presented in Table 3 and the detailed model output is presented in Tables A2 and A3 in the Appendix. The model for the English-learning infants' data yielded a main effect of Stimulus Language only, showing that English-learning infants directed longer listening times to the English than the Thai stimuli across the two conditions, one- and two-syllable words, β = 1.034, *SE* = .483, *CI*[.087, 1.982], $t(412) = 2.145$, $p = .033$. Inspection of Figure 1 suggests that the effect of language may be driven by infants in the one-syllable condition, but the Stimulus Language × Condition interaction did not reach statistical significance.

The model for the Thai-learning infants' data also yielded a main effect of Stimulus Language, with Thai infants listening longer to Thai than to English stimuli, β = -2.149, *SE* = .498, *CI*[-3.120, -4.319], $t(371.7) = -1.458$, $p < .001$. This main effect was qualified by a Condition × Stimulus Language interaction. Follow-up pairwise comparisons revealed that looking times for Thai over English in this group were significantly longer in the one-syllable condition, β = -3.173, *SE* = .706, *CI*[-4.561, -1.785], $t(371.7) = -4.495$, $p < .001$, but not in the two-syllable condition, β = -1.125, *SE* = .701, *CI*[-2.504, -.254], $t(371.7) = -1.604$, $p = .110$. Therefore, Thai infants showed a significant preference for Thai overall, but this effect was driven by the infants in the one-syllable condition.

Table 3.

ANOVA results for the Linear Mixed Effects models conducted separately for the English- and Thai-learning infants to assess the effects of Language, Condition, and Age on infants' performance in the visual preference task.

English-learning infants			
	F	*df(res)*	*p*
Condition	2.409	30.67	.131
Language	**4.330**	**381**	**.038**
Age	2.318	14.31	.150
Condition × Language	3.100	381	.079
Condition × Age	.915	13.95	.355
Language ×Age	1.766	381	.185
Condition ×Language × Age	.068	381	.794
Thai-learning infants			
	F	*df(res)*	*p*
Condition	0.433	28.4	.516
Language	**18.062**	**370**	**<.001**
Age	.839	15.22	.374
Condition × Language	**4.158**	**370**	**.042**
Condition × Age	.685	15.22	.421
Language ×Age	.942	370	.333
Condition ×Language × Age	.725	370	.395

Discussion

This study investigated English- and Thai-learning infants' early preferences for their native language based on lexical tone and word stress cues. Overall, English and Thai infants showed preferences for their native language in the preferential listening task as indicated by the main effect of Stimulus Language for both the Thai and English groups. Therefore, before their first birthday, infants demonstrate sensitivity to lexical tone information. For Thai infants this is evident in their native language preference for the tone-bearing Thai words in the one-syllable condition, and for English language infants it is evident in their preference for the absence of tone in the non-tone-bearing English words in the one- and two-syllable conditions. However, despite the continued presence and absence of tone and the added distinctive difference in word-stress in the two-syllable words, there was no native language preference for two-syllable words by Thai-learning infants.

In this task, infants had a variety of phonetic and phonotactic cues at their disposal to differentiate the word lists that did and did not belong to their language. English and Thai words differ in their phonetic and phonotactic properties, and these differences were accurately represented in our stimuli (i.e., not all phonemes that appeared in each language's word list belonged to the phonemic inventory of the other language). However, based on previous research findings it is unlikely that infants relied on these phonetic and phonotactic cues in our language differentiation task because previous studies suggest that the ability to use phonetic and phonotactic properties of words in a native language differentiation task does not emerge until around nine to 11 months (Jusczyk et al., 1993; Jusczyk & Luce, 1994; Mattys et al., 1999), so it would appear that the majority of infants in this study would be too young to differentiate English and Thai words based on their vowel and consonant segments. Nevertheless, while we acknowledge that we are unable to entirely discard the possibility that infants relied on these cues to some extent since the process of perceptual attunement begins

early in the first year (Werker, 2018), we maintain that phonetic and phonotactic information were not sufficient to support infants' language differentiation. If it were, we would have observed successful language differentiation in *both* the one- and two-syllable conditions of our task in *both* language groups, which was not the case here.

