Acquisition of Two Consecutive Neutral Tones in Mandarin-Speaking Preschoolers: Phonological Representation and Phonetic Realization

Sichen Zhang[1], *Aijun Li*[1 2]

[1]University of Chinese Academy of Social Sciences (UCASS)
[2]Corpus and Computational Linguistics Center, Institute of Linguistics, CASS
zhangsichen@ucass.edu.cn, liaj@cass.org.cn

Abstract

Building on the description of acoustic patterns of neutral tone produced by Mandarin-speaking adults in literature and the corresponding proposals of phonological-phonetic mapping mechanism, the present study examined the production of two consecutive neutral tones by Mandarin-speaking preschool children. Data was extracted from spontaneous conversation between preschoolers and their caretakers in CASS-CHILD, a child language corpus of Mandarin-speaking children. Results show the developmental process of consecutive tones production in terms of F0 contour, pitch range, and duration. Our findings are in accord with established research on the early mastery of tone features and synchronous development among lexical tone categories in prosody acquisition. Furthermore, it is revealed from this study that preschoolers perform well on acquiring phonological representation of neutral tone and sandhi. Even so, they may have difficulties in accurately realizing consecutive neutral tones in the following conditions: 1) when the preceding full syllable is Tone 2 or Tone 3, which in themselves trouble young speakers; 2) when word prosody interacts with phrase-level prosody, for example, domain-final lengthening at intonational phrase boundaries. In addition, acoustic parameters, pitch and duration in our case, interfere one another in preschoolers' off-standard realization of consecutive neutral tones.

Index Terms: neutral tone, prosody acquisition, phonological representation, phonetic realization, Mandarin Chinese

1. Introduction

In word prosody, Mandarin Chinese distinguishes four tonal phonemes in full syllables: high level Tone 1 (H), rising Tone 2 (LH), dipping Tone 3 (L), and falling Tone 4 (HL). Additionally, in multisyllabic words, a weak and unstressed syllable is said to be in neutral tone, as mentioned in [1]. The phonetic realization of neutral tone, perceived as short and lax, as well as having varied pitch contours due to the lexical tone in the immediately adjacent preceding full syllable, has been well researched in the literature, as in [2]-[5]. When it comes to the phonological representation and the mapping from phonology to phonetics, discussions on whether neutral tone has a phonologically specified tone target have carried on for decades. On the one hand, neutral tone has been prevailingly considered as lacking distinctive tonal identity in phonology. Its surface realization is derived from interpolation between tone targets of the adjacent full syllables as identified by [6]-[9]. Alternative views propose that neutral tone has an independent tone target, which is, for example, low as in [10] or static and mid as in [11]. In either case, the F0 contour of neutral tone is realized via target approximation. Therefore, regarding the production of neutral tone, it is necessary to examine both the mastery of phonological representation and the performance on phonetic realization.

With respect to prosody acquisition, previous studies focused on auditory description and/or acoustic analysis as [12]-[14], and have revealed the following facts: 1) children acquire tone features at a very early age and in advance of segments; 2) Tone 2 and Tone 3, due to their acoustic properties and perceptual salience, are typically acquired later than Tone 1 and Tone 4; 3) children perform differently on neutral-tone syllables in different morphosyntactic functions. For example, Li and Thompson [12] notes that children speaking Taiwan Mandarin frequently substitute the neutral tone in suffix *zi* with Tone 3. Zhu [13] reports the tendency of deleting the syllable of neutral-tone affix among children aged 1;6-2;0. Similarly, producing a neutral-tone syllable in its citation form is a common error. Other than these conclusions regarding acquisition of the phonology of tones, scholars also shed light on the phonetic features realized in children's tone acquisition. Zhu [13] points out that most errors are associated with pitch level and duration, which has been echoed by [14-16] in their research.

To date, although neutral tone acquisition has received much attention, the production of consecutive syllables in neutral tone by Mandarin-speaking children has not been examined. Nevertheless, as analyzed in a more recent study [8], neutral tones occurring in succession provide us with a lens to inspect the phonology-phonetics interface in the process of phonetically realizing the phonological representation of tones. The present study sets out to investigate the production of consecutive neutral tones in Mandarin-speaking preschoolers. The research objectives are as follows:

1) to observe the developmental patterns of the phonetic realization of consecutive neutral tones in terms of pitch movement, pitch range, and duration;

2) to explore the interface between children's representation of neutral tone in phonology and their performance on phonetic realization.

