Exploring the variations in disyllabic lexical tone sandhi in Xiangshan Chinese

Author Anonymous, Co-author Anonymous

Author Affiliation

author@university.edu, coauthor@company.com

Abstract

Lexical tone sandhi in Northern Wu Chinese languages typically adopts a left-dominant mechanism, usually characterized as a tone extension pattern from the leftmost tone rightwards spanning the whole sandhi domain. By default, the underlying tone of the leftmost tone should be the only major factor conditioning the sandhi outputs in a lexical sandhi domain. However, variations still exist in some languages, the causes of which have not been studied extensively.

This study examines the categorical variations in disyllabic lexical tone sandhi in Xiangshan Chinese, an under-studied Northern Wu variety. Employing both machine clustering (*k*-means in this paper) and human perceptual categorization, this study discovers 4 distinctive sandhi clusters in 106 disyllabic lexical compounds produced by 8 native Xiangshan speakers. Further analyses of the cluster distributions suggest that the historical tonal categories of the initial tones, and individual speakers, serve as the major sources of the variations. The results provide evidence supporting the reality and relevance of historical tonal categories in tone sandhi, and suggest distinct development paths for sandhi tones and citation tones. Inter-speaker variations largely explain the remaining variation and may reflect a probabilistic quantification of phonological processes or different stages of language change.

Index Terms: tone sandhi, Northern Wu Chinese, cluster analysis, variations

1. Introduction

Tone sandhi, the phonological changes of tones in conditioned contexts, generally exhibits a *left-dominant* pattern in lexical words in Northern Wu Chinese [1]–[4]. It is typically realized as a rightward tone extension of the initial tone over the whole sandhi domain, optionally followed by a default L tone insertion, e.g., in Shanghai [5]–[7]. In some languages, the underlying tone of the initial syllable is replaced with another tone before its features spread rightwards, e.g., in Wuxi [8]. In addition, pure default tone insertion in non-initial positions can also be found, e.g., in Zhenhai [9]. All the cases show a dominance of the leftmost tone, whose underlying tone determines the sandhi contour of the entire domain.

By default, the underlying tone of the leftmost tone should be the only major factor conditioning the sandhi outputs in a lexical sandhi domain. However, distinct patterns persist in some Northern Wu varieties even when the leftmost tone remains constant. For example, in Zhenhai, an LM-initial disyllabic tone sequence in a lexical domain can either surface as L-MH or LM-L. It is claimed that the former has a final stress which attracts the underlying contour tone of the initial syllable; on the other hand, the latter has an initial stress which preserves the initial tone in its original position [9]. However, there is no consistent acoustic cue signalling this “stress” distinction, and it has been further suggested that the co-existence of such two different sandhi realizations might have correlations with historical tonal categories [10]. In ancient Chinese, there were 4 distinctive tonal categories, i.e., *ping* (level), *shang* (rising), *qu* (departing), *ru* (checked), each of which further split into *yin* (High) and *yang* (Low) registers in Middle Chinese. Despite disparate tone mergers and changes, the register difference is largely preserved in most Wu dialects. The relevance of historical categories in tone sandhi processes can be found in some Wu dialects, such as Danyang, whose tone sandhi outputs are predicted by the historical rather than synchronic tonal categories of the initial tones [11], [12].

Exceptional cases, where *non*-initial tones play a role in determining the lexical sandhi outputs, also exist in Wu dialects. For example, in Chongming, disyllabic words with the same initial underlying tones show different tone sandhi types depending on whether the second tone is oblique or even [13]. Furthermore, recent experimental studies on tone sandhi in different phrasal structures in Shanghai and Wuxi confirmed that other factors, such as lexical frequency and semantic transparency, can also influence the tone sandhi application [14]. Given the complexities and inconsistencies of the lexical tone sandhi mechanisms across Northern Wu dialects, further investigation is required to understand the causes of various distinctive lexical sandhi patterns.

Xiangshan Wu Chinese, an under-studied Northern Wu dialect, has 6 distinctive citation tones, 4 of which are non-checked tones (i.e., HH, HL, LHL, LH) and the remaining 2 are checked tones (i.e., Hq, LHq) [15]. There are variations in the use of the HH and HL tones, as well as the LHL and LH tones (marked via dashed lines in Table 1) [16]. For example, the majority of syllables with historical *yinping* and *yinshang* tones is now pronounced with an HH tone, but sometimes an HL tone is produced instead for certain syllables under these categories by some speakers.

