# Production of Neutral Tone on Disyllabic Words by Two-Year-Old Mandarin-Speaking Children

Jun Gao and Aijun Li(✉)

Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China
9894575@qq.com, liaj@cass.org.cn

**Abstract.** This study examined the production of neutral tone in disyllabic words by two-year-old Mandarin-speaking children. The results showed that children were fully aware of the neutral tone sandhi rule phonologically at the age of two. However, they cannot phonetically produce neutral tone well. In particular, children made off-standard production with higher pitch register, wider pitch range and longer duration, while made correct production with correct pitch pattern but the duration ratio between the initial syllable and the final syllable is slightly larger than the adults'. The difficulty of the neutral tone production is closely related to the type of the preceding tone and the coordination of articulation for disyllabic neutral tone words.

**Keywords:** Neutral tone · Disyllabic words · Production · Two-years-olds Mandarin

## 1 Introduction

In Standard Chinese, apart from the four distinctive lexical tones, there exist weak elements in terms of neutral tone (Chao 1922, 1979), which is a special type of tone sandhi in Mandarin phonology. The neutral tone sandhi rule indicates that the neutral-tone syllable loses its lexical tone and becomes weak and short. The pitch contour of the neutral-tone syllable depends on the preceding tone within the same prosodic word. Even though words in neutral tone only take up a very small portion in the adults' lexicon, they appear in a relative higher percentage in children's lexicon. Studies on phonological development of children usually focused on the acquisition of consonants, vowels and lexical tones, with little attention paid to the acquisition of neutral tone. Li and Thompson (1977) reported that children (Taipei, in Taiwan) usually substituted full tone for neutral tone, especially for the neutral-tone words with affix -zi/tsɿ/. Zhu (2002) examined the acquisition of neutral tone of Mandarin-speaking children in Beijing in a more detailed way. Since neutral tone is pitch-related, it is interesting to know how the acquisition of neutral tone is related to the acquisition of lexical tones. The literature on Mandarin lexical tone acquisition showed that lexical

tone is acquired at around two years of age (Zhu 2002; Si 2006).[1] Hence, we were motivated to explore whether Mandarin-speaking children have acquired neutral tone at the age of two by examining the acoustic patterns of the disyllabic neutral-tone words in both F0 and durational dimensions. We also sought to discover how the acquisition of neutral tone is related to the acquisition of lexical tone and whether there are general principles in the acquisition of tonal combinations.

## 2 Neutral Tone in Mandarin

Neutral tone is related to both tone and stress system in Standard Chinese (Lu and Wang 2005). Neutral tone does not occur in the initial position of a word and is assumed to be associated with weak syllable, which is short and light. The disyllabic neutral-tone words follow a strong-weak stress pattern. The final syllable being perceived weak is mainly due to several factors: dropping of the original tone, reduced pitch range, and shorter duration (Li 2017). The duration of neutral-tone syllable is about 55%–66% of its preceding syllable (Li 2017). The perception of neutral tone is determined more by F0 than duration (Li and Fan 2015). Spectral tilt is also proved to play a role in perception, but it is less important than F0 and duration. Intensity of neutral tone syllable is not always weaker than that of full tone syllable (Lin and Yan 1980). Vowel reduction is not a reliable cue either, which is highly related to personal habit or dialect background (Lin 2012). The final part of the neutral tone syllable is likely to be voiceless and the voiceless initial part is likely to be voiced (Li 2017).

