Dynamic Prosodic Grouping—Evidence from Recitation of Classical Chinese Poetry

Jiangxin Zhang, Zhuyin Feng, Yi Xu

University College London

jiangxin.zhang.24@ucl.ac.uk, zhuyin.feng.24@ucl.ac.uk, yi.xu@ucl.ac.uk

Abstract

Traditionally, prosodic grouping in speech is assumed to be based on a static hierarchical structure. But recent evidence suggests that prosodic grouping can be altered dynamically by various factors. The present study examines a case where the prosodic structures are supposedly stable: classical Chinese poetry. We asked native speakers to recite five- or seven-syllable poems in three styles ranging from poetic to conversational. We estimated the grouping structure through boundary strength measured by the duration ratio of the current syllable over the preceding one. The results show that boundary strength is drastically altered by both speech style and line length. As speech style becomes less poetic and line length becomes longer, some of the boundaries virtually disappeared, resulting in the merger of adjacent groups. The finding suggest that prosodic grouping is dynamically determined by multiple factors, rather than by a prosodic hierarchy directly projected from morpho-syntactic structures.

Index Terms: Speech rate, Sentence length, Boundary strength, Prosodic grouping

1. Introduction

According to the widely influential Prosodic Hierarchy Theory [1][2][3], speech is organized into a set of discrete and hierarchically arranged prosodic constituents such as syllables, feet, prosodic words, phonological phrases, and intonational phrases. These units are relatively stable once mapped from syntactic and morphological structures. As a result, prosodic grouping is structurally invariant. A critical issue with this framework, however, is that the determination of the prosodic units, in practice, is rarely based on experimental phonetic evidence. This is especially problematic with units like the foot and prosodic word, whose boundaries have no generally agreed phonetic markers. And the problem is particularly severe with the foot for languages like Chinese that lack lexical stress.

1.1 The Concept of Foot

Since the 1990s, there has been a marked increase of interest in prosodic rhythm in Chinese linguistics. One of the core concepts explored in this line of research is the 'foot' [4][5]. The concept can be traced back to ancient Greek music theory. In the fifth century BCE, Damon of Athens used bêma (step) metaphorically for rhythmic units, while Aristoxenus, in Elements of Rhythmics, described the opposition between arsis (lifting) and thesis (lowering), laying the foundation for a metrical theory. Later, the foot became central to Western poetics, particularly in the rise of iambic pentameter. Shakespeare brought it to its peak, as exemplified in the opening line of Sonnet 18, 'Shall I compare thee to a summer's day,' which is shown below with markings of weak (˘) and strong (-) stresses.

˘ - ˘ - ˘ - ˘ - ˘ -

Shall I | com pare | thee to | a Sum| mer's day

Since the twentieth century, the foot has expanded from literary metrics to linguistic prosody. Jakobson argued that poetic structure is continuous with natural language rhythm [6]. Building on this, Hayes developed the Metrical Stress Theory that linked foot structure to perceived prominence and duration [7][8].

For Chinese, Zhu first introduced the Western concept of the foot, comparing it to the Chinese dun (顿) [9]. Wang further proposed a three-level structure of Chinese verse—line, foot, and rhyme [10]. Within generative phonology, Chen (1979) proposed a binary metrical foot structure and analyzed regulated verse in terms of a hierarchical metrical tree [11]. Duanmu argued that Mandarin employs moraic and syllabic trochees to account for stress distribution [12][13]. Feng extended the concept to spoken Chinese, proposing that the 'natural foot' is disyllabic, and a trisyllabic unit form a super foot [14]. Just like the Prosodic Hierarchy Theory, however, neither foot nor super foot are based on phonetic evidence.

