11 Perceptual development of phonetic categories in early infancy

Consonants, vowels, and lexical tones

Jun Gao and Rushen Shi

11.1 Introductory remark

One fundamental issue in language acquisition research concerns input-guided learning versus input-independent capacities in children. On the one hand, researchers strive to understand the way native language input shapes acquisition. On the other hand, there is a strong interest in determining how language acquisition may be affected by children’s natural capacities (independent of the specific ambient language) such as those present at birth. These questions apply to various levels of linguistic representations such as syntax, phonology and phonetics. In this chapter, we discuss key empirical findings in early phonetic development that shed light on these questions, and we report our recent experiments on infants’ perception of lexical tones during the first year of life.

11.2 Perceptual development of consonants and vowels in infants

Research in phonetic and phonological acquisition has contributed valuable empirical results on the effect of input versus children’s natural capacities. Perceptual studies with neonates and infants during the first year of life are directly pertinent. Most studies have concentrated on the perceptual development of native and non-native consonants and vowels. It has been demonstrated that infants are born with the natural capacity to perceive many phonetic contrasts, both native and non-native ones, and that their perception is gradually influenced by the sound structure of the native language during the course of the first year of life. This was shown in the classic work of Werker and colleagues (Werker et al. 1981; Werker and Tees 1984). They presented participants with consonantal contrasts in Hindi and Salish (including a Hindi retroflex-dental contrast and a Salish velar-uvular contrast, both absent in English). They found that six- to eight-month-old English-learning infants discriminated the non-English contrasts, but their discrimination declined by ten to twelve months of age. Adult English speakers also failed to discriminate the contrasts. Hindi- and Salish-learning infants, however, maintained their discrimination of their respective native contrast at ten to twelve months of age. A similar pattern of perceptual development was found for vowels. In Polka and Werker (1994), English-learning infants discriminated a German front rounded versus back rounded vowel contrast at four months of age, and the discrimination deteriorated after six months of age.

These findings suggest that infants begin acquisition with the language-general ability to perceive phonetic contrasts, and that as they begin to acquire the phonological system of the ambient language, native contrasts are maintained and the non-native ones become attenuated in perception and representation. Indeed, during the second half of the first year of life, infants begin to learn various aspects of their native phonology. Infants start representing the internal structure of native vowels, such that they respond differently to prototypical and non-prototypical tokens of a native vowel category (Kuhl 1991; Kuhl et al. 1992). Between six and nine months of age infants develop sensitivity to the phonotactic regularities (e.g., Mattys and Jusczyk 2001) and stress patterns (e.g., Jusczyk, Cutler, and Redanz 1993) of their native language. The narrowing of perception of phonetic contrasts is coherent with infants' focus on the native language phonological structures, and the experience with the ambient speech input thus influences the evolving perceptual patterns for native versus non-native sounds.

Later studies revealed a more complex picture of perceptual development of consonantal and vowel contrasts in infants. Research showed that the discrimination of native contrasts is not always about the maintenance of discrimination from early infancy. For certain contrasts, there is gradual improvement over age in infants' discrimination of native contrasts. Kuhl and colleagues (2006) found that English-learning infants can discriminate the English /r/-/l/ contrast at six to eight months of age, and importantly, they improve significantly in their discrimination of this contrast between six and twelve months of age. Likewise, Mandarin-Chinese-learning infants' discrimination of a Mandarin-Chinese affricate-fricative contrast enhances from six to twelve months of age (Tsao, Liu, and Kuhl 2006). Japanese-learning infants in Kuhl et al. (2006) and English-learning infants in Tsao, Liu, and Kuhl (2006) declined in their discrimination of those non-native contrasts during the same age period. Thus, while the lack of input leads to perceptual decline in non-native infants, continued input exposure leads to discrimination improvement in native-language infants. The ability to discriminate certain contrasts is not fully in place at birth. Facilitative learning occurs during the first year of life as infants gain experience with the native-language input.