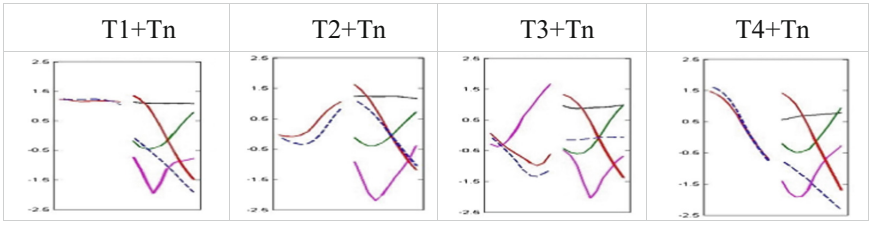


**Fig. 1.** The F0 contour patterns of all the two-tone combinations (dotted lines represent the neutral-tone combinations; solid lines represent full-tone combinations; Tn denotes four full tones and neutral tone). In the third panel, the high rising pitch contour on the first syllable is subjected to Tone3-Tone3 sandhi rule. (The F0 data are the averages of five males and five females normalized in Z-score. The X-axis is normalized duration.)

As shown by the dotted lines in Fig. 1, the neutral tone T0 is a falling tone behind Tone1 (T1, high-level tone), Tone2 (T2, high-rising tone), and Tone4 (T4, high-falling tone), while it is a mid-level tone behind Tone3 (T3, low-dipping tone). Regarding the

[1] Another study on lexical tone acquisition of our team showed that two-year-old Mandarin-speaking children still had around 13% tone errors and that the time of lexical tone acquisition was after 3.5 years of age.

falls on the neutral-tone syllable, it can be seen from Fig. 1 that there are some phonetic differences due to the offset of the preceding tone, i.e. a low fall following the high falling tone (the dotted line in the fourth panel), a mid fall following a high-rising tone (the dotted line in the second panel) and a half-low fall after a high-level tone (the dotted line in the first panel).

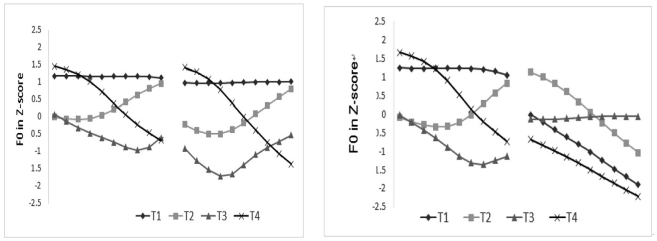


**Fig. 2.** The pitch contour patterns of all possible two-tone combinations on disyllabic prosodic words in Mandarin, full-tone combinations shown in the left panel, neutral-tone combinations shown in the right panel.

To have a clear picture of the pitch patterns of the tone combinations, we put all the full-tone combinations in one plot and all the neutral-tone combinations in the other plot as shown in Fig. 2. Again, it shows that the tonal contours of the second syllables are different in the two types of combinations: in full-tone combinations tonal contours of the second syllables maintain their citation form, but in neutral-tone combinations the second syllables exhibit a falling contour after T1, T2 and T4 and a mid level tone after T3. In addition, the pitch range of the second syllables is smaller in the case of neutral tone shown on the right panel than full tones on the left, and the pitch range of the first syllables in the neutral-tone words is larger than that in the full-tone words.

In Standard Chinese, neutral tone serves different lexical or morphosyntactic functions, as it can appear in about ten types of different contexts, such as some lexical items (lexeme type, such as /yi1 fu0/, “clothes”), stem-affix structures (affix type, such as -zi/tsɿ/), reduplications, locatives, directional complements, complement particle (de/tɤ/), aspect particles, particle (de/tɤ/), modal particles, and quantifier (ge/kɤ/). The last seven types are all related to grammatical morphemes. In the present study, we only focused on the acquisition of the first three types of neutral tone. These neutral-tone words are all disyllabic prosodic words.