1.2 Acoustic evidence for grouping

Ultimately, how continuous speech is divided into groups needs to be determined by empirical evidence, rather by intuition or anecdotal observations. There has been in fact much research on the phonetic cues that mark the boundaries of prosodic groups. One cue that has been consistently identified in multiple studies is final lengthening, namely, the lengthening of the duration of the final syllable or word of a group [15]. Another cue often accompanying final lengthening is a silent pause that occurs at a relatively large boundary [16][17]. These two cues can also be combined to jointly mark boundaries of various sizes [19][20]. Furthermore, group-internal shortening is also found to be a cue for prosodic grouping, namely, a syllable located in the middle of a prosodic group tends to be shortened [21]. Recent research has used these cues to explore various boundary-related issues. Chen [19] and Xia [22] conducted a corpus study that used syllable duration ratio—duration of current syllable divided by that of the preceding syllable, as a measure of boundary strength. Across words of many types, a substantial proportion showed no final lengthening. Instead, there was a strong tendency for those words to form larger prosodic groups with the following words that do show final lengthening. In [23] syllable duration ratio was used to assess the threshold boundary strength for Mandarin Tone 3 sandhi. Both reduced speech rates and increased sentence length raised boundary strength, which in turn lowered the likelihood of tone sandhi application. These findings raise the question as to whether speech utterances are indeed separated into neatly defined fixed groups like feet and prosodic words that are directly derived from syntactic and morphological structures [1][2].

1.3 Current study

The current study is designed to investigate a case where the prosodic grouping is supposed to be highly stable: classical Chinese poetry [11][14][24]. According to Feng [14], a canonical 5-syllable poem line conforms to a [2+3] structure, while a canonical 7-syllable line conforms to a [[2+2]+3] structure. Similar to Chen [11], Feng adopts the framework of modern metrical phonology and argues that the prosodic structure of Chinese verse adheres to the principle of the binary foot: each line obligatorily consists of two prosodic feet, and each foot typically consists of two syllables. In this study we test the alternative hypothesis that the grouping of a poetry line is not fixed, but variable with speech rate related to recitation style, such that a 5-syllable line will change from [2+3] to a holistic [5] as the speaking style changes from poetic to everyday speech, and a 7-syllable line will change from [[2+2]+3] to [4+3].

2. Methodology

2.1 Stimuli

The materials consisted of six couplets of classical Chinese poetry in traditional regulated verse patterns, including four couplets of 5-syllable lines and two couplets of 7-syllable lines, as shown in Table 1. These poem lines are all highly familiar to educated Chinese.

Table 1: *Stimuli: Couplets of popularly known classical Chinese poems.*

Original Lines & Interlinear Gloss	Translation
床 前 明 月 光， Bed front bright.moon light 疑 是 地 上 霜。 Seem COP ground.on frost	Before my bed, the moon shines bright, I wonder if it's frost on the ground.
深 林 人 不 知， Deep.forest person NEG know 明 月 来 相 照。 Bright.moon come MUT shine	In the deep woods, unknown to men, The bright moon comes to shine again.
返 景 入 深 林， Return light enter deep.forest 复 照 青 苔 上。 Again shine green.moss on	The setting sun rays pierce the deep wood, And shine upon the moss, so green and good.
岱 宗 夫 如何， Mount.Tai EXCL how 齐鲁 青 未 了。 Qi.Lu green NEG finish	O peak of peaks, how high it stands! One boundless green over spreads two states.
两 岸 猿 声 啼不住， Both.bank ape.sound cry NEG stop 轻 舟 已 过 万 重山。 light.boat PFV pass ten.thousand layer mountain	The monkeys' ceaseless cries along both shores die away, While my light boat has left ten thousand mountains far away.
朝 辞 白 帝 彩云 间， Morning leave Baidi colored.cloud between 千 里 江 陵 一 日 还。 Thousand.li Jiangling one.day return	Leaving at dawn the White Emperor crowned with cloud, I've sailed a thousand miles through canyons in a day.

2.2 Subjects and Recording Procedure

Forty native Mandarin speakers (20 females, 20 males, aged 18–25) participated as subjects. All were raised in mainland China, had no speech or hearing impairments, and were highly proficient in Standard Mandarin. Recordings were made in a quiet environment using personal devices. Each participant read the poems in three styles following the instructions given to them:

Style a: Poetic recitation, without exaggerated emotion but inter-phrase pauses allowed.

Style b: An intermediate style, clear but delivered continuously without pauses.