Table 1: *Synchronic and historical citation tone system in Xiangshan Wu Chinese.*

	ping	***shang***	***qu***	***ru***
yin	HH		HL	Hq
yang	LHL		LH	LHq

In [17]’s massive data collection across 48 sites in the Wu region, speech recordings include those from 5 different places in Xiangshan in 1970s and 1990s. According to their analysis, the disyllabic lexical words with *yinping*-*yinping* tone combination in 4 regions in Xiangshan showed 2 distinctive tone sandhi contours, which come from several repetitions of the same words produced by the same speakers. The causes of such variations, however, were not further analysed. In this study, we extend this research line by asking 2 questions: (1) What are the possible distinct variations, if any, in the lexical tone sandhi domain in Xiangshan Wu Chinese? (2) What are the possible factors contributing to such variations?

2. Methods

2.1. Data elicitation

The current analysis examined a subset of data from a larger corpus of Xiangshan Chinese collected by the first author in March and April 2020. In total, the corpus consists of 416 disyllables and 352 trisyllables with varying citation tone combinations and morphosyntactic structures, the selection of which was determined in a pilot study with a male 53-year-old Xiangshan native speaker in 2019 and 2020.

The present data were elicited from 8 native Xiangshan speakers (4 female, age range: 47-53, mean age: 50). All of them were born and raised in Xiangshan, and were reported to use Xiangshan as their major language over 75% of time in daily life. The speakers were asked to read a semi-randomized order of test tokens as naturally as possible. Each token was presented in a separate slide in a Microsoft PowerPoint presentation to avoid potential list effects. Recordings took place in a quiet room locally in Xiangshan, using an H4n Zoom recorder at a sampling rate of 44.1k Hz. To prevent possible code-switching or influence from other languages, all the instructions were given solely in Xiangshan by the first author, who is also a native Xiangshan speaker.

To look specifically at the variations in lexical tone sandhi, this paper selected 112 f_0 contours of 14 disyllabic lexical compounds produced by all 8 speakers. There are 9 distinctive syllables in the initial position of the disyllables, which come from 2 historical tonal categories (*yinping* and *yinshang*), and were all pronounced as an HH tone by the informant in the pilot study. The second syllables with the same synchronic tone almost all come from the same historical tonal category (i.e., HH from *yinping*, HL from *yinqu*, LHL from *yangping* except for 链 *'chain'* from *yangqu*, and LH from *yangqu*). The disyllables were designed to cover all the 4 possible non-checked citation tone combinations with an initial HH tone (i.e., HH-HH, HH-HL, HH-LHL, HH-LH), as shown in Table 2. We excluded 6 tokens due to unnatural production or background noise. In total, 106 f_0 contours were examined in this paper.

Table 2 *Selected disyllabic lexical compounds in this study. The first column indicates the historical tonal category of the initial tone. The first row indicates the synchronic citation tone of the second tone.*

	HH	HL	LHL	LH
yin-ping	书包 *backpack* 樱花 *cherry blossom* 青椒 *green pepper*	青菜 *a variety of Chinese cabbage*	书房 *study (N.)* 樱桃 *cherry*	青豆 *green soya bean*
yin-shang	手机 *mobile phone* 苦瓜 *bitter gourd*	海带 *kelp*	手链 *bracelet* 水牛 *buffalo* 草莓 *strawberry*	扁豆 *haricot bean*

2.2. Data analysis

The recordings were segmented manually in *Praat* [18]. For each elicited token, the boundaries of the rhymes—which include onset glides, vowels, and final nasals—of both syllables were marked. F_0 values at 10 equidistant points of each rhyme were obtained using *ProsodyPro* [19]. To normalize between-speaker variations, z-scores of logarithmic transformed f_0 were calculated based on the given speaker's mean f_0 value [20], which were then used for further analysis.

Independent and *k*-means clustering analysis and human perceptual inspection were adopted to classify f_0 sandhi contours into distinctive clusters. Machine clustering, based on interpretable similarity metrics, is able to provide relatively objective and reproducible results. In this paper, we performed *k*-means analysis for longitudinal data using the package *kml* [21] in *R* [22]. The f_0 contours were partitioned into the nearest clusters based on Euclidean distance (L2 norm) from the cluster centroids, which were refined through iterative optimization. The *k*-means clustering was run through 2 to 10 clusters with an initialization method of *kmeans-* followed by an alternation of *kmeans--* and *randomK*.