## 3 Neutral-Tone Acquisition by Mandarin-Speaking Children

Zhu’s study (2002) aimed to investigate the phonological development of normally developing Mandarin-speaking children. The study investigated 129 Mandarin-speaking children in Beijing aged from 1;6 to 4;6. The results showed that 57% of the youngest group (1;6–2;0, 21 children in total) could produce neutral tone and that except the error of the deletion of the affix -zi(/tsɿ/) almost all the errors were associated with pitch level and duration. Specifically, among the neutral-tone types like lexical

item and reduplication, 93.4% of the errors were the substitution of the neutral tone by the citation tone. For example, the neutral-tone word "hair" /tou2 fa0/ was realized as /tou2 fa4/. According to Zhu, the high falling tone in /fa4/ – which was the underlying citation tone of the syllable /fa/ – substituted the neutral tone which should have been a mid fall with much shorter duration. But from our point of view, this substitution also could be the result of lifting the onset of the pitch contour and widening the pitch range, with a mid fall realized as a high falling tone. As a result, the mechanism underlying the error may be related to pitch register and pitch range instead of substitution by the underlying citation tone. To test which hypothesis is correct, we needed to examine children's production of the neutral-tone words whose neutral tone is transformed from a rising full tone. For example, for the neutral-tone word "sun" /tai4 ia.˙0/„ the neutral tone of the second syllable is a low fall after a high falling tone while the full tone of the second syllable is a high-rising tone2. If children produce the neutral tone of the second syllable as a high fall, it means the error mechanism is raising the pitch level and widening the pitch range instead of substitution of the neutral tone with a citation tone (high rising tone). More detailed analyses were needed to testify what mechanism underlies children's production errors of neutral tone.

In Zhu's study, there were two age groups including two-years-olds: 1;6–2;0 group (21 children) and 2;1–2;6 group (24 children). Only 13 "weakly stressed syllables" were examined. And there were no acoustic analyses reported on the error types (deletion, pitch level and duration). To see whether and how the two-year-olds acquire neutral tone, the acoustic patterns of more neutral words and more children were needed.

## 4 Data

We selected two-year-old children's production data of disyllabic neutral-tone words from a large-sample corpus (CASS_CHILD_Word, Gao et al. 2013). The corpus was the picture-naming production data of Mandarin-speaking children in the urban area of Beijing. The selected data were annotated in two rounds by both authors, using Praat (http://www.fon.hum.uva.nl/praat/).

379 tokens of 97 disyllabic neutral-tone words produced by 60 Mandarin-speaking children were analyzed. All of them were two years old, aged from 2;0(01) to 2;0(30), 26 boys and 34 girls. The disyllabic neutral-tone words covered three types: the lexeme type (132 tokens), the reduplication type (66 tokens) and the affix type (181 tokens). The reduplication type contained two subtypes, the reduplicative words already in the lexicon (40) like "dad" (/pa4pa0/), "mom" (/ma1ma0/), "baby" (/pao3pao0/), "star" (/xi.˙1 xi.˙0/), and the reduplicative words created by reduplicating monosyllabic words (26). In the affix type, most words were the words with the affix zi (163). In total, the data consisted of 77 T1 + T0 (T0 meant neutral tone) words, 126 T2 + T0 words, 55 T3 + T0 words and 121 T4 + T0 words.

All neutral-tone words are categorized into three groups based on the goodness of neutral-tone production: Correct, Off-standard and Non-neutral-tone. The error of the first tone was annotated as well. Table 1 illustrates the distribution of the neutral-tone production. If calculated by tokens, 59.5% neutral tones are correctly produced, 33.5%

**Table 1.** Distribution of neutral tone production

| | | Tokens (%) | Children (%) |
|---|---|---|---|
| Correct | | 59.5% (225/379) | 93% (56/60) |
| Error | Off-standard neutral tone | 33.5% (127/379) | 73% (44/60) |
| | Non-neutral-tone | 7% (27/379) | 35% (21/60) |

were produced as Off-standard and 7% were produced as Non-neutral-tone. If calculated by population[2], the statistics were 93%, 73% and 35% respectively.

F0 and duration of each tone were extracted and manually checked by the authors. F0 data were then normalized into 10 points for later analyses.