Style c: Natural, everyday speech, also without pauses.

All participants completed three repetitions of all couplets in all styles, yielding a total of 2,160 utterances for analysis (6 couplets × 3 styles × 3 repetitions × 40 participants).

2.3 Acoustic and Statistical Analysis

All utterances were manually annotated using ProsodyPro [25]. Each syllable was segmented and labeled in a TextGrid file, and Syllable duration values were automatically saved by ProsodyPro for statistical analysis. Boundary strength was estimated using Syllable Duration Ratio (SDR):

$$SDR = \frac{d_{cs}}{d_{ps}} \qquad (1)$$

where d_{cs} is the duration of the current syllable and d_{ps} is the duration of the preceding syllable. It should be noted that, due to the use of the classical poems, the intrinsic duration differences related to syllable structure were not controlled. This limitation was addressed by examining only changes of SDR across reading styles, without direct comparison of SDRs.

To validate that stylistic variation directly affected speech rate, average syllable durations were compared across styles (Table 2). As can be seen, there were progressive decreases in syllable duration from Style a to Style c, indicating increases in speech rate.

Table 2: *Syllable duration across styles and line lengths.*

Style	Line length	Sample size	Average (s)	sd_sec
a	5-char.	480	0.423	0.074
b	5-char.	480	0.317	0.057
c	5-char.	480	0.239	0.037
a	7-char.	240	0.385	0.065
b	7-char.	240	0.300	0.054
c	7-char.	240	0.230	0.035

2.4 Statistical Modeling

Linear Mixed-Effects Models (LMMs) were fitted separately for the pentasyllabic and heptasyllabic lines, with SDR as the dependent variable. Fixed effects included poem line, repetition, reading style, syllable position, and their interactions. Subject was included as a random intercept. The emmeans package in R was used for post-hoc pairwise comparisons between styles at each syllable position, supplemented by Cohen's *d* effect sizes to assess practical significance beyond *p* values.

3. Results

3.1 Grouping Patterns in 5-syllable Lines

As shown in Figure 2, in the 5-syllable lines, first, all three reading styles exhibit relatively high boundary strength at the 5th and 10th syllables of the couplets. These positions correspond to the positions of the comma and the period, respectively. Then, the 2nd and 7th syllables in Style a show the strongest boundary strengths, with average SDRs exceeding 1.8. Style b also shows

SDRs above 1.2 at the same positions. Finally, Style c shows SDRs at or below 1.0, apart from modest peaks at the positions of the comma and the period.

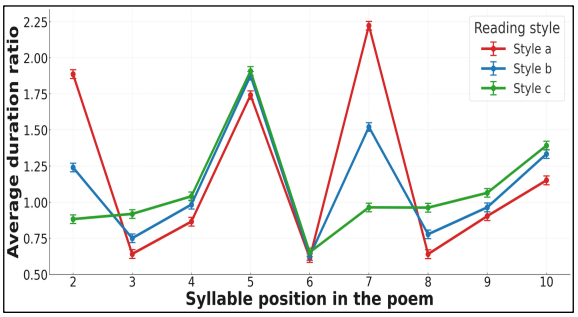

Figure 2: *Average SDRs in five-syllable lines in couplets, separated by styles, with the error bars showing 95% CIs.*

In terms of statistical significance, due to the large sample size, all comparisons reached significance ($p < 0.05$), which makes the results hard to interpret. But Cohen's *d* allowed us to focus on the comparisons with the greatest effect sizes. Figure 3 shows Cohen's *d* values for pairwise style comparisons (styles a vs. b, a vs. c, and b vs. c) at each syllable position in the 5-syllable lines. As can be seen, most syllable positions exhibit negative estimates and corresponding effect sizes, with only a few positions showing positive values. This is because effect size is calculated based on differences in SDR, and Style a often involves strong lengthening at specific syllables (resulting in a positive numerator when compared to Style c), but no or reduced lengthening at other syllables. Consequently, the negative values reflect a general weakening of boundary strength from Style a to Style c, indicating that prosodic boundary cues diminish as speech rate increases. Specifically, the 2nd and 7th syllables show highly significant stylistic differences, accompanied by extremely large positive effect sizes.

