Lexical tone perception in early infancy

Ao Chen [1], *René Kager* [1]

[1] Institute of Linguistics Uil OTS, Utrecht University

a.chen@uu.nl, r.w.j.kager@uu.nl

Abstract

In order to shed light on the perceptual reorganization of lexical tones in the first year of life, in the current study, we tested Dutch native infants (6, 9, 11-month-old) on their discrimination of Mandarin Chinese high rising tone (T2) and low-dipping tone (T3) using a visual fixation paradigm. The infants were habituated on multiple tokens produced by two female native speakers, carrying either Mandarin T2 or T3, and were then tested with two test trials in which the same tone as occurred in habituation (same trial), and the other unheard tone (novel trial) alternated. Their looking time to the same trial and the novel trial were used as indicator for discrimination. Contradictory to earlier studies, 6-month-old Dutch infants failed to discriminate between these two tones while on the other hand, 9-month-old and 11-month-old infants succeeded in doing so. However, if the infants were split in two groups according to the habituation tone, only those who were habituated on T3 were able to discriminate the contrast. The results are discussed from the perspectives of acoustical property of the contrast and perceptual bias reflecting T3 sandhi asymmetry.

Index Terms: lexical tone, language acquisition, perceptual reorganization

Introduction

In the last a few decades it has been extensively shown that infants are born with sensitivity to both native and non-native sound contrasts, but by the onset of the second year of life, the discrimination of non-native contrasts deteriorates while the sensitivity to native contrasts remains and improves [1]. Nevertheless, the success of early discrimination of non-native contrasts also depends on the acoustical quality of the contrast, i.e., the more similar the two sounds phonetically are, the more difficult they are for infants to discriminate [2].

Recently, the issues of perceptual reorganization have been extended from segment perception to lexical tone perception in early infancy [3][4][5]. In tone languages, lexical tones are realized by F0 variations, and they are used lexically to distinguish meaning. In comparison, for non-tone languages, F0 variations work on a scale larger than word such as intonation. Therefore, to study the perceptual reorganization of lexical tones with infants is informative for revealing the universality and language specificity in pitch perception.

Substantial discrepancy still remains in research regarding early lexical tone perception. Mattock and Burnham [3] tested the discrimination of Thai tones as well as non-speech tones showing the same pitch contours by native Chinese and native English infants, and found that both groups of infants succeeded in discriminating both speech and non-speech tones at 6 months regardless of language background. At 9 months, however, only Chinese infants were still able to discriminate the Thai lexical tones while the successful discrimination of non-speech tones remained for both language groups. Later [4], they found that French and English native infants were able to discriminate Thai tones both at 4 months and 6 months, which implied that the perceptual reorganization of lexical tones happened between 6 and 9 months, regardless of the specific prosodic property of their own native language. However, somehow surprisingly, though native Mandarin infants were able discriminate between Mandarin rising tone and dipping tone at both 6 months and 7-11 months, they failed to discriminate between the high-level tone and the falling tone in their native language at both 8 months and 12 months [6]. To make the picture even more confusing, another study tested the discrimination of lexical tones by Dutch infants by exposing them to either unimodal or bimodal distribution of Mandarin high-level tone and Mandarin high-falling tone [5], and found that 5-6 months infants were able to discriminate this tonal contrast no matter which distribution they were exposed to, while at 8-9 months, only those who were exposed to bi-modal distribution, which assembled the distribution of two tonal categories, were still able to do so. By 14 months, the infants failed to discriminate the tones even if they were trained with bi-modal distribution. Based on these studies, vague though, it seems that, first, the non-tone language infants' sensitivity to lexical tones deteriorates from birth to the onset of the second year; second, the acoustical property of different tones affects the discrimination.

Though informative these pioneer studies are, several points may need more scrutinization. First, the tonal contrasts used in [3][4][5] are perceptually salient, in the sense that the F0 contours of tones differ both in shape and in height, but for the acoustically less salient contrasts, whether the tone loss of the non-native infants follows the same trajectory is unknown. Second, lexical tones also exhibit phonological regularities, but the question of whether the early perception of lexical tones is interfered by tonal phonological grammar has been largely neglected. Third, the T test used in previous studies to analyze the data, assumes that the sample fits a normal distribution. However, except [5], the other studies did not mention the distributional information of their data, and hence, it is still an open question how reliable their results are.