Therefore, it is most likely that infants differentiated English and Thai words by relying on the pitch-based information that marks the most striking difference between the two languages – lexical tones. To understand how infants employed this information, we consider the results in each condition separately.

In the one-syllable condition, no word-stress information was available, so Thai and English words were primarily differentiated by the presence of lexical tones in the former and their absence in the latter language. Results for the two language groups are considered in turn.

It is not surprising that Thai infants detected the presence of lexical tones in the one-syllable words of their native language for it has been found that tone-language learning infants show an early sensitivity to lexical tone categories as early as four months of age (Yeung et al., 2013). In addition, tone language learning infants receive ample, even exaggerated, exposure to lexical tones in their early language input. Similar to the acoustic exaggeration reported for vowels in infant-directed speech in non-tone languages (Kuhl et al., 1997; Burnham, Kitamura, & Volmer-Conna, 2002) and tone languages (Liu, Kuhl, & Tsao, 2003), mothers exaggerate lexical tones in speech to tone-language learning infants (Liu, Tsao, & Kuhl, 2007; Xu Rattanasone, Burnham, & Riley, 2013).

It *is* surprising, however, that English language infants showed a native language preference for one syllable words, a preference based on the *absence* of lexical tones. While it may be difficult to envisage a preference based on the absence of a cue, it must be remembered that in this task infants could instantaneously compare and contrast the words of

the two languages by simply turning their head, so the presence of pitch variation in one case and its absence in another should have been readily apparent. If so, then it is unclear why there is a native (no tone) preference, for we know that infant-directed speech contains exaggerated pitch contours compared with adult-directed speech (Kitamura & Burnham, 2003), and that it is this pitch variation that drives infants' preference for infant-directed speech (Fernald & Kuhl, 1987). Why then did infants prefer the less pitch-variable English words? We propose that when given the opportunity to compare the words directly, infants in this study demonstrated sensitivity to and preference for the pitch contours typical of their native language, i.e., that their native language preference overrode any effect of acoustic salience.

Another unexpected finding is that age did not modulate infants' performance in the one-syllable condition. We predicted a significant increase in native language preference across our age range specifically by infants acquiring a tone language, i.e., the Thai language group. This stronger sensitivity to lexical tone information by tone-language learning infants in this task was predicted given the close relation between native language preferences and the process of phonological attunement, and the demonstrated increase in sensitivity to native lexical tone categories during tone language infants' second semester of life (Singh et al., 2018; Singh et al., 2016; Tsao, 2017), but this prediction was not supported. Instead, our findings align with the growing evidence that perceptual attunement to lexical tones follows a more complex trajectory compared to consonants and vowels (Singh & Foong, 2016). The beginnings of attunement to native tone categories in infants acquiring tone languages has been observed before six months of age (Yeung et al., 2013), which could explain the lack of any continued age-related increase in reliance on tone cues in the one-syllable condition in our six to 10-month-old Thai-learning infants. On the other hand, it has been found that non-tone language infants maintain sensitivity to lexical tones in lexical processing tasks up to 18

months of age (Hay et al., 2019; Hay et al., 2015; Singh et al., 2014), which may account for the maintained sensitivity to lexical tone cues (their presence in the Thai stimuli and absence in the English stimuli) in the one-syllable condition by the English language infants across our age range.

In the two-syllable condition, words contained both the syllable-level and word-level pitch cues signalling each language's use or non-use of lexical tone and their dominant stress pattern. In this two-syllable condition, only English-learning infants demonstrated a language preference; there was no such preference for Thai-learning infants. Results for the two language groups are considered in turn.

English language infants performed as predicted in the two-syllable condition; they preferred the English word lists. In previous studies using non-tone language stimuli, non-tone language infants have been shown to differentiate native from non-native language words based on stress-patterns by around five to seven months of age (Höhle, et al., 2009; Friederici et al., 2007; Weber et al., 2004), so we infer that a similar reliance on native word stress cues would be operational here. However, unlike our results in the one-syllable condition in which English-learning infants differentiated English from Thai based primarily on lexical tone cues, here, as the word lists differed on both tone/non-tone and weak-strong versus strong-weak stress patterns, we are unable to determine the relative degree to which infants relied on lexical tone and/or word stress cues.