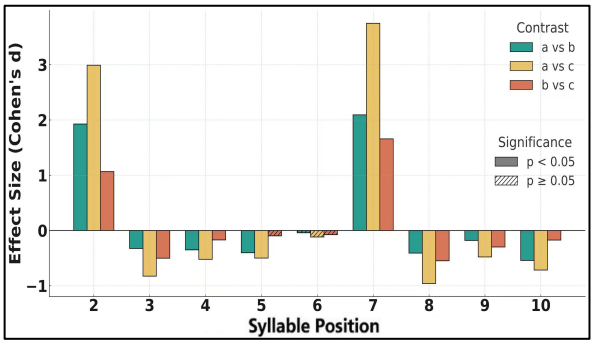

Figure 3: *p and d values of 5-syllable lines in couplets. The shading depth of the bars indicates the significance level of the corresponding p values.*

3.2 Grouping Patterns in 7-syllable Lines

Figure 4 shows average SDRs of the 7-syllable lines. Similar to the 5-syllable lines, all three reading styles exhibit relatively high boundary strengths at the line-final positions, i.e., the 7th and 14th syllables in each couplet. In addition, under Style a, the 4th and 11th syllables show the highest boundary strengths. Specifically, the average SDRs at these positions are significantly higher than those at other positions under the same style. Style b displays a similar pattern, with elevated boundary strengths at the 4th and 11th characters. In contrast, Style c, much like in the five-syllable lines, shows a more even distribution of boundary strengths across the line, apart from the 7th and 14th syllables.

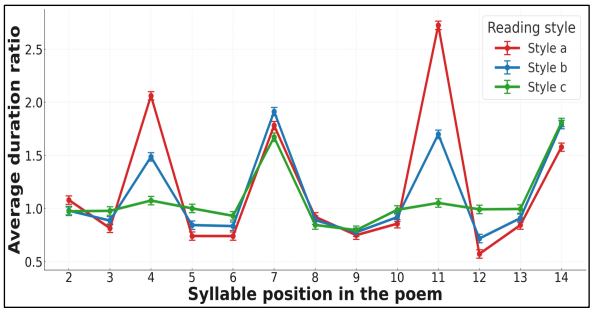

Figure 4: *Average SDRs in seven-syllable lines in couplets, separated by styles, with the error bars showing 95% CIs.*

Figure 5 shows *p* and Cohen's *d* values for pairwise style comparisons (a vs. b, a vs. c, and b vs. c) across syllable positions in 7-syllable lines. Positions 4 and 11 display highly significant stylistic differences, with all comparisons reaching $p < 0.001$, accompanied by extremely large positive effect sizes. At positions 4 and 11, therefore, style a shows the highest boundary strengths, far exceeding those of styles b and c.

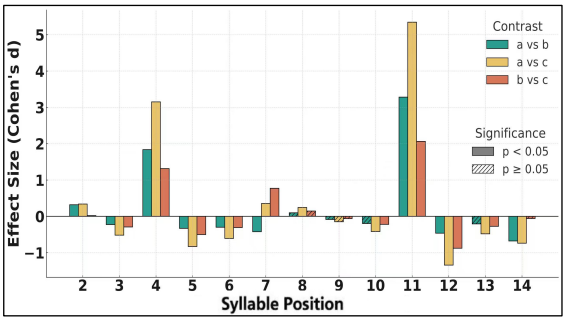

Figure 5: *p and d values of 7-syllable lines in couplets. The shading depth of the bars indicates the significance level of the corresponding p values.*

3.3 Line Length Effects on Grouping

Figure 6 compares the average SDRs at the second syllable between 5-syllable and 7-syllable lines. The average SDRs in styles a and b are higher in 5-syllable lines than in 7-syllable lines. This difference is most pronounced in style a. For instance, in 5-syllable lines, the average SDR at the second syllable under style a exceeds 1.8, whereas in 7-syllable lines it barely surpasses 1.0. These results show a clear effect of line length on boundary strength: the longer the line, the weaker the boundary.

