Temporal organization of prenuclear glides in Hefei Mandarin

Yifan Yang, Zhiheng Qian

Shanghai Jiao Tong University, China

yifanyang@sjtu.edu.cn, n1vnhil@sjtu.edu.cn

Abstract

This study investigates the temporal organization of prenuclear glides [j] and [w] in Hefei Mandarin using acoustic measurements. While the status of prenuclear glides in Chinese syllables remains debated, we employ a recently developed acoustic approach to analyze how CjV and CwV syllables are coordinated. Our results suggest that prenuclear glides in Hefei Mandarin are more likely to be part of the rime rather than forming complex onsets. The findings align with previous research on the status of Mandarin prenuclear [j] but reveal inconsistencies for prenuclear [w] compared to prior studies. This research contributes to the understanding of Chinese syllable structure and highlights the utility of acoustic methods in inferring abstract phonological structures.

Index Terms: prenuclear glide, syllable structure, temporal organization, Hefei Mandarin

1. Introduction

In Mandarin and other Chinese dialects, the canonical maximal syllable template is generally accepted as CGVX (where G represents a glide and X can be a glide or stop)[1], but the affiliation of prenuclear glides remains controversial. Various analyses have proposed different positions for the prenuclear glide: as part of the onset (e.g., [2]), part of the nucleus (e.g., [3]), an independent slot (e.g., [4, 5]), or as a secondary articulation of the preceding consonant [6]. These proposals have often been argued for based on rhyming patterns, language games, speech errors, phonotactics, etc. (see [1] for an overview).

However, recent studies have begun to examine the nature and organization of prenuclear glides in Mandarin [7, 8] and Cantonese [9] from the perspective of the coupled oscillator model of syllable structure in Articulatory Phonology [10, 11, 12, 13]. Drawing upon kinematic data, [7] and [8] found that jV in CjV syllables is more accurately analyzed as part of the rime, whereas Cw in CwV syllables can behave either as a consonant cluster or as a complex segment. Beyond Mandarin and other Chinese varieties, similar cases that have been studied in the framework of Articulatory Phonology include English onglides (and offglides) [14, 15] and Romanian diphthongs [16, 17, 18].

As noted by [7] and [8], the organization of prenuclear glides can vary across dialects. Building on this observation, the present study examines the temporal organization of prenuclear glides in Hefei Mandarin and investigates how [j] and [w] coordinate with a consonant in CjV and CwV syllables, or the organization between C and rising diphthongs, a term that is sometimes used to refer to GV sequences (e.g., [ja] and [wa]). Hefei Mandarin, a variety of Jianghuai Mandarin spoken in Hefei (the capital of Anhui Province), has 13 diphthongs, [ja, wa, jɛ, wɛ, we, ɥe, jɔ, jɯ, jã, wã, wæ̃, jĩ, ɥĩ] [19] (cf. [20]), all of which are rising diphthongs. Thus, this rich inventory makes Hefei an ideal case for exploring the status of prenuclear glides in syllable structure. Through this analysis, we aim to provide further insights into Chinese syllable structure from a novel perspective.

We adopt the theory of Articulatory Phonology to investigate how Cj and Cw are coordinated, which links to the organization of elements in syllable structure. Rather than relying on kinematic data, we employ a recently developed approach that uses measurements of acoustic production to infer temporal coordination [21]. A key advantage of this method is the relative ease of data collection, which has facilitated further research along these lines, including recent studies by [22] and [23].

Putting together, the goal of this paper is to investigate the temporal coordination between a consonant and a prenuclear glide in Hefei Mandarin, using acoustic measurements. We aim to contribute to the understanding of Chinese syllable structure and to demonstrate the application of new methods.

2. Production experiment and analysis

2.1. Participants

We recruited 10 native Hefei Mandarin speakers (5 male, 5 female) between the ages of 18 and 22. All participants were college students and were raised in Hefei. They speak Standard Mandarin at school but use Hefei Mandarin at home with their families. However, three speakers exhibited significant influence from Standard Mandarin and were therefore excluded from the analysis, leaving a final sample of seven participants (3 male, 4 female).