Further variability has been observed with respect to listeners' natural capacities and the effect of input for phonetic perception. For example, the discrimination level of the English voiced stop versus fricative d-th (/d/–/ð/) distinction stays unchanged and is equivalent for both English-learning and French-learning infants throughout the first year of life (even though the contrast is present in English but absent in French), and significant improvement was observed from age one to adulthood in English listeners only (Polka

Colantonio, and Sundara 2001), indicating a delayed effect of input on learning. There are also cases in which the discrimination of certain speech sounds is absent at birth, and infants must rely entirely on phonetic learning from the input (Narayan, Werker, and Beddor 2010). In Narayan et al. (2010) there was no discrimination of the Filipino syllable-initial alveolar- velar nasals, which are acoustically similar, in Filipino- and English-learning infants during early infancy. Filipino infants eventually learned to discriminate this contrast by ten to twelve months of age, whereas English-learning infants across ages consistently failed to make the discrimination. Narayan et al. (2010) interpreted their findings in terms of acoustic salience. That is, the acoustic cues to the contrasting nasal consonants are too weak. In this sense, the innate language-general perceptual ability shown in previous studies (e.g., Werker and Tees 1984; Polka and Werker 1994) appears to require certain basic acoustic saliency.

On the other hand, certain non-native contrasts are well discriminated from early infancy to adulthood despite missing experience, as in the case of English infants' and adults' discrimination of Zulu clicks (Best, McRobert, and Sithole 1988). Based on the perceptual assimilation model (PAM; Best, 1995), the discrimination of non-native contrasts is related to whether the sounds are assimilable to native phonetic categories and how they are assimilated to the native categories. According to this model, Zulu clicks remain discriminable to English listeners because they are non- assimilable to any English phonemic categories. It is possible that the clicks were perceived as non-speech sounds by the English listeners, and that the general auditory system was sensitive to their acoustic differences in a non-categorical fashion.

Variable results have also been reported for vowels. Polka and Bohn (1996) found that English and German adults discriminated both an English vowel contrast (dat-det) and a German vowel contrast (d/u/t-d/y/ t), even though the non-native contrasts are absent in their respective native languages. Furthermore, six- to eight-month-old and ten- to twelve-month-old English- and German-learning infants showed comparable discrimination of these native and non-native contrasts, and there was no difference in performance across those ages. Therefore, whereas the early discrim- ination reflects infants' language-general natural perceptual ability, the basis for the continued discrimination of those non-native vowel contrasts during later infancy and adulthood is unclear. It may be a manifestation of the innate natural perceptual capacities that persist. Alternatively, it may be due to listeners' perceptual assimilation of those non-native contrasts to their nearest native vowel contrasts, consistent with the view of PAM (Best 1995).

In sum, research on perceptual development of consonants and vowels revealed evidence supporting both input-independent natural perceptual capacities and input-guided phonetic learning. Both mechanisms exert effects during the course of acquisition and are contrast-dependent.

11.3 Perceptual development of lexical tones in infants

Phonemic inventories in natural languages not only include consonants and vowels but also suprasegmental categories such lexical tones. Many world languages (for example, in Asia) contain tonal contrasts for distinguishing word meaning. For example, in Mandarin *ma1* and *ma3* are minimal pairs contrasting in tones (Tone 1: high-level versus Tone 3: low-dipping) and denote different meanings (*ma1* “mother” versus *ma3* “horse”). The typical acoustic correlate for tones is the fundamental frequency (i.e., pitch) of the tone-bearing unit (usually the vowel or the syllable), although other acoustic properties such as the duration and amplitude of the tone-bearing unit may also cue tonal distinctions.

Relative to the abundant literature on early perceptual development of consonants and vowels, fewer studies have investigated infants’ perception of lexical tones. The study of lexical tones is relevant for the issue of input-driven learning versus input-independent natural capacities in the acquisition of phonetic categories, as it is interesting to know if the acquisition of lexical tones is governed by the same mechanisms as those that underlie the acquisition of consonants and vowels.

The published studies so far have yielded variable results. Mattock and colleagues (Mattock and Burnham 2006; Mattock et al. 2008) reported a similar developmental trajectory in infants’ perception of lexical tones as shown for consonants (Werker and Tees 1984) and vowels (Polka and Werker 1994). In their experiments, English- and French-learning infants discriminated the Thai low-level versus rise contrast at four and six months of age, but failed to do so at nine months of age, suggesting that infants were universal listeners of lexical tones early in life, and that the lack of tonal contrasts in English led to the decline of tonal discrimination at nine months of age. Infants who were Cantonese- and Mandarin-acquiring continued to discriminate the Thai tonal contrast at nine months of age, presumably because their native languages, which contain the similar tonal contrast, influenced their discrimination of those Thai tones.