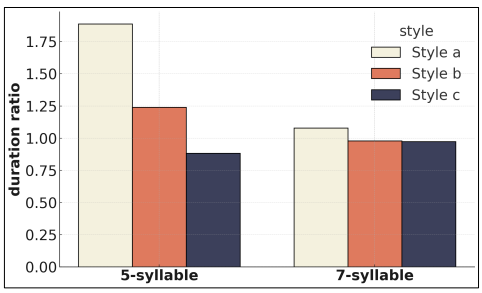

Figure 6: *Average SDRs at the second syllable in five- vs. seven-syllable lines.*

4. Discussion

4.1 Instability of the Disyllabic Foot

According to the binary-branching principle and the foot theory [11][26][27], a *couplet* of pentasyllabic classical Chinese poem is divided into two major groups, each consisting of five syllables. Each five-syllable line is further divided into a disyllabic foot and a trisyllabic superfoot, resulting in a [2+3] pattern. Consequently, the second syllable of each line (i.e., the 2nd and 7th syllables in the couplet) should correspond to a group boundary. Similarly, in a heptasyllabic verse, each couplet is first divided into two major groups of seven syllables. Each group is further divided into two disyllabic feet followed by a trisyllabic foot: [[2+2]+3], with the second syllable marking a minor boundary and the fourth syllable marking a major boundary[11].

The validity of the above theory-based predictions was examined with an established prosodic boundary marker, SDR—syllable duration ratio [19][22][23], and compared to that of an alternative prediction, namely, that the grouping of a poem line is not fixed, but variable as a function of speech rate related to recitation styles.

The results first demonstrated the validity of SDR as an indicator of boundary strength. In both pentasyllabic and heptasyllabic verses, high SDRs were found in all three reading styles at the syllable before a comma and at the final syllable of each couplet consisting of two 5- or 7 syllable lines (Figures 2 and 4). This confirms that SDR, which, in this study, combines both syllable duration and optional post-syllable silence, is an effective measurement of the boundary strength of a prosodic group.

The results also confirm the prediction that the 2nd syllable in each 5-syllable line and the 4th syllable in each 7-syllable line are both marked as strong boundaries with high SDRs in both of the poetic styles (a and b). This is consistent with both the theory-based and our own predictions. Even in style 2, where speech rate is reduced (Table 2), the prosodic grouping of pentasyllabic lines still displayed the [2+3] pattern. Similarly, in heptasyllabic lines, styles a and b both exhibit high duration ratios at the 4th and 11th syllable, indicating clear group boundaries.

However, two of our findings are inconsistent with the theoretical predictions, one of which was even beyond our own initial expectations. The first is that, for both pentasyllabic and heptasyllabic lines, the clear boundaries indicated by SDRs well above 1 at the 2nd and 4th syllables in the 5-syllable and 7-syllable lines, respectively, largely disappeared in style c, namely, the everyday speech style. This contradicts the static hierarchical predictions and suggests that prosodic grouping is not fixed but dynamic. It also challenges the concept of the disyllabic foot in Chinese. That is, a five-syllable group is not always spoken as a disyllabic group followed by a trisyllabic group, but can be spoken as a holistic 5-syllable phrase.

The second finding, which was the most surprising, is that the 7-syllable lines never exhibited a clear boundary at the 2nd syllable, as shown in Figure 4. This is the case even in style a, the most poetic style with the slowest speech rate (Table 2). Also in Table 2, the average syllable duration of 7-syllable lines in style a is much longer than that of 5-syllable lines in style c, in which the first disyllabic unit merged with the following trisyllabic unit to form a single group. This means that the theory-suggested [[2+2]+3] grouping for a heptasyllabic verse was never an option for our speakers. Instead, the initial disyllable unit is *always* combined with the following disyllabic unit to form a quadrisyllabic group.