As to the acoustical property of Mandarin lexical tones, among the four tones, i.e. high level tone (T1), high rising tone (T2), low dipping tone (T3), and high falling tone (T4), T2 and T3 have been claimed to be the most difficult contrast to acquire, both for adult L2 learners and infants [7][8]. Phonetically, the pitch contour of T2 and T3 both starts from low and has a rising part at the end. Moreover, T2 and T3 are claimed to be only partially contrastive due to the T3 sandhi rule in Mandarin, i.e. when two T3s occur together, the first T3 will change to a T2, which is perceptually same as an underlying T2 for native listeners [9]. As a result, for native listeners, a T2T3 sequence is confusing as it could either be a real T2T3 sequence or a sandhied T3T3 sequence while a T3T2 sequence is unambiguous. In addition, T2 and T3 is the only contrast that goes through phonological grammar in Mandarin. Perceptually, categorical perception experiments have shown that, native Mandarin adults perceive the T2T3 contrast in a categorical way while Dutch adults perceive them in a psycho-acoustical fashion. Interestingly, similar to native Mandarin listeners, Dutch listeners also discriminate between T2 and T3 better if they were presented with a T3T2 pair rather than with a T2T3 pair [10], as if they had knowledge of T3 sandhi. This result suggests that there might be some cognitive or auditory baises for T3 sandhi, i.e. T2T3 sounds more identical than a T3T2.

In order to explore how the perception of acoustically similar tones develop across the first year of age, and to shed light on possible innate biases in tone perception and acquisition, in the current study, we test Dutch infants on their discrimination of T2 and T3 across different ages (6, 9, 11 months old). As infants have not had substantial speech input yet, how they perceive this tonal contrast and their acquisition pattern would be very informative in revealing the natural bias in lexical tone perception.

2. Experiments

2.1 Participants

All the participants were full-term native Dutch infants. None of the infants reported ear infection for the three weeks before the experiment. There were 28 6-month-old infants (5:04-6:14), 30 9-month-old infants (7:25-9:28), 25 11-month-old infants (11:1-12:10. Another 11 6-month-old infants were tested and excluded for analysis (three for crying during the experiment, three for fussiness, three for not meeting the habituation criteria, one for equipment failure, and one for parents' interference); another four 9-month-old infants were tested but excluded for analysis (two for crying during the experiment, one for not passing the post-test, and one for failing to reach habituation criteria); another 18 11-months-old infants were tested but excluded for analysis (four for crying, four for equipment failure, six for looking time more than two standard deviation away to the mean looking time, three for failing to meet habituation criteria, one for delay in motoric development as reported by parents after the experiment).

2.2 Stimuli

Two female native Mandarin speakers produced multiple tokens of syllable /ma/ bearing Mandarin T2 and T3 respectively, together with other monosyllabic words with other tones. Two tokens of /ma/ with each tone of each speaker were selected as stimuli in the habituation phase. Another one token of each tone of one speaker were used as stimuli in the test phase. The duration of stimuli is between 517 and 814 ms, and the stimuli were balanced in intensity. The average pitch contours the stimuli of each tone produced by the two speakers were depicted in Figure 1 respectively. During the habituation and test phase, the visual stimulus was a colorful infant friendly picture.

2.3 Procedure

The visual fixation paradigm was used for the current study. There were a test cabin and a separate control room for the experimenter. During the experiment, infants sat on their parent's lap in the test cabin, and there was a 14 inch computer screen about one meter away in front of the baby displaying the visual stimuli. The auditory stimuli were presented at a comfortable volume through a hidden speaker in front of the baby. The parent listened to back ground music through headphone in order to prevent possible interaction with the infants. There was a hidden camera above screen recording the behavior of the infants, and the video was transferred to experimenter's computer in the control room. The experimenter observed and recorded the visual fixation of the infants by pressing the "looking" and "non-looking" button on a button box connected to the control computer.

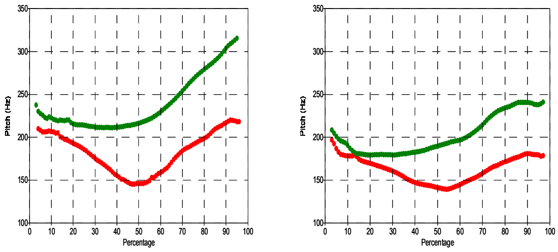

Figure 1 *Time normalized average F0 contours of the T2 and T3 tokens produced by the two speakers. Upper line represents the T2 and lower line represents the T3.*

The experiment started with a habituation phase and immediately after the habituation followed the test phase. A pretest and a posttest were used to measure general attention of the participants. During the pretest and posttest, the stimuli were moving infant friendly pictures accompanied by beeps. When the pretest finished, and once the participant focused on the screen, the experimenter initiated the first habituation trial by pressing the "looking" button, and once the participant looked away, the experimenter pressed the "non-looking" button. The looking and non-looking of the participants were always recorded by these two buttons. If the infant looked away and looked back to the screen within two seconds, the same trial continued, and if the infant looked away for more than two seconds, the current trial ended and a smiling baby face appeared on the screen to get the attention of the participant back. Once the infant looked back to the screen, the experimenter started the next trial by pressing the "looking" button again. The looking time of each look as well as the total looking time of one single trial was recorded automatically by the experimenter's control computer.