Yeung, Chen, and Werker (2013) examined infants’ perception of a contrast in Cantonese that is similar to the contrast in Mattock et al. (2008), mid-level versus rise tones. They compared the performance of non-tone-learning (English), non-native tone-learning (Mandarin), and native-tone-learning (Cantonese) infants. They found evidence of discrimination in four-month-olds of all three language groups, suggesting that infants responded as universal listeners of lexical tones. At nine months, the English-learning infants no longer discriminated the Cantonese tones, whereas the two Chinese groups continued to show evidence of discrimination. Their results, however, are difficult to interpret – the three groups of infants did not always yield the predicted pattern of responses. In particular, a preference for alternating trials (i.e., both tones presented within a trial) over non-alternating trials (the level tone in some trials and the rise tone in other trials) was predicted for successful tonal discrimination. In some of their experimental conditions, infants preferred the alternating trials over trials presenting one of the tones, but not over trials presenting the other tone. For example, Mandarin-learning infants looked longer in alternating trials than in the level-tone trials only. Their looking to the rise-tone trials and alternating-tone trials were comparably high. The English-learning four-month-olds preferred the alternating trials to the rise-tone trials, but their responses to the mid-level-tone trials and alternating-tone trials were similar. Furthermore, among the Chinese infants, one of the familiarization sub-groups (i.e., the group familiarized with the mid-level tone) did not show any discrimination during the test phase. These results seem puzzling. Nevertheless, the overall decline in English-learning infants' discrimination from four to nine months of age is consistent with the results of Mattock et al. (2008).

There is also evidence that non-tone-learning infants' discrimination of certain lexical tones persists throughout the first year of life despite no experience with tones. Liu and Kager (2014) examined the perception of the Tone 1 (high-level) and Tone 4 (fall) contrast in Mandarin in Dutch-learning infants aged five to eighteen months. Infants across the age range all successfully discriminated the contrast. Their responses were phonetic rather than phonological, since Dutch does not have lexical tones. In Shi, Santos, Gao and Li (2017) 4-, 8-, and 11-month-old infants whose native language was French, a non-tonal language, also showed no decline in discriminating Tone 1 and Tone 4. Similarly, 18-month-old English-learning infants discriminated Mandarin Tone 2 (rise) and Tone 4 (fall) in a word-learning task involving tonal mispronunciations (Singh et al. 2014). Even non-tone-speaking adults show some degree of perceiving certain tonal contrasts in Mandarin (So and Best 2010). Acoustic salience may be a factor accounting for non-tone-learning infants' sustained discrimination of these contrasts. The 4- to 11-month-old French-learning infants in Shi, Santos, Gao and Li (2017) showed a tendency to decline over age in their discrimination of the more similar Tone 2 – Tone 3 contrast. Consistent with this idea, when Liu and Kager (2014) artificially reduced the pitch differences of their naturally produced stimuli (Tone 1 and Tone 4), infants showed a decline in discriminating the tones from eight to fifteen months of age.

Tsao (2008) tested the role of acoustic salience in the discrimination of lexical tones in Mandarin-learning ten- to twelve-month-old infants, using three tonal contrasts in Mandarin, Tone 1 (high-level) – Tone 3 (low-dipping), Tone 2 (rise) – Tone 4 (fall), and Tone 2 – Tone 3. Infants discriminated the acoustically most distinct Tone 1 – Tone 3 contrast significantly better. The latter two contrasts (Tone 2 – Tone 3; Tone 2 – Tone 4) did not differ in discrimination, both poorer than Tone 1 – Tone 3. In another study, however, Mandarin-learning infants aged eight to eleven months categorically discriminated Tone 2 and Tone 4 even when the tones were embedded in variable tonal contexts (Shi 2009).

Taken together, the perception of lexical tones by non-tone-learning infants is contrast dependent, with some contrasts showing the language-universal to language-specific developmental trajectory (same as certain consonants and vowels), but with some other tones remaining discriminable throughout infancy despite lack of relative experience. The development of native-tone-learning infants is little understood. Among the few existing studies, Yeung, Chen, and Werker's (2013) results were mixed and inclusive, and Tsao (2008) only tested infants aged ten to twelve months but not younger. In addition, the Headturn Conditioned Procedure in Tsao (2008) involved training the infants on the tonal contrasts that were subsequently tested; thus, infants' spontaneous discrimination of the tones remains unclear. In the next section we report our experiment on the perceptual development of native tones during the first year of life.