4.2 Challenges to the Static Prosodic Hierarchy Hypothesis

The findings of this study therefore pose severe challenges to the standard Prosodic Hierarchy Theory [1][2], with data from the recitation of classical Chinese poems. The first challenge is that prosodic grouping is shown to be dynamic rather than static, and it is speech rate dependent. As seen from the pairwise comparisons across the three reading styles (Figures 3 and 5), most Cohen's *d* values were negative, indicating an overall decrease in boundary strength from Style a to Style c. As speech rate increases, many syllables originally located at boundary positions became group internal (Figures 2 and 4). That is, a small group is merged with the following group to form larger group. This is consistent with the observation of the corpus analysis that disyllabic, trisyllabic, and even quadrisyllabic words often merge with adjacent words to form larger prosodic groups [19] [22].

The second challenge is that the disyllabic unit is at best an optional grouping unit rather than a default basic unit [14], as it forms its own group only under specific conditions. Though not yet full clear, the conditions should include a) a sentence structure that allows it to be separated from a following unit with an even stronger grouping potential, e.g. a trisyllabic unit in a 5-syllabic poetry line, and b) a speech rate that is very slow, e.g., 3.2 syllables/s (Table 2), compared to the normal speech rate of 5-7 syllables/s [30][31]. Condition 1 is met only in the 5-syllable lines, and condition 2 is met only in the two poetic styles in the present study, where syllable 2 does exhibit final lengthening.

Finally, the current data may even challenge the foot, the prosodic word, and the prosodic phrase as distinctively definable levels of a prosodic hierarchy [1] [3]. At least for Mandarin, given that the disyllabic foot has no guaranteed boundary marking, the prosodic word as a unit consisting of feet would lose its status as a composite structure. When the 5-syllable line loses its internal boundary at a normal speech rate in the everyday speech style from a 2+3 pattern at slow speech rate, the resulting 5-syllable group becomes indistinct even between prosodic word and prosodic phrase. Thus, the phonological identity of these hypothetical units may become questionable.

5. Conclusion

We investigated prosodic grouping in Chinese by recording native speakers reciting five- and seven-syllable verses in poetic and everyday speech styles. We measured boundary strength using duration ratio of each syllable over the previous one (SDR). The results show that prosodic grouping changes dynamically with speech rate related to recitation style. Clear boundary markers in slow poetic styles (the 2nd syllable in 5-syllable lines and 4th syllables in 7-syllable lines) are weakened by increased speech rate, and largely eliminated in everyday speech style. This finding contradicts the widely accepted view that disyllables invariably form prosodic feet in Mandarin [14][28][29]. Finally, our findings even raise questions about the status of foot, prosodic word and prosodic phrase as distinctively definable prosodic units.

6. References

[1] E. O. Selkirk, "On prosodic structure and its relation to syntactic structure," *Indiana University Linguistics Club*, 1980.

