DISYLLABIC TONES IN MANDARIN PRESCHOOL CHILDREN AND CHILD-DIRECTED SPEECH

Zhiwei Wang[1], Jun Gao[2], Aijun Li[2]

[1]Beijing International Studies University (BISU), China
[2]Institute of Linguistics, Chinese Academy of Social Sciences, China
2549066586@qq.com gao-jun@cass.org.cn liaj@cass.org.cn

ABSTRACT

This study investigates the developmental performance of disyllabic tones in Mandarin-speaking preschoolers (aged 3-6) and child-directed speech (CDS). Comparing the acoustic parameters of tones in disyllabic words produced by preschoolers in four age groups and CDS, we found that: 1) the preschoolers have already acquired the tone sandhi rules phonologically; 2) their phonetic accuracy is over 95% after 4 years; the error rate of the first syllable is higher than those on the second, declining significantly with age; more errors are found in Tone 2 and Tone 3 than Tone 1 and Tone 4; 3) preschoolers' off-standard production is mainly characterized by the 'undershoot' of tonal target regarding the slope and curvature of the pitch contour; 4) pitch register, pitch range, and duration of disyllabic tones show significant difference between child-Mom dyads, but exhibit consistent developmental patterns between preschoolers and CDS, which reflects a synchronous entrainment.

Keywords: Lexical tone acquisition, Disyllabic tones, Mandarin-speaking preschoolers, child-directed speech

1. INTRODUCTION

Mandarin Chinese has four lexical tones: Tone 1 (T1, high level, HH), Tone 2 (T2, rising, LH), Tone 3 (T3, low rising, LLH), and Tone 4 (T4, falling, HL), transcribed as 55, 24, 213, and 51 in a 5-point scale [1]. The T3T3 disyllabic tone in Mandarin undergoes a phonological tone sandhi rule whereby T3T3 is realized as T2T3. When positioned in the first syllable, T3 and T4 are frequently realized as half-T3 (21) before T1\T2\T4 and half-T4 (53) before T4 due to tone coarticulation. The neutral tone in Mandarin (T5), is typically short in duration and found in the final unstressed syllable. The pitch contour of the neutral tone may vary with the preceding tones [2]. So, for normal stressed words the phonological and phonetic variations of the tone on the first syllable in disyllables are more complex than the tone on the second. Although many studies have examined Mandarin tones and tone sandhi, children's tone production and the relation between children's production speech (CPS) and child-directed speech (CDS) have been insufficiently investigated.

Previous auditory-based research suggested that Mandarin-speaking children acquire tones earlier than other linguistic elements by the age of 1.5 [3, 4]. The analysis of a children's cross-sectional corpus (CASS_Child_Word) revealed that at 18 months (1.5 yrs), Mandarin-speaking children exhibited a high tone production error rate (around 21%) in each syllable context, which decreased with age and dropped to around 2% at 42 months (3.5 yrs) [5].

However, research using the method of acoustic speech analysis yields different results [6-8]. Three-year-old children are found to be capable of perceiving T1, T2, T4 accurately, but their accuracy of tone production is much lower than adults [6]. There are also clear distinctions between children's and adults' tone production of monosyllabic words [7]. The acquisition of monosyllabic tones in Mandarin is suggested to be incomplete even at the age of 4-5 [8]. Adult-like tone (T3 and T4) and sandhi tone production is mastered after the age of 5 [9, 10].

When talking to young children, people employ a genre of child-directed speech (CDS) that is markedly different from adult-directed speech (ADS). Two hypotheses were proposed to explain the features of CDS against ADS: the hyperarticulation hypothesis [11] and the pragmatic hypothesis [12]. The former suggests that mothers enhance phonemic contrasts to facilitate speech and language acquisition in children. The latter hypothesis proposes that the acoustic differences between CDS and ADS result from mothers' affective emotions, to draw children's attention [13, 14].