11.4 The experiment

To better understand the effect of natural perceptual capacity versus input-driven learning in the early development of native lexical tones, we examined Mandarin-learning infants' perception of Mandarin tones from four to thirteen months of age. We used a habituation procedure that tested infants' spontaneous responses to different tones without any training. Two tonal contrasts, Tone 2 – Tone 3 and Tone 1 – Tone 4, were tested, allowing us to examine whether there were contrast-dependent effects in infants' tonal perception. Furthermore, we used multiple exemplars for each tone, and crucially, the exemplars for the habituated tone during the test phase were different from those during habituation. This aspect differed from Tsao (2008) and Yeung et al. (2013), in which the same exemplars were used throughout training/familiarization and the test phases. The change of exemplars across experimental phases for the same tone ensured that our task definitively tested infants' generalized knowledge about tonal categories beyond the memorization of specific exemplars heard during habituation.

In Mandarin-Chinese, there are four lexical tones, high-level (Tone 1), rise (Tone 2), low-dipping (Tone 3), and fall (Tone 4). Tone 2 and Tone 3 are generally considered acoustically similar, as they are both contour tones starting from the mid-part of the pitch range and ending higher in the pitch trajectory, although their trajectories differ. They are also different in terms of mode of phonation: Tone 3 is often produced with creaky voice. It is unknown if this characteristic plays a role in infants' tonal discrimination. The tones of the other contrast that we tested, Tone 1 – Tone 4, shared the same pitch height at the tonal onset, and their pitch trajectories diverge, with Tone 1 staying high and Tone 4 moving downward. Tone 1 is typically longer than Tone 4. The two tones thus seem to have salient acoustic differences, and they were discriminable to both infants and toddlers whose native language contains no contrastive tones (Liu and Kager 2014; Shi, Santos, Gao and Li 2017). Tsao (2008) showed that Tone 2 and Tone 3 were more difficult to discriminate than Tone 1 versus Tone 3 for Mandarin-learning one-year-olds, but the Tone 2 – Tone 3 contrast was not more difficult than the Tone 2 – Tone 4 contrast.

It is unknown where the Tone 1 – Tone 4 contrast situates relative to the other contrasts in native-tone-learning infants' discrimination. In So and Best (2010), the Tone 2 – Tone 3 and Tone 1 – Tone 4 contrasts were comparably confusable to English-speaking adults, whereas the Tone 1 – Tone 3 contrasts were better perceived. According to So and Best (2010), the Tone 2 – Tone 3 and Tone 1 – Tone 4 contrasts were comparable in their perceptual salience because both contrasts contain tones that share pitch features (e.g., pitch height at onset/offset, pitch contour, etc.). However, phonologists consider contour tones as generally more complex than level tones (Yip 2002). This may mean that Tone 2 and Tone 3 are more difficult for discrimination than Tone 1 versus Tone 4, since the former contrast involves two contour tones, whereas the latter contains one level tone and one contour tone. There is no consensus regarding what determines perceptual salience, and the answer requires more experimental work.

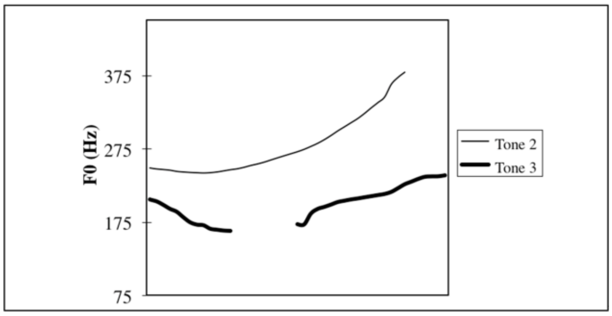
In our experiment we examined how Mandarin-learning infants' perception of Tone 1 – Tone 4 and Tone 2 – Tone 3 evolves during the first year of life. In a prior study (Shi, Gao, Achim and Li 2017) we had tested the discrimination of Tone 2 versus Tone 3 in a group of Mandarin-learning 4- to 13-month-old infants. Here we again tested this tonal contrast with two different age groups. The two particular contrasts (Tone 2 – Tone 3; Tone 1 – Tone 4) have been shown to be perceptually more confusable than other contrasts in previous studies; we thus chose them to test whether experience with the native language during the first year of life can yield improvement in infants' discrimination of the tones. We expected the Tone 2 – Tone 3 contrast to be relatively harder due to their lower acoustic salience (Tsao 2008), higher phonetic feature similarity (So and Best 2010), and greater phonological complexity (Yip 2002).