2.2. Materials

Hefei Mandarin has a rich inventory of rising diphthongs, but this study focuses on five of them, [jɔ, jɯ, wa, wɛ, we], excluding [ja], [jɛ], and [ɥe], as well as the nasalized series. The distribution of [ja] and [jɛ] is limited to alveopalatals ([tɕ, $tɕ^h$, ɕ]), and (near-)minimal pairs such as [tɕja]/*[tɕa] and [tɕjɛ]/*[tɕɛ] cannot be established for comparison in this study (see [20] for their treatment of rising diphthongs in Hefei). The diphthong [ɥe] is excluded due to its variation among the younger speakers, which aligns with the observation in previous research.

Our materials consist of 32 target syllables, categorized into three types: GV, CGV, and CV, as illustrated in Table 1. These syllables form various (near-)minimal GV/CGV pairs and CV/CGV pairs for comparison. For example, [jɔ] pairs with each of the six Cjɔ syllables, while there are four Cjɔ/Cɔ pairs: [$pjɔ^{53}$]/[$pɔ^{53}$], [$tjɔ^{53}$]/[$tɔ^{53}$], [$ljɔ^{53}$]/[$lɔ^{53}$], and [$mjɔ^{53}$]/[$mɔ^{53}$]. The tones are marked by superscript digits (where "1" indicates the lowest pitch, and "5" indicates the highest) and are carefully controlled, except for certain accidental gaps. These target syllables yield 15 GV/CGV pairs and 12 CV/CGV pairs. The number of comparable pairs is limited due to the restrictive distribution of phonotactics and tones in Hefei Mandarin.

Table 1: *Target GV, CGV, and CV syllables.*

GV (5)	**CGV** (15)	**CV** (12)
[jɔ^{53}]	[pjɔ^{53}], [tjɔ^{53}] [ljɔ^{53}], [mjɔ^{53}] [ɕjɔ^{53}], [tɕjɔ^{53}]	[pɔ^{53}], [tɔ^{53}] [lɔ^{53}], [mɔ^{53}]
[jɯ^{53}]	[ljɯ^{53}] [ɕjɯ^{53}], [tɕjɯ^{53}]	[lɯ^{53}]
[wa^{53}]	[kwa^{24}] [xwa^{53}]	[kʰa^{24}]
[wɛ^{53}]	[kwɛ^{53}] [ʂwɛ^{53}]	[pɛ^{53}], [kɛ^{53}] [ʂɛ^{53}]
[we^{53}]	[kwe^{53}] [ʂwe^{53}]	[ke^{24}] [ʂe^{53}], [se^{53}]

We embedded the target syllables in the carrier sentence [___ , tuəʔ^{5} ___ tsɿ^{21}] ("___ , read ___.") . Each sentence was repeated 10 times in a randomized order. We focus on *the second* token in each sentence in the following analysis.

2.3. Recording

The data were collected in a quiet room in Hefei in 2023. Recordings were made using a Zoom H6 recorder and Praat software [24] on a laptop computer, with a sampling rate of 44,100 Hz.

2.4. Annotation

The audio files were annotated by trained research assistants. Annotation was performed with reference to both the spectrogram and waveform, following a set of rubrics adopted in previous studies [21, 22].

First, we marked the left and right boundaries of each syllable. For GV syllables, the left boundary was identified as the appearance of a clear formant structure with strong energy. For CGV and CV syllables, when C was a plosive or affricate, the left boundary was marked at the ending point of the preceding syllable [tuəʔ^{5}] in the carrier sentence. When C was a fricative, nasal, or liquid, the left boundary was marked at the appearance of a clear formant structure and an associated change in the waveform. The right boundary for all syllables was identified as the moment when the clear formant structure disappeared and the energy dropped sharply in the waveform. Second, the boundary between C and G in CGV syllables was also annotated following the same procedure as in GV syllable annotation.

2.5. Analysis

2.5.1. Computing relevant intervals

The analysis is based on acoustic measurements following the methodology of [21]. The acoustic landmarks consist of the temporal midpoint of each segment and the endpoint of the vowel in each syllable, which is referred to as the *anchor*. Three key intervals were calculated using the time points extracted from the annotation: the left-edge-to-anchor, C-center-to-anchor, and right-edge-to-anchor intervals. These intervals

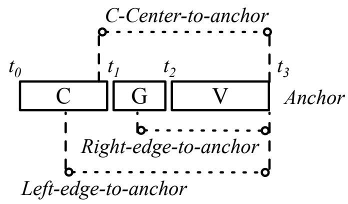


Figure 1: *Measurements of three intervals.*

are illustrated in Figure 1, where t_0, t_1, t_2, and t_3 correspond to the left edge of the syllable, the boundary between C and G, the boundary between G and V, and the right edge of the syllable, respectively.