Given that children's development of tone and tone sandhi rule has not been sufficiently explored, this study investigates the development of the disyllabic tone in Mandarin preschool children, including the distribution of production errors, the acoustic patterns, and the interaction between CPS and CDS. We predict that the error rate of the tone on the first syllable will be higher than that on the second due to the high variability of the first tone mentioned above [4]. Also, due to the changes in the CDS mentioned above [14], we predict that there will be a

synchronous entrainment effect [15] between CPS and CDS which is closely related to the development of children's speech.

2. MATERIALS AND METHODS

2.1. Participants and speech data

A total of 39 children and 11 mothers participated in this study. Mandarin is the language the mothers used exclusively with their children. The children were roughly divided into four age groups as shown in Table 1. None of the participants have a history of speech/hearing disorders.

Group	Age of the children	Children#	Moms#
G3	3 yrs (3;0-3;10, Mean: 3;4)	10 (2 males)	3
G4	4 yrs (4;0-4;11, Mean: 4;8)	11 (5males)	4
G5	5 yrs (5;0-5;10, Mean: 5;2)	12 (7 males)	2
G6	6 yrs (6;0-6;5, Mean: 6;3)	6 (3 males)	2

Table 1: Demographic information of participants.

The data were extracted from CASS_Child_Word [16], collected through a picture-naming task. If the child failed to produce the target word, the mother would say it and ask the child to imitate and repeat. Hence, some trials including both the child's and the mother's speech were selected as the recording samples in this study. There were a total of 1580 usable tokens, including 1360 produced by the children and 220 in CDS.

Two native speakers of Mandarin Chinese classified CPS as accurate, off-standard, and erroneous based on their auditory judgment. The recordings were annotated by a professional annotator in Praat. F0 was then extracted at 10 points of equal interval for each tone, then transformed into semitones with reference frequency of 54 Hz.

Ten acoustic parameters of each tone were calculated according to [17], namely Initial Pitch(pitch at time point 1), Mid Pitch (pitch at time point 5), Final Pitch (pitch at time point 10), Min Pitch 50% (minimum pitch in the second half of the tone), Max Pitch 50% (maximum pitch in the second half of the tone), Mean Pitch 50% (mean pitch in the second half of the tone), Pitch Range 50% (the absolute value of the difference between Max Pitch 50% and Min Pitch 50%), Pitch Range 100% (the absolute value of the difference between the maximum and minimum pitch in the whole tone contour), Slope 50% (directional excursion in the second half of the tone divided by the duration between Max Pitch 50% and Min Pitch 50%) and Tone Duration.

3. RESULTS

3.1. Accuracy distribution and confusion matrix in all age groups of children

Figure 1 shows the accuracy of the 16 disyllabic tone pairs produced by children aged 3;0-6;5. The overall accuracy is high, with an average of 95.36% (*Sd* = 4.99%) among the 16 tone combinations. Tone sandhi rules can cause confusion for children. Hence, the tone pairs containing T2 or T3 have lower accuracy, including T2T3 (92.31%), T4T2 (92.68%), T4T3 (93.10%), T2T2 (93.33%) (errors involving the first and the second syllables). It is also noteworthy that the accuracy of T4T4 is just 80%.

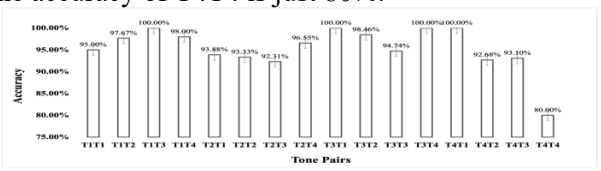

Figure 1: Accuracy of disyllabic tone pairs produced by children (regardless of age).