Stimuli. Two lexical tone contrasts, Tone 2 (rising) – Tone 3 (low-dipping) and Tone 1 (high-level) – Tone 4 (falling) were used for our experiment. The Tone 2 – Tone 3 stimuli were the same as those in Shi, Gao, Achim and Li (2017). The tone-bearing syllable was *can* (the pinyin alphabet) for the T2-T3 contrast, and *kui* for the T1-T4 contrast. The reason for choosing these syllables was that the morphemes represented by these syllables with the four tones are all unfamiliar to infants and young children, thus controlling for the factor of meaning. A Mandarin-Chinese-speaking female produced the stimuli in the infant-directed speech style in an acoustic chamber. During recording, she produced multiple exemplars of the syllables with all four tones, which ensured that the relative tone height and contours for the tones fell within the natural pitch range of the speaker. The stimuli were recorded with a 22 kHz sampling frequency, 16-bit resolution. The final selected stimuli for each target tone consisted of 13 tokens. The mean duration of the T2 tokens was 718 ms (max = 806 ms, min = 631 ms) with the standard deviation of 63 ms. The mean duration of the T3 tokens was 717 ms (max = 802 ms and min = 630 ms) with the standard deviation of 63 ms. An independent t-test showed that the duration of the tokens of the two tones did not differ, $t(24) = 0.47$, $p = 0.963$.

For T3, ten out of thirteen tokens had creaky voice, with six of the ten creaky tokens used for habituation and four used for test. The mean duration of T1 tokens was 585ms (max = 645 ms, min = 503 ms) with the standard deviation of 35 ms. The mean duration of the T4 tokens was 494 ms (max = 536 ms and min = 464 ms) with the standard deviation of 22 ms. An independent t-test showed that the tokens of T1 were significantly longer than those of T4, $t(24) = 7.861$, $p = 0.000$. All tokens were adjusted to comparable ampli- tude using Cool Edit Pro 2.0. Figure 11.1 shows example tokens of the two contrasts. In addition, we designed a visual stimulus, a colorful checkerboard-like geometrical image, which was presented along with the speech stimuli during the experiment.

Participants. Participants were a total of 62 monolingual Mandarin-Chinese-learning infants who resided in Beijing and heard standard Mandarin at home. Infants formed four groups defined by tonal contrast and age: T2-T3 younger group (n = 16, Mean: 6 months 5 days, Age Range: 5 months 26 days – 6 months 29 days); T2-T3 older group (n = 14, Mean: 8 months 22 days, Age Range: 7 months 10 days–11 months 0 days); T1-T4 younger group (n = 16, Mean: 5 months 18 days, Age Range: 4 months 15 days – 6 months 25 days); and T1-T4 older group (n=16, Mean: 11 months 20 days, Age Range: 9 months 19 days – 13 months 6 days).

(a)



(b)

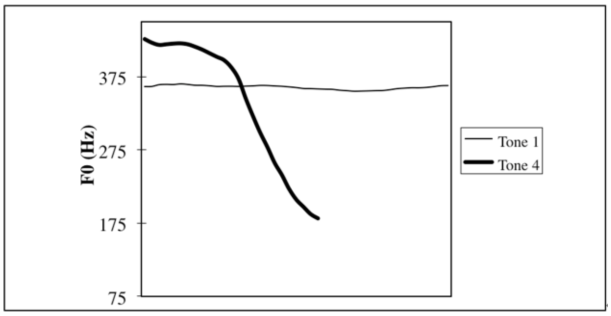


Figure 11.1 (a) Pitch trajectories of example stimuli of Tone 2 and Tone 3. The broken part in the mid-section of the Tone 3 pitch curve stands for creaky voice. (b) Pitch trajectories of example stimuli of Tone 1 and Tone 4

Apparatus. The experiment was conducted in a quiet room, where the infant sat on the mother's lap facing a computer screen. Loudspeakers were placed on both sides of the screen and played auditory stimuli simultaneously. The display screen and the loudspeakers were connected to a computer in the control room. Under the screen a camera transmitted the video of the infant to a computer in the control room. Blind to the stimuli of the experiment, the experimenter in the control room outside the testing room operated the computer to run the experiment program and coded online the infant's looking to and away from the screen. The experiment program was pre-set to pre- sent the audio and visual stimuli contingent upon the infant's looking to the screen. The program also recorded the looking-time data automatically and performed the habituation calculation online. During the experiment, the mother listened to masking music through headphones (Peltor HTM79A). She was asked not to interact with, interrupt, or influence the infant.