However, the outcome machine clusters may not always be perceptually relevant, as this method solely relies on the distance variance within the data. Therefore, these data were supplemented with human perceptual inspection, since it can offer a categorization that is perceptually more meaningful, even though it is subject to varying degrees of subjectivity and inconsistency. An independent perceptual analysis was conducted by the first author, who carefully classified the f_0 contours according to the perceptual similarities together with the visual inspection of f_0 tracks in *Praat*. Since such approach is bottom-up and involves no predefined assumptions or hypotheses, the observer should remain unbiased and not in favour of any specific outcome. The clustering results of both approaches will be reported and compared in this paper.

Based on the clustering results, we then examined several potential factors that could contribute to the variations of f_0 tone sandhi contours: historical tonal categories of the initial tone, synchronic citation tonal categories of the initial tone, synchronic citation tonal categories of the non-initial tone, speaker, and item. Other factors, such as morphosyntax, lexical frequency, and length, etc., were not considered in the current analysis, because the selection of materials was strictly controlled to include only disyllabic lexical compounds which are commonly occurring and frequently used. Due to the small sample size, only descriptive analyses were conducted at this stage, but we are planning to implement statistical analyses on the larger dataset in the future.

3. Results

3.1. Monosyllabic citation tones

The 9 distinctive monosyllables in the initial position of the selected disyllabic lexical compounds, although all pronounced with an HH tone by the informant in the pilot study, show inconsistent tone values among the 8 speakers (also discussed in Section 1). Three out of the 8 speakers (S4, S5, S8) pronounced some tokens with an HL tone, as shown in Table 3. The tokens differ among the speakers, indicating a random rather than item-based nature of the citation tonal variations. In total, there are 18 tokens with an initial HL synchronic citation tone in the current dataset.

Table 3 *Syllables pronounced with an HL tone*

Speakers	Tokens produced with HL tone
S4	扁 *flat*, 水 *water*, 青 *green*
S5	书 *book*, 苦 *bitter*, 青 *green*, 樱 *cherry*
S8	扁 *flat*, 手 *hand*, 青 *green*

3.2. Disyllabic lexical tone sandhi patterns

As discussed in Section 2.2, both human perceptual analysis and *k*-means analysis were adopted to explore the categorical variations in the selected disyllabic lexical compounds. The perceptual analysis revealed 4 distinctive sandhi f_0 contours in the data (Figure 1A). We refer to the 4 clusters as HMML, MMMH, MHHL, and HHHH, with the first and the last pairs of letters respectively describing the contours of the first and the second sandhi tones.

As for the *k*-means clustering results, the *kml* package provides non-parametric criteria for selecting the optimal number of clusters, which select the solution that maximizes the ratio of between-cluster variance to within-cluster variance [23]. The optimal clustering according to Ray Turi criteria [24] gives 3 clusters (Figure 1B, row 1), which are similar to the HMML, MMMH, and MHHL clusters in the perceptual analysis. The "missing" HHHH cluster was mostly (87%) partitioned into the cluster C, a rise-fall contour. The close distance between the HHHH and MHHL clusters, whose first 75% portion of the contours almost overlap with each other, could be the reason why they are treated as similar contours by the machine which relies heavily on the distance matrix.

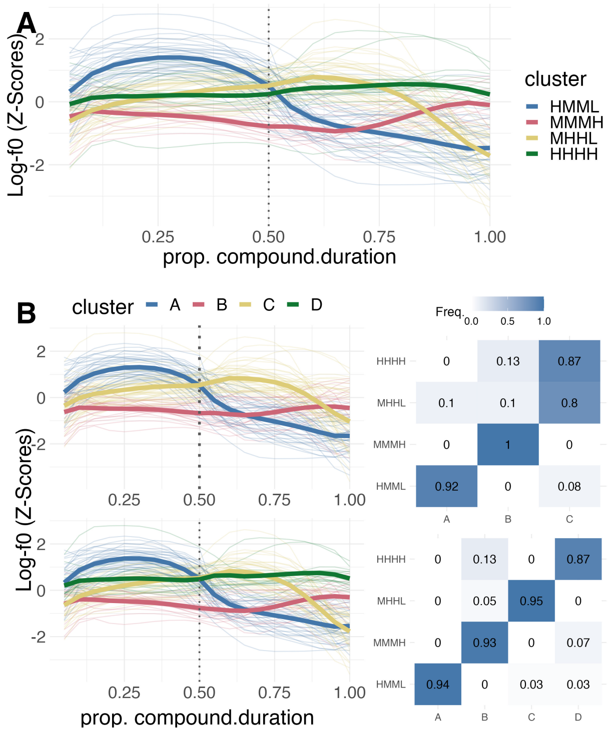

Figure 1 *Cluster results of the perceptual analysis (A) and k-means clustering (B). Solid lines are cluster means and faint lines are individual f_0 contours. Dotted vertical lines indicate syllable boundaries. Heatmaps at right show the distribution of the k-means clusters (x-axis) against the perceptual clusters (y-axis).*