The confusion matrix of the 16 disyllabic tone pairs is presented in Table 2. The first tone tends to be mispronounced when followed by T1. Similarly, the tone pairs with T4 as the second tone tend to substitute whatever first tone with T1 or T3. Concerning the tone pairs with T2 or T3 as the second tone, the error patterns are more diverse. Among these tone pairs, the second tone is more frequently mispronounced. T2 tends to be replaced by T3 or T1 on the first syllable, and T3 by T1 or T4. On the second syllable, T2 tends to be replaced by T3. There is no case where both the first and the second tone are wrong at the same time.

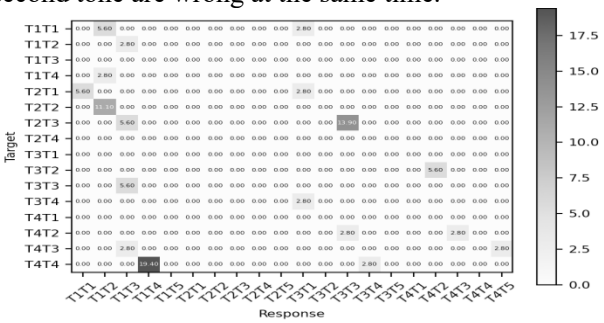

Table 2: The confusion matrix of disyllabic tones produced by children (regardless of age).

Off-standard production is defined as the case where the phonological category of tones is correct, but there are still problems with the phonetic realization. Specifically, for T1 (HH), the children's off-standard production shows a F0 contour similar to the rising or falling tone. For T2 (LH), T3 (LLH), the children's off-standard production is of the insufficient slope or curvature, showing a rather flat F0 contour. Similarly, for T4 (HL), the pitch slope is smaller with a lower start and insufficient falling.

3.2. Accuracy and error rate by age groups

Figure 2 shows the accuracy of disyllabic tones produced by children in each age group, in which S1 and S2 respectively stands for the first and second syllable. Statistical results indicate that the accuracy of CPS has significant positive correlations with age ($r1$=0.85, $p1$=0.001<0.05; $r2$=0.79, $p2$=0.001<0.05).

Figure 3 shows the error and off-standard rates of disyllabic tones produced in each age group. There is a clear downward tendency in the error rate of S1 from G3 to G5, and it tends to stabilize between G5 and G6. The off-standard rate of S1 declines from G3 to G6. The error rate of S2 is higher in G3, which indicates an anomaly in G4. However, in G5 and G6, the error rate stabilizes. The off-standard rate of S2 decreases from G3 to G6. Statistical results indicate that the error and off-standard rates of CPS have significant negative correlations with age (error rate: $r1$= -0.86, $p1$=0.047<0.05; $r2$=-0.99, $p2$=0.000<0.05; off-standard rate: $r1$=-0.65, $p1$=0.04<0.05; $r2$=-0.92, $p2$=0.000<0.05). Both the error and off-standard rates are higher for S1 than S2. Before the age of 4, the error rate of S1 was higher than S2, and the two rates of G4 were both less than 5%.

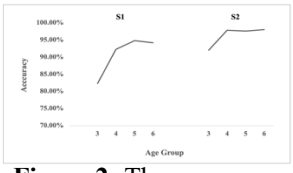

Figure 2: The accuracy of disyllabic tones in CPS.

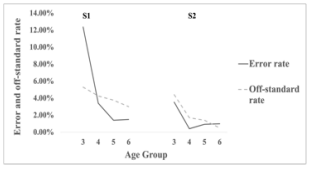

Figure 3: The error and off-standard rate of disyllabic tones in CPS.

Figure 4 shows the error and off-standard rates of CPS. In terms of S1, T2 has the highest error rate, followed by T4 and T3. As for S2, T3 is correct and those of the other tones are below 2% on average. The off-standard rate of T3 on S1 is the highest, followed by T2, and T4 and T1; and that of T3 on S2 is the highest, followed by T1, T4 and T2. Except T1, the error rate and off-standard rate of T2, T3, T4 on S2 are lower than on S1.