The total looking time of the first three trials in the habituation phase was used as a baseline to measure habituation. Once the total looking time of three consecutive trials was less than 65% of the total looking time of the first three habituation trials, the habituation criteria was met, and the test phase started automatically. In the test phase, the infants were presented with one "old" trial, which was /ma/ with the same tone as they had heard in the habituation phase, and another "novel" trial which was /ma/ carrying the other tone that they had not heard yet in the habituation phase. The tones that were used in habituation phase and the order of the "old" and "novel" trials in the test phase were counter-balanced among the participants.

If the infants were able to detect the difference between the two tones, then in the test phase, when they heard the novel trial, their listening time should be recovered due to hearing something new, i.e. they would look at the screen for a longer time when presented with the novel trial than when presented with the old trial.

2.3 Results and discussion

The videos of all the participants were recoded offline after the experiment before submitted for analysis. After recoding, the raw looking time of the "old" and "novel" trials was logarithmically converted to correct the skewness of the distribution of the raw looking time data. The log transformed looking time of all the age groups fits normal distributions ($D_{6\ months}$ (56) = 0.12, p>0.2; $D_{9\ months}$ (60) = 0.072, p>0.2, $D_{11\ months}$ (50) = 0.078, p>0.2). Statistics hereafter are based on the log transformed looking time (LGLT).

2.3.1 6-month-old infants

First, a 1-tailed paired T test was carried out between LGLT of the "old" and "novel" trials, between LGLT of the "old" trial and the last trial in habituation (last hab), and between LGLT of the "last hab" and the "novel" trial. None of the pairs revealed significant difference: $T_{old\text{-}novel}$ (27) = -0.391, p>0.05; $T_{last\ hab\text{-}old}$ (27) = -0.036, p>0.05; $T_{last\ hab\text{-}novel}$ (27) = -0.328, p>0.05.

Second, the LGLT of the participants habituated on T2 and habituated on T3 was analyzed separately. As previous analysis has shown that LGLT of the last habituation trial and the novel trial did not reveal any significant difference, here only the LGLT of the old trial and the LGLT of the novel trial were submitted for statistical test. A paired T test showed that, regardless of on which tone they were habituated, there is no significant difference between the LGLT of the old trials and that of the novel trials: T_{habT2} (13) = -1.25, p>0.05; T_{habT3} (13) = 0.563, p>0.05. Figure 2 gives overall mean LGLT of the three trial types of the 6 months old infants, and the mean LGLT of the old trial and the novel trial separated by the habituation tone.

As could be read from the figures, although the infants tended to look longer to the novel trial, the looking time difference failed to reach significance statistically, which implies that 6-month-old Dutch infants were not able to discriminate between Mandarin T2 and T3, no matter on which tone they were habituated.

2.3.2 9-month-old infants

Same statistical tests as used for the 6-month-old infants were carried out for the 9-month-old infants. The LGLT difference between of the last habituation trial and the old trial, as well as between last habituation trial and the novel trial, failed to reach significance: $T_{lasthab\text{-}old}$(29) = 1.505, p>0.05; $T_{lasthab\text{-}novel}$(29) = -0.005, p>0.05. However, LGLT difference between the novel trial and the old trial did reach significance: $T_{old\text{-}novel}$(29) = -1.704, p<0.05 (1-tailed). When the participants are separated into two sub-groups based on the tone that they were habituated on, only those who were habituated on T3 showed significantly longer LGLT to the novel trial: T_{habT2} (15) = 0.792, p>0.05; T_{habT3} (13) = -3.597 p<0.005. Figure 3 gives the overall mean LGLT of the three trial types, and the mean LGLT of the old trial and the novel trial separated by the habituation tone.

It is evident that the overall significance between the LGLT of the old trial and that of the novel trial was caused by the huge difference in LGLT among those infants who were habituated on T3: only infants that were habituated on T3 looked significantly longer to the novel trial in the test phase, which suggests that they were able to discriminate between Mandarin T2 and T3. It seems that with the same amount of variation, to form a tonal category of T3 is somehow easier than to form a category of T2, and as a result, to discriminate a T2 from the T3 category is easier than the other way around.