In Figure 1, the C-center of a consonant cluster is determined by averaging the midpoints of the segments in the cluster (assuming CG is a cluster), while the left and right edges of the cluster correspond to the midpoints of the leftmost and rightmost segments in the cluster, respectively. It is important to note that in GV and CV syllables, the three intervals are identical.

Among the crucial time points, t_0, t_1, and t_3 were extracted from the annotated TextGrids, whereas t_2 (the boundary between G and V) was determined mathematically. We adopted the approach proposed by [25], which calculates the *Dynamic Quality Change* (DQC) of a diphthong to analyze its internal structure. In brief, for a diphthong, the change in vowel quality between each pair of consecutive time points q_i and q_{i-1} was calculated using the equation in (1) after [25, pp. 48]; namely, the Euclidean distance between q_i and q_{i-1}.

$$DQC = ||\mathbf{q}_i - \mathbf{q}_{i-1}|| \times 10,000 \quad (1)$$

In (1), q_i is the vowel quality at time point i, defined by three coordinates (x_i, y_i, z_i). Following [26] and [25], we define $SF1$, $SF2$, and $SF3$ as the lowest three formant frequencies, with SR serving as the base frequency used as a reference. Thus, x, y, and z are defined as the logarithmic distance from $SF1$ to $SF2$, from SR to $SF1$, and from $SF2$ to $SF3$, respectively [26, pp. 2121]. We adopted 185 Hz as the SR of female speakers and 155 Hz as the SR of male speakers [26, pp. 2121]. Based on the calculated DQC trajectory, the peak of the trajectory is identified as the boundary between the two elements within the diphthong, denoted as t_2 [25, pp. 48] (cf. [27] for an alternative method to determine the boundary between G and V). See Figure 2 for an illustration, where the dashed line indicates the locus of the peak. In some cases, the peak of the DQC trajectory is located at the final portion of the vowel due to instability or miscalculations of formant values near the end. We visually inspected these cases and selected the peak accordingly.

2.5.2. Comparison of target pairs and stability analysis

It has been argued that whether a prevocalic consonant sequence forms a complex onset can be diagnosed by the presence or absence of a C-centering pattern in the temporal organization of consonants (e.g., [10, 11, 28]). In this pattern, the mean of the midpoints of the consonantal gestures, which forms the C-center of the cluster, is temporally aligned with a following anchor, such as the end of the following vowel. Therefore, if CG also exhibits a C-centering effect, the coordination pattern would suggest that CG forms a complex onset and G is more consonantal by nature.

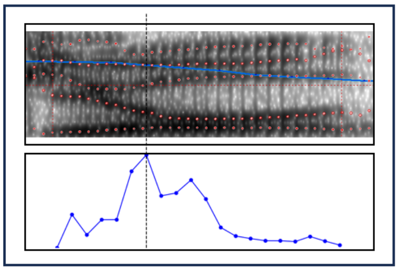


Figure 2: *Illustration of the DQC trajectory ([tjɔ⁵³]).*

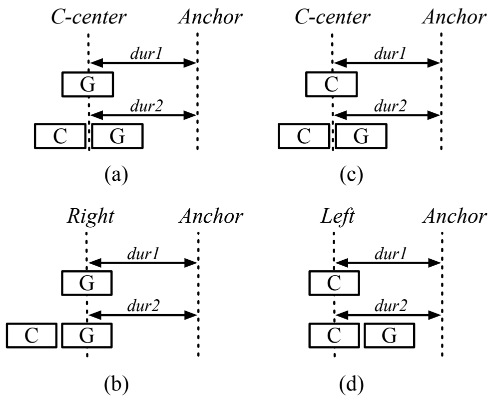


Figure 3: *Comparison of target pairs.*

This diagnosis is typically based on stability analysis. To evaluate the temporal organization, we conducted two comparisons for each (near-)minimal GV/CGV pair (Figure 3a, b) and CV/CGV pair (Figure 3c, d) based on the calculated intervals, after similar practice of [21] and [22].