Turning to the Thai language infants, contrary to our predictions, they showed no native language preference when presented with the two-syllable words. On the basis of the results in the one-syllable condition, Thai infants would have been expected to use the absence (in the English words) and presence (in the Thai words) of lexical tone. On the basis of previous studies, they would have also been expected to use the word-stress cue (Höhle, et al., 2009; Friederici et al., 2007; Weber et al., 2004). They used neither. It appears that Thai

language infants' expected ability to differentiate words on the basis of word-stress pitch cues is impaired by the presence of lexical tone pitch cues, and that their ability to differentiate words on the basis of lexical tone pitch cues (see the one-syllable condition here) is impaired by the presence of word-stress pitch cues. As each cue can be used separately, it appears that when two pitch-based cues co-occur in the signal in this simultaneous preference task, Thai language infants' differentiation abilities are impaired.

It is noteworthy that while there were no statistically significant differences between English infants' performance in the one-syllable and two-syllable conditions, the size of the preference for two-syllable native words was significantly reduced compared to the one-syllable words (see Figure 1 and Table A2 in the Appendix). This performance pattern can be explained by the same mechanism as above for the Thai-language infants; English language infants' language differentiation abilities may well be impacted by the co-occurrence of two pitch-based cues in the signal. This impact may have been weaker for the English learning infants than for Thai learning infants given Thai infants' stronger sensitivity to lexical tones because of their greater experience with lexical tones in their daily language input. This explanation, however, should be considered with caution given the lack of a significant Condition by Stimulus Language interaction in the English group. Therefore, we also acknowledge that it is possible that Thai infants' ability to rely on word stress information follows a different developmental trajectory compared to English infants. There is, as yet, no evidence regarding tone language infants' ability to differentiate languages on the basis on word-stress, either alone or in combination with other cues. Clearly further research is required, starting with studies of tone-language infants' perception of and differentiation on the basis of word-stress.

Our procedures were based on a previous study by Burnham and colleagues with English-learning infants, but we failed to replicate their findings. In an identical paradigm to

this study, Burnham et al. (1999) presented English-learning infants with one- and two-syllable words in English and Thai and found evidence for English-language preference for two-syllable words in seven-month-old, but not six- and nine-month-old infants, and no preference for one-syllable words across the three age groups. At this point, we are unable to identify precisely the source for these divergent findings, but we note several important design differences between our studies. Burnham et al. administered the one- and two-syllable conditions in a within-subjects design, and they reported a significant decrease in infants' attention across the two conditions. Even though the order of the conditions was counterbalanced, performance in the second task may have been unreliable given the reduced attention and possible fatigue. Furthermore, treatment of age as a categorical variable in the previous study, followed by separate analyses for each age group, could have resulted in decreased statistical power to detect a language preference across ages and conditions.

Furthermore, we note that despite our efforts to ensure that experimental set up and procedures were identical across the two labs, we could not discard entirely possible lab-specific effects on infants' performance. For instance, it was not possible to guarantee identical coding precision by the experimenters in the two sites, identical sound levels for stimulus presentation, and other possible differences in the physical lab space. Unfortunately, our design conflates language group and lab, so it is impossible to control for lab effects directly in our statistical analyses. Future replications can shed light on such lab effects, but we maintain that despite these possible limitations, cross-lab experiments offer a unique opportunity for cross-linguistic and cross-cultural comparisons and for increasing the representation and generalisability of infancy research to a wider infant population (see Frank et al., 2017 for a discussion).

Infants developing in non-tone language environments prefer to listen to their native language in the first months of life. The grain-size of information on which they base this

preference is initially quite coarse (prosody, intonation, word-stress) but becomes finer (phonetic, phonotactic) with increasing exposure to their native language. In this study, we have shown that both tone and non-tone language six- to 10-month-old infants use the presence (in Thai words) versus the absence (in English words) of the fine-grain size cue, lexical tone, to differentiate languages. However, as yet, we do not know the developmental trajectory for any native language preference in tone language infants. If young (five to seven months) tone language infants use lexical tone to differentiate languages, then from a structural point of view this would be in accord with studies with non-tone language infants of this age who also use pitch-based cues (prosody, intonation, word-stress) to differentiate languages. However, from a functional point of view this poses an interesting question. Tone language infants may similarly use pitch-based cues to differentiate languages, but over and above the non-phonemic pitch-based cues (prosody, intonation, word-stress), there would be one pitch-based cue, lexical tone, that, for them, is phonemic. The effect of using phonemic versus non-phonemic cues in native language preference early in infancy remains unknown. Therefore, over and above showing that both tone and non-tone language six- to 10-month-old infants use lexical tone cues to differentiate languages, this study opens the door to further studies that will benefit not only our knowledge of language development in tone-language but also in non-tone language environments.