Procedure. The habituation paradigm was adopted. Each infant was habituated with one of the two tones in a contrast. Seven tokens of the tone were presented randomly and repeatedly across trials during the habituation phase. The experimental program recorded online the looking time of each trial. The looking time of each sliding window of three consecutive trials was compared online with the looking time of the first window of three trials. The habituation criterion was reached if the looking time in a later window declined to 50 percent or lower of the first window of trials, and the experiment proceeded into the test phase automatically. In the test phase there were two trial types, Same and Different. The Same type presented six novel tokens of the same tone that had been presented in the habituation phase. The Different type presented six tokens of the contrasting tone. That is, the test stimuli were all new, with the Same exemplars belonging to the habituated tonal category, and the Different exemplars belonging to the other tone that had never appeared during habituation. For both the habituation and test phases, the inter-stimulus-interval (ISI) within a trial was 1000 ms, and the maximum trial length was 21s. Each trial was initiated upon the infant's looking, and was terminated if he or she looked away for more than two seconds or if the maximum trial length was reached. When a trial stopped, an attention-getter, an animation of a jumping star, popped up automatically to attract the infant's attention back to the screen. The visual stimulus, a colorful checkerboard-like geometrical image occurred simultan- eously with the speech stimuli during each trial. In addition, a pre-trial and a post-trial were presented at the beginning and the end of the experiment. These trials presented a zooming picture of a cat. During the pre-test trial, the cat image was accompanied by the following speech: *Zhe shi shenme?* ("What's this?") *Mao* ("cat"), *mao, mao; zhe shi mao* ("this is a cat"), *mao, mao; yi zhi mao* ("a cat"), *mao, mao*". During the post-test trial, the auditory stimulus was only the word *Mao*, which was presented repeatedly. The pre- trial served to acquaint the infant with the equipment. The post-trial helped us judge whether infants were still on task toward the end of the experiment, as looking time should increase in the post-trial because the stimuli were distinct from those in the preceding trials.

Design. Infants were divided into two main groups, one for the T2-T3 contrast, and the other for the T1-T4 contrast. Within each contrast group, half of the infants were habituated with one tone, and the other half with the other tone. All of them then heard new exemplars of both tones in different test trials. Same and Different test trials were relative to the particular habituation tone. For example, for the T2 habituation infants, T2 was the Same test trial type, and T3 the Different type. The reverse was the case for the T3 habituation infants, with T3 being the Same and T2 being the Different test trials. The first test trial was either the Same type or Different type, counterbalanced across infants.

The looking time during the test trials was the dependent variable. The rationale of the habituation paradigm was that once infants became habituated with one tone, they should show renewed interest upon hearing a different tone in the test phase if they could discriminate the tones. In our design, the exemplars for both the Same and Different test tones were novel. We therefore predicted that if infants could categorize the tones of a contrast, they should look significantly longer in the Different than in the Same test trials, even though all test stimuli were novel. If infants could not categorize the contrasting tones, looking time to the Same and Different test trials should not differ.

Results. Looking time during the test trials was analyzed in a 2x2x2 mixed ANOVA, with Trial Type (Same, Different) as the within-subject factor, Age (younger, older) as the between-subject factor, and Contrast (T2-T3, T1-T4) as the between-subject factor. The results showed a significant main effect of Trial Type ($F(1, 58) = 7.519$; $p = 0.008$). There was no effect of Age, $F(1,58) = 0.62$, $p = 0.434$, and no effect of Contrast, $F(1, 58) = 1.13$; $p = 0.292$. Furthermore, we found no significance in any of the interactions. Figure 11.2 shows that looking times were significantly longer in Different than in Same

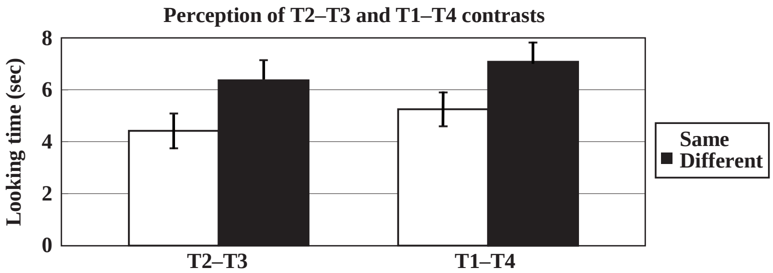


Figure 11.2 Results of both younger and older Mandarin-learning infants for the Tone 2 – Tone 3 (left two columns) and for Tone 1 – Tone 4 (right two columns) contrasts. Looking times (means and standard errors) were significantly longer in Different than in Same test trials