We then tested the outcomes of k-means clustering by setting the number of clusters to 4. The outcome results are very similar to those from the perceptual analysis (Figure 1B, row 2). The distribution witnesses a neat one-to-one correspondence between the machine and human perceptual clusters, with over 87% overlap for each cluster pair (i.e., cluster A and HMML, cluster B and MMMH, cluster C and MHHL, cluster D and HHHH).

The *k*-means clustering and human perceptual analysis led to mutually compatible results when the number of cluster is pre-defined for the *k*-means clustering. The automatization of choosing the optimal number using *kml* package can be problematic in this study, as it completely leaves out one perceptually distinctive category. To this end, we will use the results of the 4 clusters for further analysis. The mismatched cases between the two clustering methods were further examined. It should be noted that certain degrees of ambiguity are unavoidable, as such purely form-driven categorization essentially forces gradient f_0 contours into distinct categories. In this case, we believe the machine clustering can deal better with those "marginal" data points in each category via rigorous calculation of distance. Therefore, we largely rely on the 4 cluster results given by *k*-means analysis and only changed the cluster label when the investigated f_0 contour is perceptually very different from the category to which it had been assigned.

3.3. Possible factors in tone sandhi variation

We first examined the distribution of all the 4 f_0 clusters across the synchronic citation tonal categories of the non-initial tones. By doing so, we aimed to explore whether those f_0 sandhi trajectories align with the anticipated left-dominant pattern in lexical compounds as attested in other Northern Wu dialects. That is, if such scenario is also relevant in Xiangshan, the citation tones of the non-initial tones should not play a role in varying the f_0 sandhi categories of the whole disyllables.

All the 4 f_0 sandhi clusters are found in each tonal category of the non-initial syllable (Figure 2A). Although the frequencies vary slightly across the 4 tonal categories, it is obvious that there is no one-to-one correspondence between the tonal categories and the f_0 clusters, suggesting that the citation tones of the non-initial tones cannot be the major source of such variations in sandhi outputs. The historical status of these tonal categories was not considered, as they are largely equal to the synchronic categories (see Section 2.1).

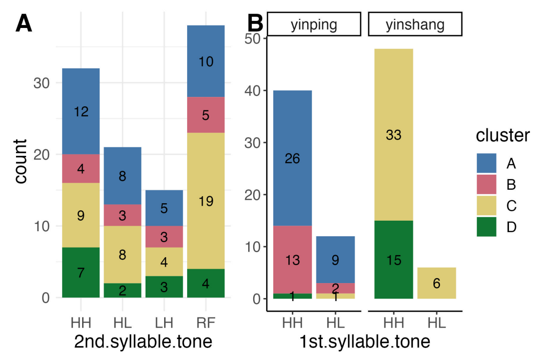

Figure 2 *Distribution of clusters across different citation tones of the second syllable (A) and the first syllable (B)*

To understand whether the various f_0 sandhi clusters can be explained by the initial tones, we then looked at the distribution of clusters across different historical and synchronic tonal categories of the initial tones. As shown in Figure 2B, an asymmetry of sandhi patterns is found between the *yinping* and *yinshang* tonal categories. While tokens with *yinping* initial tones mostly belong to clusters A and B (97.5% for HH and 91.7% for HL), those with *yinshang* initial tones have clusters C and D as their only sandhi clusters. The distinctiveness and exclusiveness in distributions imply a significant contribution of historical tonal categories to the variations of f_0 sandhi clusters.

Such trends do not vary too much across synchronic citation tonal categories of the initial tones, suggesting that the synchronic tonal categories do not serve as a major source of such sandhi variations. The only noticeable differences exist in the second most predominant sandhi clusters: cluster B for *yinping* only occurs twice for HL initial tone, and cluster D for *yinshang* is not available for HL initial tone at all. However, the relatively small number of HL tones in this dataset does not allow us to conclusively determine whether such difference in distributions of the 2 sandhi outputs is truly relevant.