2. Methods

2.1. Participants

Data used in this study were drawn from CASS_CHILD, a multimedia Mandarin child speech corpus established at Chinese Academy of Social Sciences (CASS) [17]. Participants of CASS_CHILD were 23 Mandarin-speaking children (13 males and 10 females), ranging from age of 1;6 to slightly over 6;0, as well as their caretakers, including their parents, grandparents, and babysitters. All participants are native

Mandarin speakers, and currently live in Beijing. Two thirds of the families are Beijing locals, and the rest exhibit mild dialectal influence. Participants have not been diagnosed with any speech/hearing disorders.

2.2. Procedure

Recording was conducted in a sound-proof room (3.25 m*3.5 m*2.4 m) decorated as child-friendly playroom with plenty of toys, books, and flashcards. Children and caretakers were asked to interact freely over the props prepared. Meanwhile, their conversations were collected by wireless pin microphones (AKG SR400/SR40) clipped 15 cm from mouth on their collars. Speech data were sampled at 22 kHz with a 16-bit resolution in stereo two channels (children's speech in the right channel and caretakers in the left).

2.3. Data analysis

All stereo files were converted into mono channels with Cool Edit Pro 2.0, processed by automatic segmentation program to generate syllable- and phone-level transcriptions, manually checked later by well-trained annotators using Praat [18]. Children's utterances containing consecutive neutral tones were selected, and further narrowed down to those containing two neutral tones, in which the second one is the particle *de*. *De* may function as a possessive particle, like in *māma de* ("mother's"), or a nominalizer, like in *nǎipíng shang de* ("that on the milk bottle"). The number of syllables in the prosodic word preceding the neutral tones varied from 1 to 3. 121 tokens in the tonal sequence of TnN1N2 at the intonational-phrase (IP) boundary were obtained, where Tn stands for the stressed full syllable carrying Tone 1 (H), Tone 2 (LH), Tone 3 (L), or Tone 4 (HL); N1 for the first neutral tone; and N2 for the second, that is, *de*. F0 values were taken at ten points in equal intervals of the syllable finals, and durations of the final in every syllable were extracted, both using Praat scripts.

It is worth noting that the distribution of the data was rather scattered in spontaneous conversations, which means very few utterances had the same tonal combinations, and the numbers of utterances contributed by each speaker were different. Even if several utterances were produced by the same child, they could be recorded at their different ages. What's more, a considerable number of tokens contained creaky voice, resulting from the IP boundary position, breathy voice, a frequently-occurring phonation type in children speech, as well as overlaps need to be cleaned. Therefore, the distribution of data presented significant difficulties in statistical analysis. Our results and analysis were organized and presented based on these considerations.

3. Results

3.1. Perceived accuracy of the preschoolers' production of two consecutive neutral tones

Preschool children's production of the TnN1N2 sequences was assessed based on native-speakers' auditory judgments in terms of 1) the accuracy of the phonological representation, i.e., whether the syllables in neutral tone underwent neutral-tone sandhi, losing their citation tone and varying their pitch contours according to the lexical tone in the preceding full syllable; 2) the accuracy of the phonetic realization, i.e., whether the neutral tones were realized short and lax, with correct acoustic properties of duration and pitch, including pitch movement, pitch register, and pitch range as defined in [14]. Results were grouped by tone categories, as shown in Table 1, and by age at 6-month intervals, as shown in Table 2. The percentage numbers in the parentheses indicate the percentage taken by the tokens in each tone or age group.

Table 1: *Errors in the production of two consecutive neutral tones by tone categories of the preceding full syllable*.

Tn	Tokens	Phonologically Incorrect	Phonetically Inaccurate
T1	49	0	1 (2.0%)
T2	27	3 (11.1%)	3 (11.1%)
T3	15	0	3 (20.0%)
T4	30	0	0
Total	*121*	*3 (2.5%)*	*7 (5.8%)*

Table 2: *Errors in the production of two consecutive neutral tones by age group.*

Age group	Tokens	Phonologically Incorrect	Phonetically Inaccurate
1;6-1;11	6	0	0
2;0-2;5	17	0	2 (11.8%)
2;6-2;11	24	1 (4.2%)	2 (8.3%)
3;0-3;5	20	0	2 (10.0%)
Younger	67	1	6
3;6-3;11	12	0	0
4;0-4;5	24	1 (4.2%)	0
4;6-4;11	6	0	1 (16.7%)
5;0-5;5	6	1 (16.7%)	0
5;6-5;11	5	0	0
6;0+	1	0	0
Older	54	2	1
Total	*121*	*3 (2.5%)*	*7 (5.8%)*