## 5 Results

### 5.1 Pitch

40.6% (154/379) of the data were judged wrong, i.e. not produced as neutral tone. These errors were made by 47 children, 22 boys and 25 girls. The errors could be divided into two groups, off-standard neutral tone and non-neutral-tone. For the off-standard-neutral-tone errors, the pitch movement of the final syllable followed the neutral-tone sandhi rule, but the pitch register was higher and the pitch range wider, which made the final syllable perceptually strong. This meant that children were aware of the sandhi rule. For the non-neutral-tone errors, the pitch movement of the final tone did not follow the neutral-tone sandhi rule. In these errors, some final syllables were produced in the citation tone, which made final syllables being perceived metrically strong. Other final syllables were produced as a level tone being perceived relatively weak.

**Off-Standard-Neutral-Tone Error**

127 word tokens (produced by 44 children, 21 boys and 23 girls) were judged as soff-standard-neutral-tone errors (82.5% of 154 errors). Among them, there were 27[3] Tone1-initial words (28.1% of 96 Tone1-initial words), 57 Tone2-initial words (45.8% of 117 Tone2-initial words), 22 Tone3-initial words (42.3% of 52 Tone3-initial words) and 21 Tone4-initial words (18.4% of the 114 Tone4-initial words).

It could be seen that off-standard-neutral-tone errors were more likely to occur in Tone2-initial and Tone3-initial disyllabic neutral-tone words (the right panel in Fig. 3). Compared to the correct productions (the left panel in Fig. 3), in off-standard-neutral-tone errors, for Tone2-initial disyllabic words (T2 + T0), the pitch contour on the final

[2] For some children, they could only have correct production on some neutral-tone words. As a result, the total of the children of the three production types exceeded 60.

[3] This was counted based on children's real production of the first tone rather than on orthography. Children made substitution errors on the first tone, so the number of Tone 1 on the first syllable was counted according to both correct production and substitutions. All the numbers hereafter were counted based on children's real productions.

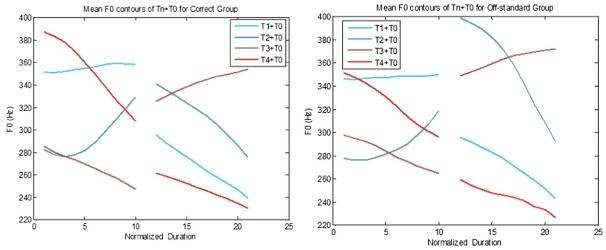


**Fig. 3.** The left panel is the mean F0 of the correct productions of Tn + T0 and the right panel is for those of the off-standard productions.

syllable had a higher onset and larger pitch range and for Tone3-initial disyllabic words (T3 + T0), the pitch contour on the final syllable had a higher register. Besides, T4 in T4 + T0 and T2 in T2 + T0 were not as steep as in the correct productions.

**Non-Neutral-Tone Error**

14 words (produced by 12 children, 6 boys and 6 girls) were produced as non-neutral-tones (9.1% of 154 errors) with the final syllable bearing level tone (the left panel in Fig. 4) despite whatever the initial tone was. The erroneously produced words were shirt, insect, wheel, baby, older sister, ear, moon, place and leopard. Of the 14 words, 6 words were produced with the first-tone substituted as Tone1.

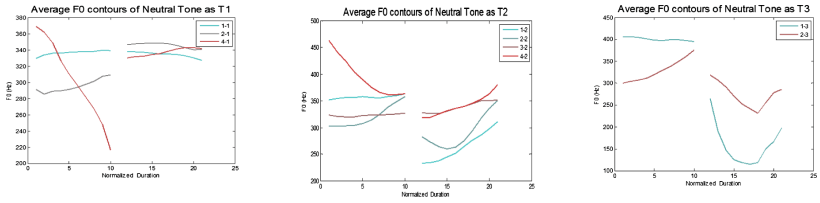


**Fig. 4.** The pitch contour patterns of the non-neutral-tone error with the final syllable bearing level tone Tone1 (left), high rising tone Tone2 (mid) and low rising Tone3 (right)

10 words (produced by 8 children, 5 boys and 3 girls) were produced as non-neutral tones (6.5% of 154 errors) with the substitution of the neutral tone by the citation Tone2 (the mid panel in Fig. 4), despite whatever the initial tone was. The reason why children produced neutral tone as Tone 2 might be that the final syllables of most of the 10 words (i.e. *cherry, lantern, clothes, steamed bun, sun*) carry a high rising tone when produced in isolation.