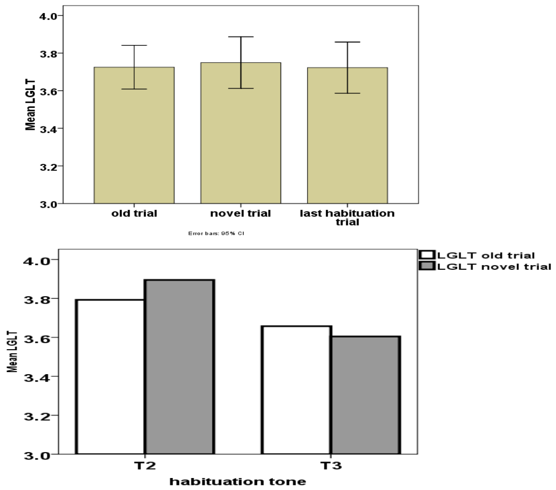

Figure 2 *overall mean LGLT of the three trial types of the 6 months old infants, and the mean LGLT of the old trial and the novel trial separated by the habituation tone.*

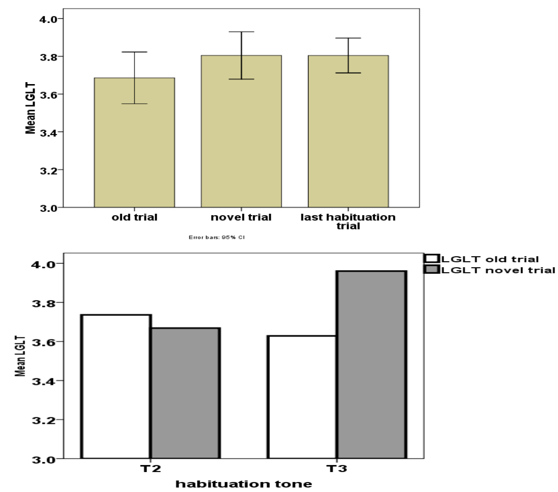

Figure 3 the overall mean LGLT of the three trial types, and the mean LGLT of the old trial and the novel trial separated by the habituation tone.

2.3.3 11-month-old infants

Same statistical tests as used for the 6-month-old infants were carried out for the 11-month-old infants. Same as 9-month-old group, the LGLT of the novel trial is significantly longer than that of the old trial: $T_{old\text{-}novel}$ (24) = -3.417, p<0.05. Again, the LGLT difference between of the last habituation trial and the old trial failed to reach significance: $T_{lasthab\text{-}old}$(24) = 1.448, p>0.05. But different from the 9-month-old group, the LGLT of the novel trial is significantly longer than the last habituation trial: $T_{lasthab\text{-}novel}$(24) = -1.648, p<0.05. When the participants are separated into two sub-groups based on the tone that they are habituated on, once again, only when infants were habituated on T3, the LGLT difference between the old trial and novel trial reached significance: T_{habT2} (11) = -1.638, p>0.05; T_{habT3} (12) = -3.236 p<0.05. Same as the 9-month-old group, the overall significant LGLT difference between the old trial and novel trial was caused by those who were habituated on T3, as their looking time to the novel trial was much longer than that to the old trial. Figure 4 gives the overall mean LGLT of the three trial types, and the mean LGLT of the old trial and the novel trial separated by the habituation tone.

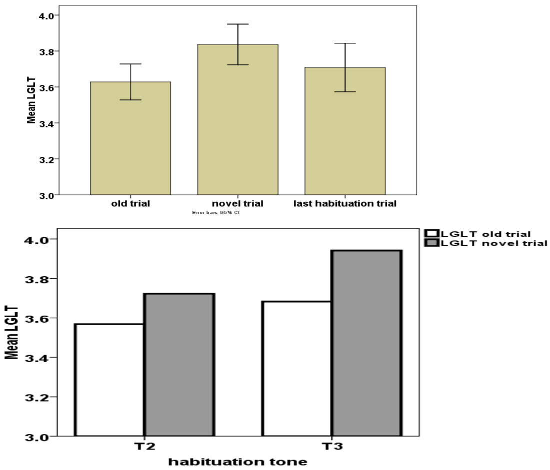

Figure 6 *the overall mean LGLT of the three trial types, and the mean LGLT of the old trial and the novel trial separated by the habituation tone.*