For GV/CGV pairs, we compared the C-center-to-anchor interval with the right-edge-to-anchor interval of the target syllables in each pair. If the C-center-to-anchor is more stable, i.e., *dur1* and *dur2* exhibit less variability in Figure 3a, this suggests a rightward shift of G. Conversely, greater stability in the right-edge-to-anchor interval indicates a leftward shift of C (Figure 3b).

For CV/CGV pairs, we compared the C-center-to-anchor interval with the left-edge-to-anchor interval of the target syllables in each pair. If the C-center-to-anchor interval is more stable, this suggests a leftward shift of C (Figure 3c). However, if the left-edge-to-anchor interval is the least variable across both types of syllables in a target pair, this suggests that the prenuclear glide is "pushed" even further rightward, making it less likely to be part of the onset (Figure 3d).

Finally, if the C-center-to-anchor intervals in both GV/CGV and CV/CGV pairs show greater stability, this would indicate a C-centering effect of CG, implying that CG patterns like a complex onset.

The Relativized Standard Deviation (RSD) of each interval was calculated to assess stability patterns, using the formula in (2), where σ and μ represent the standard deviation and mean, respectively. The utility of RSD in studying temporal coordination has been supported by various authors (e.g., [29, 30, 21]). For each GV/CGV pair, the RSD was calculated across all tokens for each of the seven speakers. Taking the pair [jɔ⁵³]/[pjɔ⁵³] as an example, we pooled the C-center-to-anchor interval of all the tokens for the same speaker and calculated the RSD. Similarly, we pooled the right-edge-to-anchor interval for the same speaker and calculated its corresponding RSD. The same procedure was applied to CV/CGV pairs, except that we computed the RSD for the C-center-to-anchor and left-edge-to-anchor intervals instead.

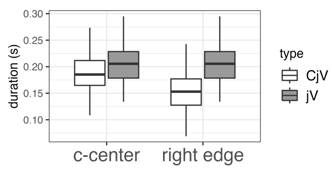
$$RSD = \frac{\sigma}{\mu} \times 100 \tag{2}$$

The results of the stability and duration analyses were further examined using mixed-effects regression modeling with the *lme4* package [31]. All statistical analyses and visualizations were conducted in R [32].

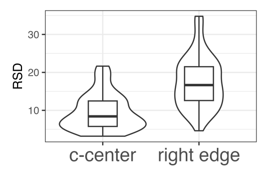
3. Results

3.1. Palatal glide [j]

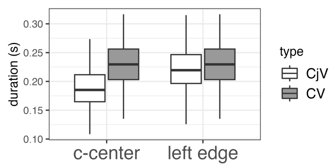
We first present the results for the jV/CjV and CV/CjV syllable pairs. Figure 4a illustrates the duration comparison between jV and CjV syllables, pooled across pairs and speakers. A visual inspection suggests that the C-center-to-anchor interval shows greater stability across these syllables, since the C-center-to-anchor durations for jV syllables closely align with those of CjV syllables. In contrast, the comparison between CV and CjV syllables reveals less variability in the left-edge-to-anchor interval, as shown in Figure 4c.



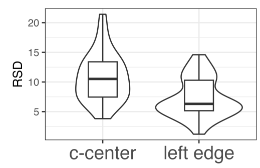
(a) *Duration: jV/CjV*



(b) *RSD: jV/CjV*



(c) *Duration: CV/CjV*



(d) *RSD: CV/CjV*

Figure 4: *Duration and RSD: jV/CjV, CV/CjV pairs.*

The findings are further supported by the RSD values presented in Figure 4b and 4d, as well as by mixed-effects modeling. The dependent variable in the models was the RSD values. The best model for both pairs included a fixed-effects structure of an intercept, interval type (C-center or right-edge/left-edge). The random effects included by-syllable pair and by-speaker random slopes and intercepts. The results indicate statistically clear smaller RSD values for the C-center-to-anchor interval in the jV/CjV pairs (baseline: C-center; $\hat{\beta} = 8.01, SE = 1.27, p = .00$; see Table 2), and smaller RSD values for the left-edge-to-anchor interval in the CV/CjV pairs (baseline: C-center; $\hat{\beta} = -3.52, SE = 1.30, p < .05$; see Table 3).