3 words (produced by 3 children, 1 boy and 2 girls) were produced (1.9% of 154 errors) with the neutral tone produced as Tone 3 (the right panel in Fig. 4). The reason for this wrong production might be that children added pragmatically a final successive boundary tone to the neutral tone. A fall neutral tone followed by a low rise made it perceived as Tone3. The initial tones of the three words included two in Tone1 and one in Tone2.

**Errors on the Initial Tone**

Fewer errors occurred on the initial tones. There were two kinds of errors in the initial tones. One was the off-standard production and the other was substitution. The initial Tone2 and Tone4 were more likely to be subjected to off-standard production and substitution as Tone1.

Even though children produced the initial tone wrongly into another tone, on most occasions they had the right neutral tone (or at least the right pitch contour) on the final syllable according to the 'new' initial tone. As such, in the off-standard cases of the initial tone, children could still produce the neutral tone on the final syllable correctly. These indicated that children were full aware of the neutral tone sandhi rule.

In these disyllabic words, the initial Tone2 was rarely produced wrongly as Tone3 and the initial Tone3 was rarely produced wrongly as Tone2. These were different from the cases of the monosyllabic tones where Tone2 and Tone3 are the frequent substitution targets for each other (according to the results of another study of ours).

**Correct Productions of Neutral Tone**

Figure 5 showed the pitch patterns of children's correct production of neutral tone grouped by the first tone. The mean contours were shown in the left panel of Fig. 3. Table 2 presented the maxima and minima of mean F0 of the first syllables and the neutral-tone syllables for correct and off-standard neutral-tone words (see Fig. 3).

In terms of neutral-tone syllables, the tonal range of the off-standard neutral-tone tokens was 48 Hz wider and the tonal register was 9 Hz higher than those of the correct neutral-tone tokens. In terms of the first syllable, the tonal range of the correct neutral-tone tokens was 54 Hz wider and the tonal register was 9 Hz higher than those of the off-standard neutral-tone tokens. The pattern of the Correct group was similar to the adults' production patterns (see Fig. 2).

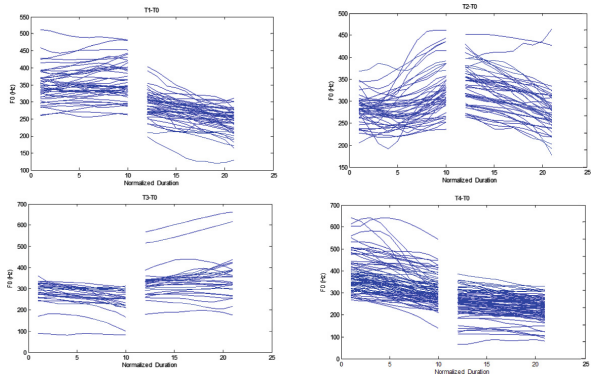


**Fig. 5.** Correct production of neutral tone grouped with the first tone.

**Table 2.** Mean F0(Hz) of Correct and Off standard productions.

| | CorrSyll1 | CorrNeu | OffSyll1 | OffNeu |
|---|---|---|---|---|
| F0max | 387 | 354 | 351 | 399 |
| F0min | 247 | 230 | 265 | 227 |
| F0average | 317 | 292 | 308 | 313 |
| F0range | 140 | 124 | 86 | 172 |