As can be seen from Tables 1 and 2, errors in the phonological representation of consecutive neutral tones rarely occurred, only accounting for 2.5% of the data. However, inaccuracy in the phonetic realization occurred at a much higher rate of 5.8%. With respect to tone categories of Tn, the preceding full syllable, incorrect phonological representations of neutral tones were only found after Tone 2 (LH) and Tone 3 (L). Failures in phonetic controlling also concentrated in the Tone 2 and Tone 3 groups, contributing 85.6% of the total failures. Regarding the developmental features, inaccurate productions mostly occurred in age groups under 3;6. Thus, we roughly regroup the children into two: the younger group from 1;6 to 3;0, and the older group from 3;0 to 6;0+.

3.2. F0 contours of the preschoolers' production of two consecutive neutral tones

Taking a step further from auditory judgments, in this section, we present the analysis of F0 contours of two consecutive syllables in neutral tone produced by preschool children in comparison to adults' production as reported in [8]. Given the scarcity of phonological errors, hereafter we focus on the tokens with correct phonological representations, which include "good" tokens realized with accurate phonetic features and "off-standard" ones.

Figures 1 and 2 respectively visualize the good productions of the TnN1N2 sequence by preschool children in the younger and older groups. The F0 data of each token were normalized into a five-points scale. Z-scored normalization were not

adopted due to the restriction of data on normalizing for each speaker.

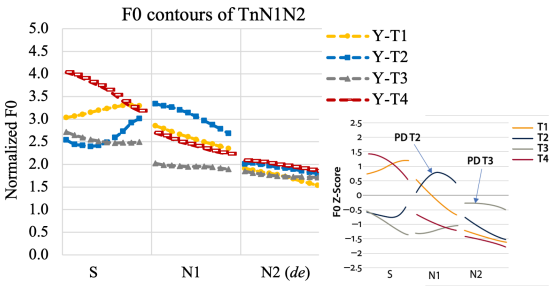

Figure 1: *F0 contours of the younger age group's good productions of the TnN1N2 sequence (the lower-right panel is the adults' patterns given by [8, p. 60]).*

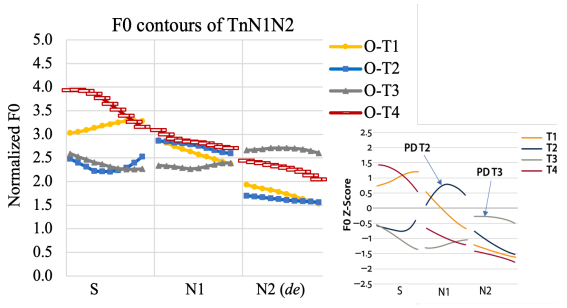

Figure 2: *F0 contours of the older age group's good productions of the TnN1N2 sequence (the lower-right panel is the adults' patterns given by [8, p. 60]).*

As seen from Figures 1 and 2, the general F0 contours in the T1N1N2, T2N1N2 and T4N1N2 sequences produced by both preschooler groups were in line with adults' patterns. In these sequences, pitch values converged to a mid-tone target at the end of N1, and to a low tone target at the end of N2. A continuous falling can be observed from N1 to N2. To elaborate, the F0 peak of Tone 1 was realized within the syllabic boundary of the full syllable, labeled as S. The peak of Tone 2 was delayed and realized in N1, but F0 failed to achieve a tone target as high as expected in adults' production according to [8]. As for the T3N1N2 sequence, the younger group failed to realize a high tone target associated with the delayed peak of S in N2. The F0 contour gradually declined from the beginning of S to the end of N2. In contrast, the F0 of N2 produced by the older group showed a post-low raising from N1, which was the realization of the delayed high tone component of S.

Among the phonetically off-standard productions of the TnN1N2 sequence, it is necessary to conduct the analysis in a token-by-token manner, as plotted in Figure 3.