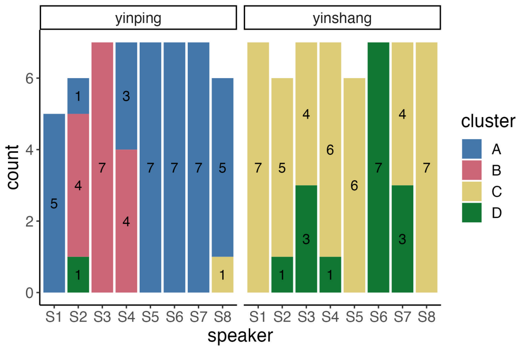

Figure 3 *Distribution of clusters across speakers*

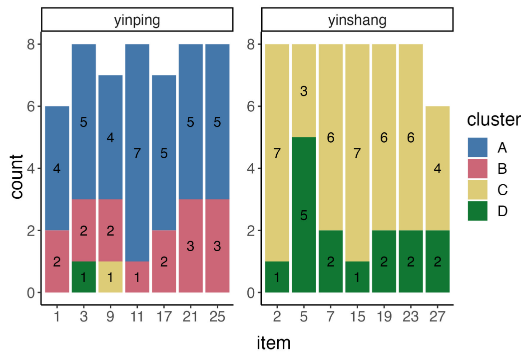

Figure 4 *Distribution of clusters across items*

Finally, the distributions of f_0 sandhi clusters across speakers, as well as items, were analysed. As shown in Figure 3, inter-speaker variations appear to be largely accountable for the presence of 2 distinct sandhi contours associated with each historical tonal category. Five out of the 8 speakers mostly produced cluster A contour for tokens with *yinping* initial tones, 4 of which (S1, S5, S6, S7) only produced this contour. The remaining 3 speakers produced cluster B contour as the most frequent pattern instead. On the contrary, for tokens with *yinshang* initial tones, 3 speakers (S1, S5, S8) only produced cluster C contour, and S6 only produced cluster D contour. The remaining 4 speakers produced a mixture of clusters C and D.

Items, on the other hand, show much less influence on the variant sandhi patterns. No item was produced as one single cluster contour consistently across the speakers (Figure 4). Within the *yinping* category, clusters A and B are found as the first and second most predominant patterns for each token, which conforms to the overall trend found across all the tokens. The same is true for tokens under the *yinshang* category, but with clusters C and D instead. Overall, this suggests that the selection of sandhi output is not item-based.

4. Discussion

Our research discovered 4 distinctive sandhi contours in disyllabic lexical compounds with an initial H-register synchronic citation tone in Xiangshan Chinese: (a) HMML, a general fall; (b) MMMH, a general rise; (c) MHHL, a rise-fall; and (d) HHHH, two level tones.

The distributions of the 4 sandhi clusters further suggest that the historical tonal categories of the initial tones, as well as individual speakers, contribute most to such categorical variations. In particular, disyllabic lexical compounds starting with a *yinping* historical tonal category mostly surface as an HMML or an MMMH contour, whereas those with an initial *yinshang* tonal category favour the other two sandhi outputs: MHHL and HHHH. Zooming into each historical tonal category, the choice between the two sandhi outputs is mostly speaker-specific. That is, some speakers prefer to produce an HMML for a *yinping*-initial disyllable, others choose the MMMH contour as the predominant pattern instead.

The results underline the importance of historical tonal categories of the initial tones in lexical sandhi applications in Xiangshan, and suggest distinctive development paths for the sandhi forms and the citation tones. While monosyllabic citation tones have undergone substantial tone merging, thus reducing the number of tones in the inventory, divergent sandhi patterns are still preserved in the disyllabic lexical sandhi domain. In this sense, the sandhi tones are *preservatory* in the diachronic tone change path and provide evidence confirming the reality of the historical tonal categories. With variant use of synchronic monosyllabic tones, Xiangshan provides unique evidence for the reality of historical tonal categories in a synchronically variable tonal system.