Acknowledgements: The first author's work was supported by the European Union's Horizon 2020 Marie Sklodowska-Curie individual Fellowships European Programme under Grant Agreement No 798908 Optimising IDS, and she receives support from the Basque Government through the BERC 2018-2021 program, and from the Spanish Ministry of Science and Innovation through the Ramon y Cajal Research Fellowship, PID2019-105528GA-I00. We would like to thank Dr Benjawan Kasisopa for her assistance with stimulus selection and experimental set up and Mr Johnson Chen for creating the

experimental software. We also thank Ms Hana Zjakic and Ms Nawasri Chonmahatrakul for their assistance with data collection. We are grateful to all the infants and their parents for their valuable time and interest in this research.

Conflict of interest statement: The authors declare no conflict of interest related to this work.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.

Bates, D. (2005). Fitting linear mixed models in R. *R News, 5*(1), 27–30.

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.

Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition, 65*(1), 33–69. https://doi.org/10.1016/S0010-0277(97)00040-1

Burnham, D., Kitamura, C., & Lancuba, V. (1999). The development of linguistic attention in early infancy: The role of prosodic and phonetic information. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 1197–1200). San Francisco.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435. https://doi.org/10.1126/science.1069587

Burnham, D. & Mattock, K. (2007) The perception of tones and phones. In M.J. Munro & O.-S. Bohn (eds) *Language Experience in Second Language Speech Learning. In honor of James Emil Flege*. Amsterdam: John Benjamins (Series: Language Learning and Language Teaching). Chapter 15, pp 259-280.

Burnham, D., Kim, J., Davis, C., Ciocca, V., Schoknecht, C., Kasisopa, B., & Luksaneeyanawin, S. (2011). Are tones phones? *Journal of Experimental Child Psychology, 108*(4), 693-712.

Cabrera, L., Tsao, F. M., Liu, H. M., Li, L. Y., Hu, Y. H., Lorenzi, C., & Bertoncini, J. (2015). The perception of speech modulation cues in lexical tones is guided by early language specific experience. *Frontiers in Psychology, 6*(AUG), 1–14. https://doi.org/10.3389/fpsyg.2015.01290

Chen, A., & Kager, R. (2016). Discrimination of lexical tones in the first year of life. *Infant and Child Development, 25*, 426–439. https://doi.org/10.1002/icd.1944

Chen, A., Stevens, C. J., & Kager, R. (2017). Pitch perception in the first year of life , a comparison of lexical tones and musical pitch. *Frontiers in Psychology, 8*(March), 1–8. https://doi.org/10.3389/fpsyg.2017.00297

Colombo, J., & Bundy, R. S. (1983). Infant response to auditory familiarity and novelty. *Infant Behavior and Development, 6*(2–3), 305–311. https://doi.org/10.1016/S0163-6383(83)80039-3

Cutler, A., & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Dunst, C., Gorman, E., & Hamby, D. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning, 5*(1), 1-13.

Fernald, H., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherse speech. *Infant Behaviour and Development, 10*, 279–293.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., ... & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421-435.

Friederici, A. D., Friedrich, M., & Christophe, A. (2007). Report brain responses in 4-month-old infants are already language specific. *Current Biology, 17*, 1208–1211. https://doi.org/10.1016/j.cub.2007.06.011

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

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

Gandour, J., & Candour, M. J. (1982). The relative frequency of tunes in Thai. *Papers in South East Asian Linguistics*, 9, 155.

Green, P., & Macleod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Götz, A., Yeung, H. H., Krasotkina, A., Schwarzer, G., & Höhle, B. (2018). Perceptual reorganization of lexical tones: effects of age and experimental procedure. *Frontiers in Psychology, 9*, 477.