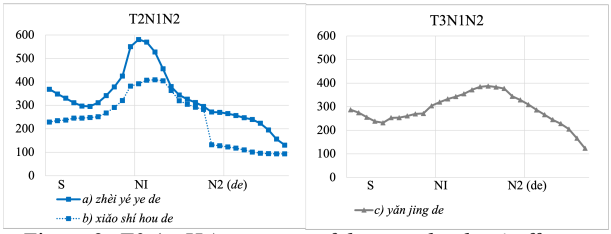

Figure 3: *F0 (in Hz) contours of the preschoolers' off-standard productions of the TnNIN2 sequence.*

These tokens shared a context in common, where the nuclear accent, that is, sentence stress, fell on the TnN1N2 prosodic word. Instead of stressing the full syllable by raising/lowering the high/low tone target and expanding the pitch range, the preschool speakers chose to stress N1. As can be seen in Figure 3, the pitch range of N1 was enlarged in both T2N1N2 and T3N1N2 sequences. The latter token even displayed an F0 contour divergent from that of good, or adult-like production. The high tone target of F0 peak in Tone 3 was realized in N1 rather than N2, and the low tone target of neutral tone was reached at the end of N2.

3.3. Pitch range of the preschoolers' production of two consecutive neutral tones

The token-by-token analysis of off-standard productions in Section 3.2 leads us to analyze the realization of pitch range. Considering the complexity for doing individual normalization as mentioned in Sections 2.3 and 3.2, the patterns of pitch range were normalized through calculating the ratio of the F0 range of N1 and that of N2 respectively to the F0 range of preceding full syllable(s) in the same prosodic word, labelled as N1: S(s) and N2: S(s) below in Table 3, Figure 4, and Table 4.

Table 3: *Pitch range patterns of the preschoolers' good productions of the TnN1N2 sequence.*

Ratio / Sequence	N1: S(s)	N2: S(s)
T1N1N2	74.7%	52.1%
T2N1N2	99.2%	80.4%
T3N1N2	75.2%	107.0%
T4N1N2	78.3%	61.3%

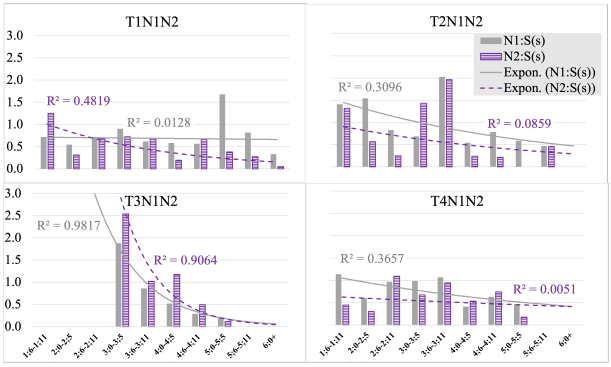

Figure 4: *Pitch range patterns of the preschoolers' good productions of the TnN1N2 sequence.*

As listed in Table 3, a gradual compression in pitch ranges from S to N1 and to N2 was a general tendency in the T1N1N2 and T4N1N2 sequences. In the T3N1N2 sequence, the F0 peak delay of Tone 3 was realized in N2 with an expanded pitch range. Likewise, the pitch range of N1 in the T2N1N2 sequence, which carried the delayed F0 peak of Tone 2, was hardly compressed compared to S. With respect to the developmental patterns displayed in Figure 4, pitch range features of the T1N1N2 and T4N1N2 sequences were relatively stable in preschoolers' production, with a few exceptions that could be attributed to the small amount of data in certain age groups. In the T2N1N2 and T3N1N2 sequences, the falling tendencies of pitch range ratios were observed along with the growth and development of preschool children.

Table 4: *Pitch range patterns of the preschoolers' off-standard productions of the TnN1N2 sequence.*

Token \ Ratio	N1: S(s)	N2: S(s)
a) *zhèi yé ye de*	220.2%	108.8%
b) *xiǎo shí hou de*	137.6%	42.7%
c) *yǎn jing de*	110.2%	289.1%

For the off-standard tokens analyzed in Section 3.2, the pitch ranges of neutral-tone syllables which realized the delayed F0 peak of S, which means N1 in T2N1N2 and N2 in T3NIN2, were greatly expanded. This is a typical strategy used in the preceding full syllable to stress the nuclear accent in adult speech.

3.4. Duration of the preschoolers' production of two consecutive neutral tones

In this part, the duration patterns of two consecutive syllables in neutral tone produced by preschool children is observed through calculating the ratio of the vowel length of N1 and that of N2 respectively to the average vowel length of preceding full syllable(s) in the prosodic word, notated as N1:S(s) and N2:S(s) below. Additionally, the same values of two consecutive neutral tones produced by adults were shown as a reference. The data were from [19] and the structure of two consecutive neutral tones sequence was kinship terms, which were reduplicated with the second syllable unstressed, followed by the possessive particle *de*. Results are illustrated in Figure 5.