trials for each contrast. Because infant looking behavior can be quite variable, with some infants being overall long-lookers and others overall short-lookers, we log-transformed the raw looking times so as to reduce such variability. The same ANOVA was also conducted on the log-transformed data. The result pattern was identical to that of the raw data, with a significant main effect of Trial Type, $F(1, 58) = 12.999$; $p = 0.001$, and no other significant main effect nor interaction. These results indicate that Mandarin-learning infants perceived the two tonal contrasts in Mandarin successfully at both early and later stages of the first year of life. That is, T2-T3 and T1-T4 were equally perceptible to infants from four to thirteen months of age. T2-T3, a contrast generally considered the most acoustically similar among Mandarin tones, did not show a different pattern of development than T1-T4.

11.5 General discussion

The results of previous studies in the literature provide evidence for the existence of both input-independent discrimination and input-guided learning of phonetic categories. The natural capacity to discriminate certain phonetic categories is universally available at birth; the discrimination is maintained if the input language continues to support those contrasts, but the discrimination gradually declines if the contrast is absent in the input. This pattern was shown for various consonants and vowels. In the limited studies on lexical tones, early tonal discrimination and later decline were observed in non-tone-learning infants (Mattock and Burnham 2006; Mattock et al. 2008; Shi, Santos, Gao and Li 2017; Yeung et al. 2013). Our present study demonstrated the continuing ability to categorize native tone contrasts in Mandarin-learning infants from four to thirteen months of age, consistent with the idea that infants are born as universal listeners, and that input experience serves to maintain the perceptual sensitivity to native tonal contrasts.

Besides the evidence of maintenance in phonetic development, previous research on consonants has shown that experience with the ambient language can exert an enhancement effect for certain contrasts (e.g., Kuhl et al. 2006; Tsao, Liu, and Kuhl 2006; Narayan, Werker, and Beddor 2010). With respect to lexical tones, it is unclear whether there is input-driven facilitation of tonal discrimination over age. Our present study tested tonal contrasts that are presumably less salient perceptually (including the most similar T2-T3 contrast in Mandarin), offering a potential opportunity for observing gradual learning from input exposure. However, we found that Mandarin-learning infants perceived the tonal contrasts at both younger and older ages during the first year of life, showing no evidence of improvement.

The effect of perceptual salience for early phonetic development has been much discussed in the field. Acoustically, more distinct contrasts are assumed to be easier for discrimination, for both native and non-native contrasts; conversely, acoustically similar contrasts should be more difficult. Supporting evidence was reported in experiments that tested infants' discrimination of different tonal contrasts (e.g., Tsao 2008). T2 and T3 in Mandarin are generally considered the most similar contrast in terms of pitch patterns; nevertheless, the creaky mode of phonation in T3 might be helpful cues. T1-T4 may arguably be more distinct in their pitch patterns, and their durations clearly differ. Even non-tone infants and toddlers can discriminate this contrast (Liu and Kager 2014; Shi, Santos, Gao and Li 2017). On the other hand, the T2-T3 and T1-T4 contrasts seemed to be equally confusable for adult non-tone listeners in the study of So and Best (2010), who explained this result on the basis of their comparable degree of similarity in phonetic features (i.e., pitch-based features such as High and Low). Thus, if the pitch features of T2-T3 are considered as LH-LL and the T1-T4 as HH-HL, the tones within each contrast share the onset feature. However, the citation form of T3 has a final rise, making it LLH in pitch contour, which is why T2 are T3 are generally regarded as similar in the field. In our experiment, the perception of these two contrasts did not differ for Mandarin-learning infants. They discriminated both contrasts equally well, suggesting that acoustic cues beyond pitch patterns (such as creaky phonation) may contribute importantly to the comparable perceptual salience of the two contrasts. Future studies should examine whether these two contrasts differ from the most distinct contrast T1-T3 in infants' perceptual development.

The comparable discrimination for the T2-T3 and T1-T4 contrasts in our study cannot be explained in terms of contour tones versus level tones described in phonological theory (Yip 2002). The T2-T3 contrast contains two contour tones, with T3 being a complex contour tone. The T1-T4 contrast, on the other hand, contains one level tone and one falling contour, which should be easier for discrimination than the T2-T3 contrast. Our results are not consistent with this prediction. Contour tones are not necessarily more difficult for perception than level tones. Relative perceptual salience for tonal contrasts appears to depend on the exact acoustic-phonetic differences of the contrasting tones.