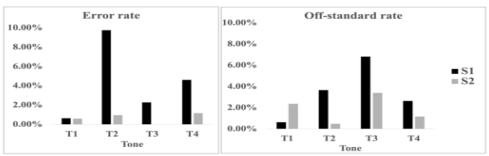

Figure 4: The error and off-standard rates of the tones on S1 and S2 in CPS.

3.3. Developmental patterns of the acoustic properties of disyllable tones in CPS and CDS

3.3.1 Pitch

This section analyses the patterns of disyllabic tones in children's correct production and in CDS. Figure 5 shows the mean pitch contours of the four tones on S1 and S2 in both CPS (in different age groups) and in CDS.

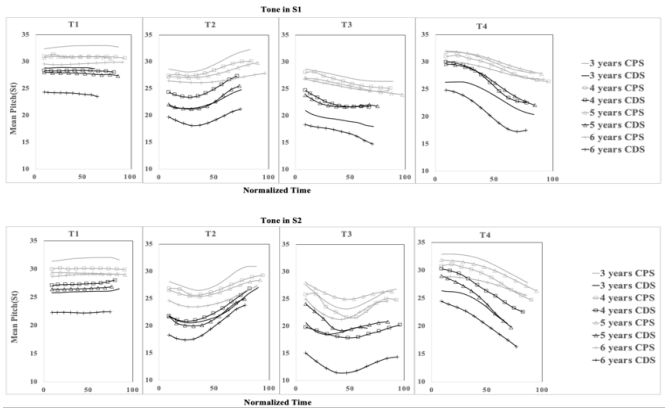

Figure 5: Pitch contours of disyllabic tones in CPS and CDS.

Comparing CPS and CDS, the pitch contours of T2 and T4 on the S1 position in CPS are flatter than those in CDS with lower absolute values of the rising and falling slope. The realization of the low target of T3 in CPS is poorer than that of CDS on both S1 and S2. Among different age groups in CPS, the overall pitch registers are downward with age. The contour of falling part of T3 and T4 on S1 grows with age, gradually approaching CDS's performance. As children age grows, S2 demonstrates three developments in tone production: the falling of T4, the forward movement of the pitch feature points of T2, and an increased curvature of contour tones. Finally, comparing the CDS for each age group of children, except for G3, the pitch of CDS for the other groups declines with age, following the same pattern as CPS.

Figure 6 depicts the developmental pattern of pitch range and pitch register in CPS and CDS. The mean pitch range of S2 is larger than that of S1 in both CPS and CDS, and that in CDS is larger than that in CPS for both syllables. Other than the CDS for 3-year-olds, the mean pitch register of S1 and S2 exhibits a decreasing trajectory in both CPS and CDS as the children grow older.

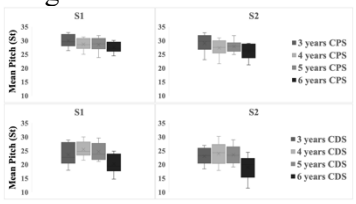

Figure 6: Pitch register and pitch range of S1 and S2 in CPS and CDS.

3.3.2 Duration

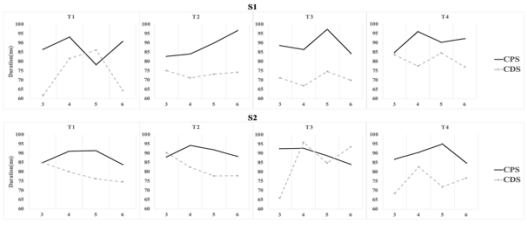

Figure 7: The mean duration of CPS and CDS.

Figure 7 shows the mean duration of disyllabic tones in CPS and CDS. The duration of each tone produced in CPS is longer than that in CDS. This phenomenon is more prominent in S1 than in S2.

3.4. Statistical analysis of the acoustic properties of disyllabic tones in CPS and CDS

For each of the acoustic parameters listed in Section 2.1, a multivariate analysis of variance (MANOVA) was performed using genre (CPS and CDS) as between-subjects factor, syllable context, and tone as within-subject factors and the acoustic parameter as the dependent variable.