The results suggest that Cj does not exhibit a C-centering effect, and the prenuclear glide is shifted further to the right.

Table 2: *The fixed effects of jV/CjV RSD comparison.*

	est.	SE	df	t	Pr(> |t|)
(Intercept)	10.16	2.17	1.68	4.67	0.06
right-edge	8.01	1.27	7.19	6.30	0.00

Table 3: *The fixed effects of CV/CjV RSD comparison.*

	est.	SE	df	t	Pr(> |t|)
(Intercept)	10.76	1.37	4.37	7.86	0.00
left-edge	−3.52	1.30	4.490	−2.72	< .05

Table 4: *The fixed effects of wV/CwV RSD comparison.*

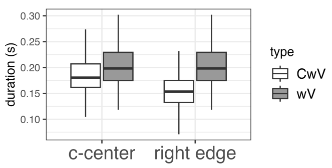
	est.	SE	df	t	Pr(> |t|)
(Intercept)	9.23	0.77	8.20	12.00	0.00
right-edge	6.65	1.08	8.33	6.18	0.00

Table 5: *The fixed effects of CV/CwV RSD comparison.*

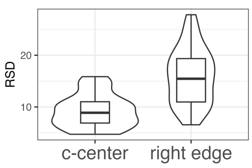
	est.	SE	df	t	Pr(> |t|)
(Intercept)	11.70	1.08	8.39	10.81	0.00
left-edge	−4.14	1.10	8.742	−3.78	0.00

3.2. Labiovelar glide [w]

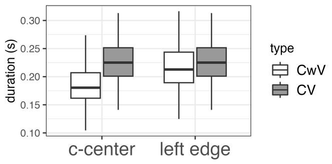
Similarly, Figure 5a illustrates the duration comparison between wV and CwV syllables, pooled across pairs and speakers, suggesting that the C-center-to-anchor interval shows greater stability across these syllables. Conversely, the comparison between CV and CwV syllables indicates less variability in the left-edge-to-anchor interval, as shown in Figure 5c.



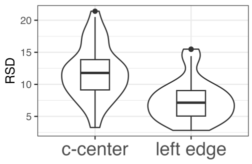
(a) *Duration: wV/CwV*



(b) *RSD: wV/CwV*



(c) *Duration: CV/CwV*



(d) *RSD: CV/CwV*

Figure 5: *Duration and RSD: wV/CwV, CV/CwV pairs.*

Following the same procedure, the results of mix-effects modeling indicate statistically-clear smaller RSD values for the C-center-to-anchor interval in the wV/CwV pairs (baseline: C-center; $\hat{\beta} = 6.65, SE = 1.08, p = .00$; see Table 4), and smaller RSD values for the left-edge-to-anchor interval in the CV/CwV pairs (baseline: C-center; $\hat{\beta} = -4.14, SE = 1.10, p = .00$; see Table 5).

Again, the results suggest that Cw does not exhibit a C-centering effect either, and the prenuclear glide [w] is also shifted further to the right.

4. Discussion and closing remarks

The results suggest a similar organizational pattern of Cj and Cw. The analyses of the GV/CGV pairs (where G = [j] or [w]) reveal a potential rightward shift of G, as indicated by the greater stability of the C-center-to-anchor interval (see Figure 3a, b; Figure 4a, b; Table 2, 4). Additionally, the CV/CGV pairs demonstrate a left-edge effect, with the left-edge-to-anchor duration exhibiting greater stability (see Figure 4c, d; Figure 5c, d; Table 2, 4). Taken together, these results suggest that the prenuclear G in both CjV and CwV is "pushed" further rightward, making it more likely to be part of the rime, rather than acting more like a complex onset that exhibits a C-centering effect.

This method slightly differs from what was used in [21] and [22], who focused only on the CV/SCV pair to determine whether word-initial sibilants (S) form part of a complex onset. In the study of English of [21], for example, the duration and RSD analyses indicate that the C-center-to-anchor interval is more stable, suggesting a complex-onset pattern of SC clusters in English. However, in the case of CGV syllables, previous research has found that G may behave either like a complex onset with its preceding C, or as a vocalic element in the rime [8]. Therefore, we need to examine both GV/CGV and CV/CGV pairs to determine whether both G and C exhibit shifts in CGV syllables.