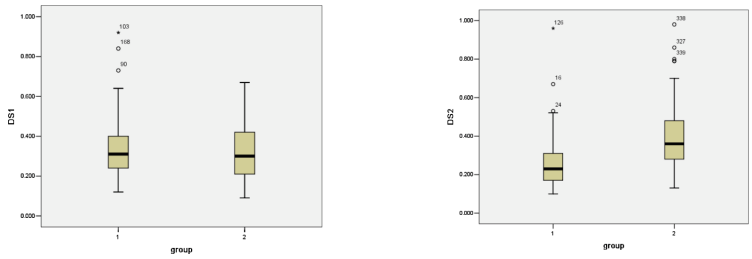


**Fig. 6.** Duration distribution for first syllable (DS1) and neutral tone syllable (DS2) for two groups. Group1 stands for the correct productions and group2 for the off-standard neutral-tone productions.

### 5.2 Duration

Duration distributions of the Correct and Off-standard groups were plotted in Fig. 6.

**Correct Production of Neutral Tone**
For correct productions of neutral tone, the mean duration of the first syllable was 331 ms (Sd. 129 ms) and that of the second syllable was 249 ms (Sd. 106 ms). The second syllable was shorter than the first syllable. On average, the duration of the final syllable was 0.86 of the first syllable. The ratio was larger than the ratio in adults (0.55–0.66) between the initial and the final syllables in disyllabic neutral-tone words.

**Off-Standard Production of Neutral Tone**
For off-standard productions, the mean duration of the first syllable was 324 ms (Sd. 130 ms) and that of the second syllable was 394 ms (Sd. 165 ms). The second syllable was longer than the first syllable. The duration of the first syllable in incorrect productions was similar to that in correct productions. But the second syllable in incorrect productions was much longer than that in correct productions. On average, the duration of the final syllable was 1.43 of the initial syllable. It meant that the final syllable was much longer than the first syllable.

ANOVA analysis showed that the durations of the first syllable were not significantly different between two groups ($F = 0.214$, $p = 0.64$), but the durations of the neutral tone syllable were significantly different between two groups ($F = 0.050$, $p = 0.00$).

### 5.3 Summary

Out of 97 words in the experiment, “ear” /ǁr3tou0/ and “eye” /ya.ˑ˙3tɕi.ˑ˙0/ are more likely to be produced with off-standard-neutral-tone error, as the pitch on the final syllable had a higher register. At the same time, “pomegranate”/⟦ 2liǁu0/, “grape”/p‿u2t‿ao0/, “hair”/t‿ou2fa0/ and “skirt”/tɕ‿i .ˑ˙2ts 0/ are more likely to be produced with off-standard-neutral-tone error, as the pitch on the final syllable had a much higher onset and wider pitch range.

In another study (Gao et al. 2017) on the acquisition of lexical tones by Mandarin-speaking children, we showed that the difficulty ranking of lexical tones on monosyllabic words is T3 > T2 T1, > T4. Low-dipping tone is the most difficult tone, followed by rising and level tone; falling tone is the easiest tone. T3 is commonly mispronounced as T2 or produced in a non-canonical way. T2 is usually produced wrongly as T3. T1 is usually produced wrongly as T4 or T2. T4 is usually produced wrongly as T1 or produced in a non-standard way.

Comparing the production patterns of neutral tone and lexical tones, we concluded that children were fully aware of the pitch movement patterns in neutral tone at two years of age. They knew how the pitch of the neutral tone was going based on the preceding tone, because even when they incorrectly produced the previous tone into another tone, they produced the right pitch contour on the neutral tone in accordance with the ‘new’ preceding tone. And children also could apply the neutral-tone sandhi rule to the new reduplicated words which they had never heard, which were created by reduplicating a monosyllabic word.