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child Development, 86*(1), 10–22. https://doi.org/10.1111/cdev.12269

Hay, J. F., Cannistraci, R. A., & Zhao, Q. (2019). Mapping non-native pitch contours to meaning: Perceptual and experiential factors. *Journal of Memory and Language, 105*, 131-140.

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 and Development*, *32*(3), 262-274.

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 37-58.

Jusczyk, P. W. (1997). Finding and remembering words: Some beginnings by English-learning infants. *Current Directions in Psychological Science, 6*(6), 170–173.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675–687. Retrieved from http://www.jstor.org/stable/1131210?seq=1#page_scan_tab_contents

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants′ sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*(3), 402–420. https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., & Luce, P. A. (1994). Infants′ sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*(5), 630–645. https://doi.org/10.1006/jmla.1994.1030

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy, 4*(1), 85–110. https://doi.org/10.1207/S15327078IN0401_5

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews. Neuroscience, 5*(11), 831–843. https://doi.org/10.1038/nrn1533

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.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V, Ryskina, V. L., … Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package "lmerTest." R Package Version, 2(0).

Liu, L., & Kager, R. (2014b). Perception of tones by infants learning a non-tone language. *Cognition, 133*, 385–394.

Liu, L., & Kager, R. (2018). Monolingual and bilingual infants' ability to use non-native tone for word learning deteriorates by the second year after birth. *Frontiers in Psychology, 9*(March), 1–12. https://doi.org/10.3389/fpsyg.2018.00117

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), 1–10. https://doi.org/10.1111/1467-7687.00275

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology, 43*(4), 912.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241–265.

Mattock, Karen, 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. https://doi.org/10.1016/j.cognition.2007.07.002

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and Prosodic Effects on Word Segmentation in Infants. *Cognitive Psychology, 494*, 465–494.

Molnar, M., Gervain, J., & Carreiras, M. (2014). Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. *Infancy, 19*(3), 326–337. https://doi.org/10.1111/infa.12041

Nazzi, T, Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology. Human Perception and Performance, 24*(3), 756–766. https://doi.org/10.1037/0096-1523.24.3.756

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*(1), 1-19.

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

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception & Performance, 19*, 421–435.

Potisuk, S., Gandour, J., & Harper, M. P. (1996). Acoustic correlates of stress in Thai. *Phonetica, 53*(4), 200-220.

Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic , rhythmic , and intonational cues Franck Ramus. *Annual Review of Language Acquisition,* 2, 1–14.

Shi, R., Santos, E., Gao, J., & Li, A. (2017). Perception of similar and dissimilar lexical tones by non-tone-learning infants. *Infancy, 22*(6), 790–800. https://doi.org/10.1111/infa.12191

Singh, L., & Fu, C. S. L. (2016). A new view of language development: The acquisition of lexical tone. *Child Development, 87*(3), 834-854. https://doi.org/10.1111/cdev.12512

Singh, L., Fu, C. S. L., Seet, X. H., Tong, A. P. Y., Wang, J. L., & Best, C. T. (2018). Developmental change in tone perception in Mandarin monolingual, English monolingual, and Mandarin–English bilingual infants: Divergences between monolingual and bilingual learners. *Journal of Experimental Child Psychology, 173*, 59–77. https://doi.org/10.1016/j.jecp.2018.03.012

Singh, L., Poh, F. L. S., & Fu, C. S. L. (2016). Limits on monolingualism? A Comparison of monolingual and bilingual infants' abilities to integrate lexical tone in novel word learning. *Frontiers in Psychology, 7*(MAY), 1–16.

Singh, L., Tam, H. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science, 17*(1), 94–109. https://doi.org/10.1111/desc.12097

Tsao, F. (2017). Perceptual Improvement of Lexical Tones in Infants : Effects of Tone Language Experience. *Frontiers in Psychology, 8*(April), 1–14. https://doi.org/10.3389/fpsyg.2017.00558

Weber, C., Hahne, A., Friedrich, M., & Friederici, A. D. (2004). Discrimination of word stress in early infant perception: electrophysiological evidence. *Cognitive Brain Research, 18*(2), 149–161. https://doi.org/10.1016/j.cogbrainres.2003.10.001

Werker, J. F. (2018). Perceptual beginnings to language acquisition. *Applied Psycholinguistics, 39*(4), 703–728. https://doi.org/10.1017/S0142716418000152

Werker, J. F., & Hensch, T. K. (2015). Critical Periods in Speech Perception: New Directions. *Annual Review of Psychology, 66*, 173–196. https://doi.org/10.1146/annurev-psych-010814-015104

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 349-355.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech-perception - evidence for perceptual reorganization during the 1st year of life. *Infant Behavior and Development, 7*(1), 49–63. https://doi.org/Doi 10.1016/S0163-6383(84)80022-3

Werker, J. F., & Tees, R. C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology, 46*(3), 233–251.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*(1), 25-47.