The results are partially consistent with [7] and [8], who investigated the temporal organization of prenuclear [j] and [w] in Standard Mandarin and Taiwan Mandarin. Their kinematic data consistently show a "GV overlap" pattern of CjV syllables in both varieties of Mandarin, indicating that [j] is part of the rime [8, pp. 44]. This finding also aligns with previous findings on English glides, where [j] is argued to be more vocalic [14].

However, Cw shows variations in previous research. In [7] and [8], some speakers exhibited a C-centering pattern, while other speakers showed a right-edge effect of CV/CwV pairs, indicating a potential overlap of C and [w]. This indicates that Cw could either be a complex onset or secondary articulation C^w, differing from the findings of the current study on Hefei Mandarin. Additionally, [27] investigated the perceptual center in Mandarin syllables and concluded that the prenuclear glide (particularly [w]) in CGV syllables behaves more like a vowel in the rime. Thus, the conflicting results of Cw across various studies highlight the need for further investigation.

In conclusion, the present study, through acoustic analysis, finds that the prenuclear glides [j] and [w] in CGV syllables behave more like vowels in the rime rather than as part of the onset. The results for Cj align with previous research, reinforcing the understanding of its nature and role in Chinese syllables. Moreover, this study demonstrates the utility of acoustic methods in investigating syllable structure. The ease of data acquisition gives the method the potential to be applied to similar studies. In addition, [33] also pointed out that learners may acquire certain aspects of syllable structure from acoustic input, and therefore, it should be possible to infer abstract phonological representations through acoustic data. However, the pattern of Cw remains inconsistent across studies, suggesting the need for further investigation. Future research incorporating articulatory data and simulations could help determine whether these conflicting findings result from dialectal variation or other factors.

5. Acknowledgements

We would like to thank Huifang Kong for her assistance with data acquisition, and Yunjia Lyu, Qianyi Yang, and Tianjian Zhu for segmenting and annotating the acoustic data. All errors are our own.

6. References

[1] J. Zhang, "Chinese syllable structure," in *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, Jan. 2023.

[2] Z. Bao, "On the nature of tone," Doctoral Dissertation, Massachusetts Institute of Technology, 1990.

[3] H. S. Wang and C.-L. Chang, "On the status of the prenucleus glide in Mandarin Chinese," *Language and Linguistics*, vol. 2, no. 2, pp. 243–260, 2001.

[4] Y.-H. Lin, "Autosegmental treatment of segmental processes in Chinese phonology," Doctoral Dissertation, The University of Texas at Austin, 1989.

[5] J. van de Weijer and J. Zhang, "An X-bar approach to the syllable structure of Mandarin," *Lingua*, vol. 118, no. 9, pp. 1416–1428, Sep. 2008.

[6] S. Duanmu, *The phonology of standard Chinese*, 2nd ed., ser. Oxford linguistics. Oxford; New York: Oxford University Press, 2007.

[7] F.-F. Hsieh, G.-S. Li, and Y.-C. Chang, "Temporal organization of onglides in standard Chinese and Taiwanese Mandarin: A cross-dialectal study," *Journal of the Acoustical Society of America*, vol. 140, pp. 3107–3107, Oct. 2016.

[8] F.-F. Hsieh, "A critical review of recent approaches to onset weight —With special reference to the "medial" (onglides) in Chinese," *Studies in Prosodic Grammar*, vol. 2, no. 2, pp. 32–51, 2017.

[9] P.-R. Chen, F.-F. Hsieh, and Y.-C. Chang, "C-G vs. C-V timing differences in Hong Kong Cantonese," in *ISSP 2024 - 13th International Seminar on Speech Production*. ISCA, May 2024, pp. 35–38.

[10] C. P. Browman and L. Goldstein, "Some Notes on Syllable Structure in Articulatory Phonology," *Phonetica*, vol. 45, no. 2-4, pp. 140–155, Jul. 1988.