[2] E. Selkirk, "*Phonology and Syntax: The Relation Between Sound and Structure*," Cambridge, MA, USA: MIT Press, 1984.
[3] M. Nespor and I. Vogel, "Prosodic Phonology," 2nd ed. Berlin, Germany: Mouton de Gruyter, 2007.
[4] Z. L., Wang and X. S. Wang, "The development and dilemmas in the study of rhythm in modern Chinese poetry," *Journal of Wuhan University: Humanities Edition*, 64(2), 94–98. (in Chinese), 2011.
[5] Y. Yufang and W. Bei, "Acoustic correlates of hierarchical prosodic boundary in Mandarin," in *Proc. Speech Prosody*, pp. 295–298, 2002.
[6] R. Jakobson, "Linguistics and poetics," in *Style in Language*, T. A. Sebeok, Ed. Cambridge, MA, USA: MIT Press, pp. 350–377, 1960.
[7] B. Hayes, "*A metrical theory of stress rules*" (Doctoral dissertation). Massachusetts Institute of Technology, 1981.
[8] B. Hayes, "*Metrical Stress Theory: Principles and Case Studies*," Chicago, IL, USA: Univ. of Chicago Press, 1995.
[9] G. Q. Zhu, "*On Poetry,"* National Book Publishing House (in Chinese), 1943.
[10] L. Wang, "*Chinese Poetics and Prosody,"* Shanghai: New Knowledge Press (in Chinese), 1958.
[11] M. Y. Chen, "Metrical structure: Evidence from Chinese poetry," *Linguistic Inquiry*, vol. 10, no. 3, pp. 371–420, 1979.
[12] S. Duanmu, "Rime length, stress, and association domains," *Journal of East Asian Linguistics*, vol. 2, no. 1, pp. 1–44, 1993.
[13] S. Duanmu, "Evidence for stress and metrical structure in Chinese," in *The Cambridge Handbook of Chinese Linguistics*, C.-R. Huang, Y.-H. Lin, I.-H. Chen, and Y.-Y. Hsu, Eds. Cambridge, U.K.: Cambridge Univ. Press, pp. 361–382. doi: 10.1017/9781108329019.020, 2022.
[14] S. L. Feng, "On the natural metrical foot in Chinese," *Chinese Language and Linguistics*, 1, 40–47. (in Chinese), 1998.
[15] D. K. Oller, "The effect of position in utterance on speech segment duration in English," *The journal of the Acoustical Society of America*, 54(5), 1235-1247, 1973.
[16] C. W. Wightman, S. Shattuck-Hufnagel, M. Ostendorf and P. J. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *The Journal of the Acoustical Society of America*, 91(3), 1707-1717, 1992.
[17] C. Petrone, , S. Fuchs and L. L. Koenig, "Relations among subgl ottal pressure, breathing, and acoustic parameters of sentence-lev el prominence in German," *The Journal of the Acoustical Society of America, 141*(3), 1715-1725. https://doi.org/10.1121/1.4976073, 2017.
[18] Y. Yang and B. Wang, "Acoustic correlates of hierarchical prosodic boundary in Mandarin," In *Proceedings of Speech Prosody 2002*, International Conference, 2002.
[19] S. Chen, "*Durational performance on polysyllabic words in speech grouping: A corpus study of Mandarin Chinese*," Master's thesis, Univ. College London, London, U.K., 2023.
[20] Y. Xu, "Timing and coordination in tone and intonation—An articulatory-functional perspective," *Lingua*, vol. 119, no. 6, pp. 906–927, doi: 10.1016/j.lingua.2007.09.015, 2009.
[21] Y. Xu and M. Wang, "Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin," *Journal of Phonetics*, vol. 37, no. 4, pp. 502–520, 2009.
[22] Y. Xia, "*An acoustic exploration of the prosodic grouping in Mandarin*," Master's thesis, Univ. College London, London, U.K., 2023.
[23] Yang, M. Wang, and Y. Xu, "Boundary threshold of Mandarin tone sandhi," *Chinese Journal of Phonetics*, 2025.
[24] S. L. Feng, "The prosodic mechanism of construction and evolution in Chinese poetry," *Studies in Chinese Poetry*, (1), 44–61. (in Chinese), 2011.
[25] Y. Xu, "ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis," In *Proceedings of Tools and Resources for the Analysis of Speech Prosody* (TRASP 2013), Aix-en-Provence, France. 7-10, 2013.
[26] S. L. Feng, "*Prosodic Syntax,"* Shanghai: Shanghai Education Press. (in Chinese), 2000.
[27] S. Duanmu, "Phonology of Chinese (Mandarin)," in *Encyclopedia of Language and Linguistics*. Amsterdam, The Netherlands: Elsevier, pp. 45–62, 2005.
[28] J. Cao, "Phonetic and linguistic clues for prosodic segmentation in Chinese rhythm," in *Proc. 5th Phonetic Conf. China*, pp. 184–187, 2001.
[29] S. Duanmu, "*The Phonology of Standard Chinese*," 2nd ed. Oxford, U.K.: Oxford Univ. Press, 2007.
[30] W. R. Tiffany, "The effects of syllable structure on diadochokinetic and reading rates," *Journal of Speech and Hearing Research,* vol. 23, pp. 894-908, 1980.
[31] A. Eriksson, "Aural/acoustic vs. automatic methods in forensic phonetic case work," in *Forensic Speaker Recognition*: Springer, 2012, pp. 41-69.