As expected, significant main effects of tone were found in all comparisons (all $p<0.05$). There were significant main effects of genre in Pitch Range 50% (St) (CDS>CPS, $p<0.001$), Pitch Range 100% (St) (CDS>CPS, $p<0.001$), and Tone Duration (ms) (CPS>CDS, $p<0.001$), which means the pitch range of CDS is significantly wider than that of CPS, and the duration of CPS is significantly longer than that of CDS. Significant interactions between tone and genre were found for three acoustic parameters: the significance of CPS in Pitch Range 100% (St) (CPS: T4>T2=T3>T1, $p<0.05$; CDS: T4>T2>T3>T1, P<0.001), and Slope 50% (St/ms) (CPS: T2>T3>T1>T4, $p<0.05$; CDS: T2>T3>T4>T1, $p<0.001$) is lower than that of CDS.

4. DISCUSSION AND CONCLUSION

While children generally produce disyllabic tones well and their accuracy improves with age, the tone pairs containing T2 and T3 were found to be relatively difficult to acquire, and the T4T4 pairs had the lowest accuracy. The errors of replacing the T4 on the first syllable with T1 were mostly made by the lower age groups (3-4 years). These error patterns may be attributed to the proportion of tone combinations we selected from CASS_Child_Word. Furthermore, in the T4T4 tone pair, the physical limitations of children's pronunciation prevent F0 from dropping to an appropriate level. As a result, children may realize T4 at the position of S1 as T1, influenced by S2. The positive correlation between the accuracy and age indicates developmental improvements. The higher error and off-standard rates in S1 compared to S2 can be attributed to the "right-dominant-like" tone sandhi in Mandarin (tone sandhi of neutral tone word is a left-dominant-like one, which is not concerned in this paper), which echoes the results of previous studies on similar tone sandhi patterns [18].

The significance interaction between tone and genre shows that although children produced appropriate tone contours for all tones, children aged 3 to 6 do not produce adult-like disyllabic tones in terms of pitch range, slope, and length, consistent with previous study [9].

Comparing CPS and CDS, we found that their pitch registers share a same downward trend with children's age. This similarity in developmental patterns can be explained by the entrainment effect where the parent makes effort to communicate with children effectively and convey pragmatic information.

To conclude, this study investigates the production of disyllabic tones by Mandarin-speaking preschoolers and analyses the development of acoustic patterns in both CPS and CDS. We found that preschoolers have a phonological mastery of the tone sandhi rules. After the age of 4, the phonetic accuracy of children's disyllabic tone production has exceeded 95%, but more errors are made in Tone 2 and Tone 3 than in Tone 1. Since Mandarin has a kind of "right-dominant-like" sandhi pattern for normal stressed word, the first (left) syllable is more likely to be pronounced erroneously than the second (right). The error rate decreases dramatically with age. Moreover, we analyzed the phonetically off-standard pattern of CPS, which is characterized primarily with the "undershoot" of tonal goal with respect to the slope and curvature of pitch contours. T1 has a steeper slope, while T2, T3 and T4 have insufficient slope and curvature. Significant differences were also identified in the pitch register, pitch range, and duration of disyllabic tones across the four age groups of children and CDS. CPS and CDS follow similar developmental patterns with age, reflecting a synchronous entrainment. However, current evidence is insufficient to verify or contradict both the hypotheses. The results of this study therefore support our prediction that the error rate of the first syllable would be higher than that of the second, and that there is a synchronous entrainment between CPS and CDS, which is closely related to the developmental trajectory of child speech.

The present study would be helpful in the following aspects. Firstly, the confusion matrix in our study will be beneficial for the analysis of the speech of atypical children. Secondly, the understanding of the developmental pattern of the suprasegmental characteristics of Mandarin contributes to the studies and practices on Mandarin teaching to child learners with special needs and in ethic areas, as well as to teach Chinese as a second language for children.