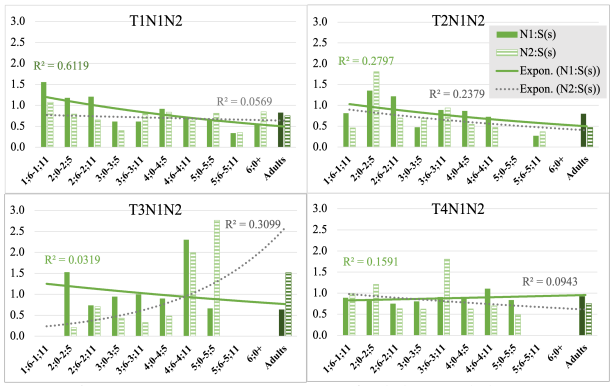

Figure 5: *Duration patterns of adults' and the preschoolers' good productions of the TnN1N2 sequence.*

As shown in Figure 5, most columns stood beneath the ratio of 1.0, and most columns representing N2 were shorter than N1. This reveals a stepwise declining tendency from S to N1 to N2. However, as indicated by the adults' pattern of the T3N1N2 sequence, N2 was produced longer in duration with the realization the F0 peak delay of S. Along with the growth of age and development of linguistic abilities, the duration ratio of N1 and N2 to S(s) continues to decline. The slope of declination appears to be rather flat because the sequences were at an IP boundary, and hence influenced by the boundary-related effects of boundary tone and final lengthening. Besides, in the T3N1N2 sequence, the N2:S(s) ratio displayed an increasing tendency, showing that preschoolers were developing towards the adult-like realization of this duration feature.

4. Discussion and Conclusions

The present study set out to investigate the production of two consecutive syllables in neutral tone by Mandarin-speaking preschool children. The structure of TnN1N2, in which N2 is a particle *de* functioning as a possessive marker or nominalizer, was selected as the target sequence. Utterances with this sequence that occurred at the intonational-phrase boundary were extracted from an established child language corpus. We were interested in the developmental characteristics in children's acquisition of tones, and the underlying explanations.

According to the results presented above, firstly, the findings in this study are consistent with established knowledge on the acquisition of word prosody in the perspective of phonology. The phonological representation of neutral tone, like other features in word prosody, is mastered by children at a very early age. Phonological errors such as mispronounced tones or wrong sandhi rules were rare in our results since the age of 1;6, known as the word-spurt phase.

Despite the generally successful acquisition of the phonological representation, there are gaps between the preschoolers' phonetic realization and adult-like acoustic properties. It is found in our study that preschool children, especially those in the younger group, had difficulty in realizing the delayed F0 peak on N2 in the T3N1N2 sequence. Their production of N1 in T2N1N2 was also off standard, and in this case, failed to achieve a high enough tone target in the middle of the syllable as in the adults' pattern. This finding confirms previous conclusion in [14] that Tone 2 and Tone 3 are acquired later than Tone 1 and Tone 4. By the age of 1;6, even though Tone 3 and Tone 4 productions sounded good in isolation, some children would lose their control in the process of phonetic realization when the phonological context became complex. The acoustic properties in the "off-standard" examples include F0 contours (see Figure 1 and Figure 2), pitch range (see Figure 3), and duration (see Figure 4).

Furthermore, this study also observed the influence of phrase-level prosody on the production of word prosody. As mentioned in Section 3.4, final-lengthening effect plays a role in the developmental patterns of children's production of the duration of neutral-tone syllables. Further research examining the interaction of word prosody and phrase-level prosody is already in progress.

Additionally, it would have been ideal if this study was not restricted by the small amount and scattered distribution of data. Besides, as [20] puts forward, neutral tone is a continuum with internal variation of acoustic properties. Further explorations on children's production of neutral tones in various morphosyntactic structures would be necessary.

In conclusion, the present study investigated Mandarin-speaking preschool children's production of two consecutive neutral tones. The results contribute to our understandings of the developmental features in preschool children's tone acquisition, as well as the underlying phonology-phonetics mapping process.

5. Acknowledgements

This work was supported by the Independent Topic Selection Project of Cultural Experts, "Four Batches" of Talents, and the CASS Innovation Project awarded to Aijun Li. We would also like to thank our colleague Dr. Jun Gao for the construction of CASS_CHILD and Dr. Zhiqiang Li for the comments.