Our study further demonstrates that infants can categorize lexical tones. In the test phase of our experiment both Same and Different trials presented novel stimuli, unlike previous studies, which presented the same exemplars throughout the experiment for the same-tone category but new exemplars for the contrasting tone. Hence, our infants' responses to the test stimuli could not be simply due to a stronger interest in new versus old stimuli. Rather, they perceived the Same-trial new stimuli as belonging to the same tonal category of the habituation exemplars, and their stronger interest in the Different-trial stimuli suggests that they perceived them as belonging to a contrasting tonal category. In this sense, their tonal perception showed a certain degree of abstractness.

In conclusion, based on the findings from studies in perceptual development of phonetic categories, especially those from infant studies, we now know more about infants' initial state of speech-processing capacities and the role of input for later phonetic development. Input-independent processing and input-guided learning are both involved during acquisition. Furthermore, the perceptual system functions similarly for segmental categories (consonants and vowels) as well as for suprasegmental categories such as lexical tones, suggesting that they belong to a common phonetic-phonological system, which are subject to the same underlying mechanisms of acquisition and processing.

Author notes

The experiment reported in this chapter formed part of the doctoral thesis of the first author. The data were presented in the 2010 Speech Prosody and 2011 BUCLD meetings. This research was supported by grants from the National Social Science Fund of China (Project No.: 08AYY02) and from the Natural Sciences and Engineering Research Council of Canada (NSERC).
Corresponding authors for this article: Rushen Shi, shi.rushen@uqam.ca; Jun Gao, 9894575@qq.com

Bibliography

Best, C. T. (1995) “A direct realist view of cross-language speech perception” in Strange, W. (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 171–204.

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

Harrison, P. (2000) “Acquiring the phonology of lexical tone in infancy”, *Lingua*, 110(8), pp. 581–616.

Jusczyk, P. W., Cutler, A., and Redanz, N. (1993) “Preference for the predominant stress patterns of English words”, *Child Development*, 64, pp. 675–687.

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

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

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (1992) “Linguistic experience alters phonetic perception in infants by 6 months of age”, *Science*, 255, pp. 606–608.

Liu, L., and Kager, R. (2014) “Perception of tones by infants learning a non-tone language”, *Cognition*, 33(2), pp. 385–394.

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

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

Mattys, L., and Jusczyk, P. W. (2001) “Phonotactic cues for segmentation of fluent speech by infants”, *Cognition*, 78, pp. 91–121.

Narayan, C., Werker, J. F., and Beddor, P. (2010) “The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination”, *Developmental Science*, 13(3), pp. 407–420.

Polka, L., and Bohn, O.-S. (1996) “A cross-language comparison of vowel perception in English-learning and German-learning infants”, *Journal of Acoustical Society of America*, 100, pp. 577–592.

Polka, L., Colantonio, C., and Sundara, M. (2001) “A cross-language comparison of /d/–/ð/ perception: Evidence for a new developmental pattern”, *Journal of the Acoustical Society of America*, 109(5), pp. 2190–2201.

Polka, L., and Werker, J. F. (1994) “Developmental changes in perception of non-native vowel contrasts”, *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), pp. 421–435.

Shi, R. (2009) “Contextual variability and infants’ perception of tonal categories”, *Chinese Journal of Phonetics*, 2, pp. 1–9.

Shi, R., Gao, J., Achim, A., and Li, A. (2017). “Perception of lexical tones in native Mandarin-learning preverbal infants and toddlers”, *Frontiers in Psychology*, 8, pp. 11–17.

Shi, R., Santos, E., Gao, J., and Li, A. (2017). “Perception of similar and dissimilar lexical tones by non-tone-learning infants”, *Infancy*, 22(6), 790–800.

Singh, L., Hui, T. J., Chan, C., and Golinkoff, R. M. (2014) “Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation”, *Developmental Science*, 17(1), pp. 94–109.

So, C., and Best, C. (2010) “Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences”, *Language and Speech*, 53(2), pp. 273–293.

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

Tsao, F. M., Liu, H. M., and Kuhl, P. K. (2006) “Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants”, *Journal of Acoustical Society of America*, 120(4), pp. 2285–2294.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (1981) “Developmental aspects of cross-language speech perception”, *Child Development*, 52(1), pp. 349–355.

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

Yeung, H. H., Chen, K. H., and 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), pp. 123–139.

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

[This chapter was published in Prosodic studies: Challenges and prospects 2020]