Even though they are fully aware of the neutral-tone rule phonologically, they cannot produce it well enough phonetically. Children still need improvement on pitch register, pitch range and syllable duration in their productions of neutral tone. One reason might be in articulation. Children’s articulatory organs and the coordination between these organs have not reached the level of maturity as seen in adults. They tend to produce higher pitch onset and pitch register, wider pitch range and lengthened duration than adults in neutral-tone syllables. Another reason might be due to the fact that the child-directed speech is slower with higher register and wider pitch range.

The reason that Tone2(LH)-initial and Tone3(LLH)-initial neutral-tone combinations are more prone to have a higher onset or register and wider pitch range of the pitch on the final syllable might be that the H target of the initial tone is realized on the second syllable. The peak delay is quite common when neutral tone is involved (Li 2003; Chen and Xu 2006).

In our data, compared to errors in consonants and vowels, lexical tones on the first syllable had fewer errors. This means that lexical tones are acquired earlier than consonants and vowels. For lexical tones, even though children still made errors on monosyllabic syllable (according to the results of another study of ours) and some on the initial tone, they generally showed good awareness of the neutral-tone rule. It seems that rules are acquired earlier than tones. This may be due to the fact that the domain within which neutral-tone sandhi rule applies is prosodic word, to which children are very sensitive in their language acquisition process.

**Acknowledgements.** The study is supported by two grants from the Key Project of National Fund of Social Sciences, No. 15ZDB103 and No. 13CYY025.

## References

Chao, Y.R.: Gwoyeu Romatzyh or the national romanization. In: Wu, Z.J. (ed.) Linguistic Essays by Yuenren Chao, pp. 61–72. The Commercial Press, Beijing (1922)

Chao, Y.R.: Beijing Kouyu Yufa (A Grammar of Spoken Chinese). The Commercial Press, Beijing (1979)

Zhu, H.: Phonological Development in Specific Contexts: Studies of Chinese-Speaking Children (Child Language and Child Development). Multilingual Matters Ltd., Clevedon (2002)

Si, Y.: A case study of the acquisition of speech sounds by a Putonghua-speaking child. Contemp. Linguist. **8**(1), 1–16 (2006)

Lu, J.L., Wang, J.L.: Guanyu qingsheng de dingjie (Analysis on the nature of neutral tone. Contemp. Linguist. **7**(2), 107–112 (2005)

Li, A.: Phonetic correlates of neutral tone in different information structures. Contemp. Linguist. **19**(3), 348–378 (2017)

Li, A., Fan, S.: Correlates of Chinese neutral tone perception in different contexts. In: Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015), Glasgow, Scotland (2015)

Li, C.N., Thompson, S.A.: The acquisition of tone in Mandarin-speaking children. J. Child Lang. **4**(2), 185–199 (1977)

Lin, M.C., Yan, J.Z.: Beijinghua Qingsheng de Shengxue Xingzhi. Dialect **3**, 166–178 (1980)

Lin, M.: Experimental Study on Chinese Intonation. China Social Sciences Press, Beijing (2012)

Li, Z.: A perceptual account of asymmetries in tonal alignment. In: Kadowaki, M., Kawahara, S. (eds.) Proceedings of the North East Linguistic Society 33, pp. 147–166. UMass GLSA, Amherst (2003)

Chen, Y.Y., Xu, Y.: Production of weak elements in speech evidence from F0 patterns of neutral tone in standard Chinese. Phonetica **63**, 47–75 (2006)

Gao, J., Li, A., Xiong, Z., Shen, J., Pan, Y.: Normative database of word production of Putonghua-speaking children - Beijing articulation norms project: CASS_CHILD_WORD. In: Proceedings of Oriental-COCOSDA 2013 (2013)

Gao, J., Li, A., Zhang, Y.: The production of tones in monosyllabic and disyllabic words by Mandarin-speaking children. In: Proceedings of the ISSP, OC, Tianjin, pp. 16–19 (2017)

[This paper was published at Springer Nature Switzerland AG 2018]