[11] C. P. Browman and L. Goldsteint, "Gestural specification using dynamically-defined articulatory structures," *Journal of Phonetics*, vol. 18, no. 3, pp. 299–320, 1990.

[12] C. P. Browman and L. Goldstein, "Competing constraints on intergestural coordination and self-organization of phonological structures," *Les Cahiers de l'ICP. Bulletin de la communication parlée*, no. 5, pp. 25–34, 2000.

[13] L. Goldstein, D. Byrd, and E. Saltzman, "The role of vocal tract gestural action units in understanding the evolution of phonology," *Action to language via the mirror neuron system*, pp. 215–249, 2006.

[14] B. W. Gick, "The articulatory basis of syllable structure: A study of English glides and liquids," Doctoral Dissertation, Yale University, 1999.

[15] F.-Y. Hsieh, "A gestural approach to the phonological representation of English diphthongs," Doctoral Dissertation, University of Southern California, 2017.

[16] S. N. Marin, "Vowel to vowel coordination, diphthongs and Articulatory Phonology," Doctoral Dissertation, Yale University, 2007.

[17] S. Marin and L. Goldstein, "A gestrual model of the temporal organization of vowel cluster in Romanian," in *Consonant Clusters and Structural Complexity*, P. Hoole, L. Bombien, C. Mooshammer, and B. Kühnert, Eds. De Gruyter, 2012, pp. 177–203.

[18] I. Chitoran and S. Marin, "Vowels and diphthongs - The articulatory and acoustic structure of Romanian nuclei," in *Contemporary Studies in Romance Phonetics and Phonology*, M. Gibson and J. Gil, Eds. Oxford University Press, 2018.

[19] J. Li, *Hefeihua yindang [The phonetic files of Hefei dialect]*. Shanghai: Shanghai Educational Press, 1997.

[20] H. Kong, S. Wu, and M. Li, "Hefei Mandarin," *Journal of the International Phonetic Association*, vol. 53, no. 3, pp. 1145–1166, Dec. 2023.

[21] K. Durvasula, M. Q. Ruthan, S. Heidenreich, and Y.-H. Lin, "Probing syllable structure through acoustic measurements: case studies on American English and Jazani Arabic," *Phonology*, vol. 38, no. 2, pp. 173–202, May 2021.

[22] R. Walker and Y. Yang, "Temporal coordination and markedness in Moenat Ladin consonant clusters," in *Proceedings of the Annual Meetings on Phonology*, vol. 10, 2023.

[23] Y. Yang and R. Walker, "Temporal organization in sibilant-stop clusters in Moenat Ladin," *Isogloss. Open Journal of Romance Linguistics*, vol. 11, no. 3, pp. 1–22, 2025.

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2022. [Online]. Available: http://www.praat.org/

[25] F. Ling, "Research on calculating the variation of chinese monophonic and compound sounds," *Fangyan*, vol. 37, no. 1, pp. 44–54, 2015.

[26] J. D. Miller, "Auditory-perceptual interpretation of the vowel," *The journal of the Acoustical society of America*, vol. 85, no. 5, pp. 2114–2134, 1989.

[27] Y.-J. Lin and K. De Jong, "The perceptual center in Mandarin Chinese syllables," *Journal of Phonetics*, vol. 99, p. 101245, Jul. 2023.

[28] C. P. Browman and L. Goldstein, "Competing constraints on intergestural coordination and self-organization of phonological structures," *Les Cahiers de l'ICP. Bulletin de la communication parlée*, no. 5, pp. 25–34, 2000.

[29] J. Shaw, A. I. Gafos, P. Hoole, and C. Zeroual, "Syllabification in Moroccan Arabic: evidence from patterns of temporal stability in articulation*," *Phonology*, vol. 26, no. 01, p. 187, May 2009.

[30] M. Q. Ruthan, K. Durvasula, and Y.-H. Lin, "Temporal coordination and sonority of Jazani Arabic word-initial clusters," *Proceedings of the Annual Meetings on Phonology*, vol. 7, Jun. 2019.

[31] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, 2015.

[32] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2024. [Online]. Available: https://www.R-project.org/

[33] K. Durvasula, "A simple acoustic measure of onset complexity," in *Proceedings of the International Congress of Phonetic Sciences*, 2023, pp. 838–842.