Xu Rattanasone, N., Burnham, D., & Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *Journal of Phonetics, 41*(5), 332-343.

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language, 68*(2), 123–139. https://doi.org/10.1016/j.jml.2012.09.004

Yip, M. (2002). *Tone*. Cambridge University Press.

Appendix

Table A1.

Detailed output for the Linear Mixed Effects model assessing the effects of Group, Language, Condition, and Age on English and Thai infants' performance in the visual preference task.

	ß	*SE*	*df*	*t*	*p*
(Intercept)	0.280	0.111	133.15	2.519	.013
Group (Thai-learning)	-0.671	0.158	165.19	-4.238	.001
Condition (2-syl)	-0.338	0.153	141.49	-2.215	.028
Language (Thai)	-0.388	0.138	757.19	-2.817	.005
Age	0.030	0.110	167.37	0.275	.784
Group (Thai-learning) × Language (Thai)	1.035	0.201	757.19	5.147	.001
Group (Thai-learning) × Condition (2-syl)	0.627	0.220	153.74	2.851	.005
Condition (2-syl) × Language (Thai)	0.341	0.190	757.19	1.790	.074
Group (Thai-learning) × Age	0.070	0.160	168.24	0.439	.661
Condition (2-syl) × Age	-0.111	0.155	180.74	-0.716	.475
Language (Thai) × Age	-0.155	0.136	757.19	-1.14	.255
Group (Thai-learning) × Condition (2-syl) × Language (Thai)	-0.763	0.277	757.19	-2.758	.006
Group (Thai-learning) × Condition (2-syl) × Age	-0.117	0.224	155.61	-0.520	.604
Group (Thai-learning) × Language (Thai) × Age	-0.030	0.199	757.19	-0.149	.882
Condition (2-syl) × Language (Thai) × Age	0.051	0.194	757.19	0.263	.793
Group (Thai-learning) × Condition (2-syl) × Language (Thai) × Age	0.115	0.277	757.19	0.416	.678

Table A2.

Detailed output for the Linear Mixed Effects model for the English-learning infants to assess the effects of Language, Condition, and Age performance in the visual preference task.

	ß	*SE*	*df*	*t*	*p*
(Intercept)	10.026	0.496	412.00	20.222	.001
Condition (2-syl)	-1.744	0.686	412.00	-2.544	.011
Language (Thai)	-1.962	0.701	412.00	-2.798	.005
Age	0.246	0.489	412.00	0.504	.614
Condition (2-syl) × Language (Thai)	1.724	0.969	412.00	1.778	.076
Condition (2-syl) × Age	-0.656	0.698	412.00	-0.939	.348
Language (Thai) × Age	-0.782	0.691	412.00	-1.132	.258
Condition (2-syl) × Language (Thai) × Age	0.258	0.987	412.00	0.261	.794

Table A3.

Detailed output for the Linear Mixed Effects model for the Thai-learning infants to assess the effects of Language, Condition, and Age performance in the visual preference task.

	ß	*SE*	*df*	*t*	*p*
(Intercept)	6.617	0.618	73.58	10.709	.001
Condition (2-syl)	1.430	0.882	62.96	1.623	.110
Language (Thai)	3.270	0.731	371.66	4.471	.001
Age	0.429	0.647	70.01	0.662	.510
Condition (2-syl) × Language (Thai)	-2.136	1.003	371.66	-2.13	.034
Condition (2-syl) × Age	-1.128	0.905	59.42	-1.246	.218
Language (Thai) × Age	-0.932	0.726	371.66	-1.284	.200
Condition (2-syl) × Language (Thai) × Age	0.840	0.986	371.66	0.851	.395