6. References

[1] Y-R. Chao, *A Grammar of Spoken Chinese*. Berkeley: University of California Press, 1968.

[2] Y-R. Chao, *Mandarin Primer: An Intensive Course in Spoken Chinese*. Cambridge: Harvard University Press, 1948.

[3] M. Lin and J. Yan, "Beijinghua qingsheng de shengxue xingzhi [Acoustic characteristics of neutral tone in Beijing Mandarin]," *Fangyan [Dialect]*, vol. 3, pp. 166–178, 1980.

[4] M. Lin, *Hanyu yudiao shiyan yanjiu [The experimental study of intonation in Mandarin Chinese]*. Beijing: China Social Sciences Press, 2012.

[5] M. Yip, *Tone*. Cambridge, UK: Cambridge University Press, 2002.

[6] M. Yip, "The tonal phonology of Chinese," Ph.D. dissertation, Dept. Linguistics and Philosophy, MIT, Cambridge, MA, 1980.

[7] Z. Li, "The phonetics and phonology of tone mapping in a constraint-based approach," Ph.D. dissertation, Dept. Linguistics and Philosophy, MIT, Cambridge, MA, 2003.

[8] A. Li and Z. Li, "Prosodic realization of tonal target and F0 peak alignment in Mandarin neutral tone," *Language and Linguistics*, vol. 23, no. 1, pp. 47–81, Jan. 2022.

[9] J. Lu and J. Wang, "Guanyu qingsheng de jieding [Analysis on the Nature of Neutral Tone]," *Dangdai Yuyanxue [Contemporary Linguistics]*, vol. 2, pp. 107–112, 2005.

[10] N. Sun, "F0 features of tone and boundary tone and their phonetic realization in Mandarin Chinese," Ph.D. dissertation, Dept. Linguistics, Beijing, Chinese Academy of Social Sciences, 2006.

[11] Y. Chen and Y. Xu, "Production of weak elements in speech–Evidence from f0 patterns of neutral tone in standard Chinese", *Phonetica*, vol. 63, pp. 47–75, 2006.

[12] C. N. Li and S. A. Thompson, "The acquisition of tone in Mandarin-speaking children," *Journal of Child Language*, vol. 4, no. 2, pp. 185–199, Jun. 1977.

[13] H. Zhu, *Phonological Development in Specific Contexts: Studies of Chinese-speaking Children*. Clevedon, UK: Multilingual Matters Ltd, 2002.

[14] J. Gao and A. Li, "Production of neutral tone on disyllabic words by two-year-old Mandarin-speaking children," in *Studies on Speech Production: 11th International Seminar, ISSP 2017, Tianjin, China, October 16–19, 2017, Revised Selected Papers*, Q. Fang, J. Dang, P. Perrier, J. Wei, L. Wang, and N. Yan, Eds., Cham, Switzerland: Springer, 2018, pp. 89–98.

[15] R. Shi, J. Gao, A. Achim, and A. Li, "Perception and representation of lexical tones in native Mandarin-learning infants and toddlers," *Frontiers in Psychology*, vol. 8, Jul. 2017.

[16] N. Rhee, A. Chen, and J. Kuang, "Going beyond F0: The acquisition of Mandarin tones," Journal of Child Language, vol. 48, no. 2, pp. 387–398, 2021.

[17] J. Gao, A. Li, and Z. Xiong, "Mandarin Multimedia Child Speech Corpus: CASS_CHILD," in *2012 Int. Conf. on Speech Database and Assessments*, Macau, China, Dec. 2012, pp. 7–12.

[18] *Praat: Doing Phonetics by Computer*. (2020, Version 6.1.54). P. Boersma and D. Weenink. Accessed: Oct. 10, 2021. [Online]. Available http://www.praat.org/

[19] A. Li, Z. Li, G. Huang, and L. Zhang, "Tonal target and peak delay in Mandarin neutral tone," in *Oriental COCOSDA 2018*, Miyazaki, Japan, May 2018.

[20] J. Huang and F. Shi, "Hanyu qingsheng yinjie biaoxian de duoyangxing [The diverse representation of Chinese neutral-tone syllables]," *Yuyan Wenzi Yingyong[Applied Linguistics]*, vol. 1, pp. 76–85, Feb. 2019.

[This paper was published at Interspeech 2022, 3108-3112]