2.3.4 Cross-age analysis

As the 9-month-old group and 11-month-old group have shown the same pattern in discrimination, a cross-age analysis was carried for these two groups. The raw looking time difference between the old trial and the new trial (difference score) was calculated for the 9-month-old group and f 11-month-old group respectively, and the difference score of both groups fits a normal distribution ($D_{9\ months}(30) = 0.076$, $p>0.2$, $D_{11\ months}(25) = 0.068$, $p>0.2$), and Levene's test of homogeneity of variance revealed no significance ($F(2, 55) = 2.504$, $p>0.05$). A one-way ANOVA was carried out taking the difference score as dependent variable and age as between-group factor, and no significant difference is found between the two groups: $F(1) = 0.958$, $p>0.1$. This result suggests that native Dutch infants perceive Mandarin T2 and T3 in the same pattern from 9 to 11 months.

3. General discussion

As we could see from the results above, native Dutch infants are not able to discriminate between Mandarin T2 and T3 at 6 months, but somehow surprisingly, their ability to discriminate this tonal contrast improves by 9 months, and the same perception pattern could be observed until at least 11 months. Nevertheless, even at 9 months and 11 months, they succeed in discrimination only if they are habituated on T3.

Several points are worth mentioning regarding the results. First, contradicting previous research, in the current study, the 6-month-old infants are not able to discriminate between non-native tonal contrasts. As Dutch infants do succeed in discriminating the Mandarin T1 versus T4 contrast [5], their failure in discriminating T2 versus T3 could plausibly be due to the acoustical property of the T2-T3 contrast. As mentioned earlier, perceptually T2 and T3 is the most difficult contrast for both native Mandarin children [8] and adult L2 learners [7]. Indeed, the F_0 contours of the two tones start from almost the same height, while in the second half of the F_0, both tones rise in parallel, so it could be that the acoustical differences between the tones are too small for infants at 6 months to pick them up. Second, clear discrimination asymmetry has been observed in these experiments. Perceptual asymmetry among infants is not rare: coronal consonants have been claimed to be underspecified as compared to labial consonants, which hinders discrimination of labial from coronal but not vice versa [11]. Similarly, if Dutch infants were to discriminate between Mandarin T2 and T3, they need to first build up a tonal category based on the input in the habituation phase. For Dutch infants, the T2 category may well be underspecified: for example, frequently occurred Dutch monosyllables such as *ja* ("yes") and *nee* ("no") can be realized with a range of rising contours that acoustically resemble Mandarin T2. Dutch infants' experience with rising contours does not exclude T3 from falling into this category. Last and most importantly, the behavior of infants is in a way similar to the performance of Dutch adults, i.e. they discriminate between T2 and T3 better if they are first presented with a T3. Taken together, it seems that compared to a T3T2 sequence, a T2T3 sequence is more likely to be perceived as identical, and hence perceptually a T2T3 may be closer to a T3T3 than a T3T2. As T3 sandhi is also positional and asymmetrical, our results imply that there may be a perceptual basis for the T3 sandhi rule in Mandarin.

4. References

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

[2] Narayan, C. R., Werker, J. F., and Beddor, P. S. (2010). "The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination." Developmental Science, 13(3):407-420, 2010

[3] Mattock, K., and Burnham, D. "Chinese and english infants' tone perception: Evidence for perceptual reorganization." Infancy*,* 10(3): 241-265, 2006

[4] Mattock, K., Molnar, M., Polka, L., and Burnham, D. "The developmental course of lexical tone perception in the first year of life." Cognition, 106(3):1367-1381, 2008

[5] Liu, L. "Is perceptual reorganization affected by statistical learning? Dutch infants' sensitivity to lexical tone", Master thesis, Utrecht University, 2010

[6] Shi, R., Gao, J., and Li, A. "Perception of lexical tones in infants", poster, Boston University Conference on Language Development 36, 2011

[7] Gandour, J. "Tone perception in far eastern languages.", Journal of Phonetics, 11(2):149-175, 1984

[8] Li, C. N., and Thompson, S. A. (1977). "The acquisition of tone in mandarin-speaking children", Journal of Child Language, 4(02):185-199.

[9] Hume, E., & Johnson, K. "The impact of partial phonological contrast on speech perception". Congress of Phonetic Sciences XV proc, 2003

[10] Chen, A. and Kager, R., "The perception of lexical tones and tone Sandhi in L2: success or failure", Congress of Phonetic Sciences XVII proc, 2011

[11] Dijkstra, N. & P. Fikkert (subm.) "Universal Constraints on the Discrimination of Place of Articulation? Asymmetries in the Discrimination of 'paan' and 'taan' by 6-month-old Dutch Infants". BUCLD 35 *Proc, 2010*