For future work, we will verify the results with more age-grouped test data from the corpus. Tone pairs containing neutral tones would also be considered. We are conducting a perceptual experiment for Mandarin preschoolers, to explore the relationship between tone perception and production.

5. ACKNOWLEDGEMENTS

This work was supported by the Project of Cultural Experts and "Four Batches" of Talents directed by Aijun Li and CASS Innovation Project.

6. REFERENCES

[1] Chao, Y.-R. 2002. *Linguistic Essays*. The Commercial Press.

[2] Cao, J. 1992. On neutral-tone syllables in Mandarin Chinese. *Canadian Acoust.* 20, 49–50.

[3] Li, C., Thompson, S. 1977. The acquisition of tone in Mandarin-speaking children. *Journal of Child Language* 4, 185–199.

[4] Hua, Z., Dodd, B. 2000. The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language* 27, 3–42.

[5] Gao, J., Li, A., Zhang, Y. 2017. The production of tone in monosyllabic and disyllabic words by Mandarin-speaking children. *Studies on Speech Production–11th International Seminar*, 2017, Tianjin, China, October 16–19, 2017, Revised Selected Papers 89–98.

[6] Wong, P. 2012. Acoustic characteristics of three-year-olds' correct and incorrect monosyllabic Mandarin lexical tone productions. *Journal of Phonetics* 40, 141–151.

[7] Wong, P., Schwartz, R. G., Jenkins, J. J. 2005. Perception and production of lexical tones by 3-year-old, Mandarin-speaking children. *Journal of Speech, Language & Hearing Research* 48, 1065–1079.

[8] Wong, P. 2013. Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *The Journal of the Acoustical Society of America* 133, 434–443.

[9] Wong, P., Strange, W. 2017. Phonetic complexity affects children's Mandarin tone production accuracy in disyllabic words: A perceptual study. *PloS one.* 12.

[10] Xu Rattanasone, N., Tang, P., Yuen, I., Gao, L., Demuth, K. 2018. Five-Year-olds' acoustic realization of Mandarin tone sandhi and lexical tones in context are not yet fully adult-like. *Frontiers in Psychology* 9, 817.

[11] Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., Lacerda, F. 1997. Cross-language analysis of phonetic units in language addressed to infants. *Science* 277, 684–686.

[12] Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., Cristia, A. 2015. Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science* 26, 341–347.

[13] Tang, P., Xu R. N., Yuen, I., Demuth, K. 2017. Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America.* 142. 493-503.

[14] WANG, L., Kalashnikova, M., Kager, R., LAI, R., WONG, P. 2021. Lexical and Prosodic Pitch Modifications in Cantonese Infant-directed Speech. *Journal of Child Language*. 48. 1-27.

[15] Edlund, J., Heldner, M., Hirschberg, J. 2009. Pause and gap length in face-to-face interaction. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH.* 2779-2782.

[16] Gao, J., Li, A., Xiong, Z., Shen, J., Pan, Y. Normative database of word production of Putonghua-speaking children — Beijing articulahon norms project: CASS_CHILD_WORD. *2013 International Conference Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE)*, Gurgaon, India, 2013, pp. 1-6.

[17] Wong, P. 2018. Mothers do not enhance tonal contrasts in child-directed speech: Perceptual and acoustic evidence from child-directed Mandarin lexical tones. *The Journal of the Acoustical Society of America* 143, 3169.

[18] Li, X., Mok, P. P. K. 2020. The acquisition of tone sandhi of the Xiamen dialect. *Speech Prosody*, 479–483.

[19] Liu, Y., Li, A., Dang, J., Zhou, D. 2021. Semantic and acoustic-prosodic entrainment of dialogues in service scenarios. In *Companion Publication of the 2021 International Conference on Multimodal Interaction (ICMI '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 71–74.

[20] Pickering, M. J., Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27, 169–190.