Various interpretations can be made of the inter-speaker variations. One possibility is that there are *variable rules* deriving the sandhi outputs from the same initial tone, each of which is associated with a probability, which then gives rise to variations and gradience [25]. This aligns well with our results, where many speakers show a co-existence of two sandhi outputs for the same initial historical tonal category, each associated with a different probability. Alternatively, the speakers can be viewed as being at different stages of language change. For example, in the 2 sandhi realisations of *yinping*-initial lexical disyllables, the HMML output can be derived in a left-to-right association from an initial HL synchronic tone, whereas the MMMH output is less transparent and might come from an earlier stage. If it is truly the case, speakers who only produced HMML may be viewed as in an earlier sandhi change state, while those who produced more MMMH contour may be in a newer stage. Minimally, the presence of different sandhi forms presents a challenge to Generative Phonology derivations of tone rules. We leave the actual derivation mechanisms for future research.

5. References

[1] M. S. Wright, *A metrical approach to tone sandhi in Chinese dialects*. University of Massachusetts Amherst, 1983.

[2] A. O. Yue-Hashimoto, 'Tone sandhi across Chinese dialects', *Wang Li Meml. Vol. Engl. Vol.*, pp. 445–474, 1987.

[3] M. Y. Chen, *Tone Sandhi: Patterns across Chinese Dialects*, 1st ed. Cambridge University Press, 2000. doi: 10.1017/CBO9780511486364.

[4] J. Zhang, 'A directional asymmetry in Chinese tone sandhi systems', *J. East Asian Linguist.*, vol. 16, pp. 259–302, 2007.

[5] B. Xu, Z. Tang, and N. Qian, 'Xinpai Shanghai fangyan de liandu biandiao [Tone sandhi in New Shanghai]', *Fangyan Dialects*, vol. 1981, pp. 145–155, 1981.

[6] R. J. You, 'Shanghaihua yindang [The sound archive of Shanghai]', *Xiandai Hanyu Fangyan Yinku*, 1994.

[7] B. Ling and J. Liang, 'The nature of left- and right-dominant sandhi in Shanghai Chinese—Evidence from the effects of speech rate and focus conditions', *Lingua*, vol. 218, pp. 38–53, Jan. 2019, doi: 10.1016/j.lingua.2018.02.004.

[8] M. K. Chan and H. Ren, 'Wuxi tone sandhi from last to first syllable dominance', *Acta Linguist. Hafniensia*, vol. 21, no. 2, pp. 35–64, 1988.

[9] P. Rose, 'Acoustics and phonology of complex tone sandhi: An analysis of disyllabic lexical tone sandhi in the Zhenhai variety of Wu Chinese', *Phonetica*, vol. 47, no. 1–2, pp. 1–35, 1990.

[10] Z. Li, 'Dual prominence in tone mapping: A case of tone movement', in *Proceedings of the 24th West Coast Conference on Formal Linguistics*, Cascadilla Proceedings Project Somerville, 2005, p. 245.

[11] M. K. Chan, 'Contour-tone spreading and tone sandhi in Danyang Chinese', *Phonology*, vol. 8, no. 2, pp. 237–259, 1991.

[12] M. K. Chan, 'An autosegmental analysis of Danyang tone sandhi: Some historical and theoretical issues', *Wuyu Yanjiu Stud. Wu Dialects*, pp. 145–184, 1995.

[13] M. Y. Chen and H. Zhang, 'Lexical and postlexical tone sandhi in Chongming', *Stud. Chin. Phonol.*, vol. 1, pp. 13–52, 1997.

[14] H. Yan, *The Nature of Variation in Tone Sandhi Patterns of Shanghai and Wuxi Wu*, vol. 4. in Frontiers in Chinese Linguistics, vol. 4. Singapore: Springer Singapore, 2018. doi: 10.1007/978-981-10-6181-3.

[15] Y. Shi, B. Post, and F. Nolan, 'The phonetics and phonology of corrective focus in Xiangshan Wu Chinese: interaction with disyllabic tone sandhi', in *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS)*, Prague, Czech Republic, 2023, pp. 1643–1647.

[16] Y. Shi, 'Tones and Disyllabic Tone Sandhi in Xiangshan Chinese Dialect', University of Cambridge, 2020.

[17] P. Rose, 'Dialect-geographical Acoustic-Tonetics: Five Disyllabic Tone Sandhi Patterns in Cognate Words from the Wu Dialects of ZhèJiāNg Province', in *Interspeech 2018*, ISCA, Sep. 2018, pp. 2733–2737. doi: 10.21437/Interspeech.2018-1130.

[18] P. Boersma and D. Weenink, 'Praat: doing phonetics by computer'. 2018.

[19] 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, 2013, pp. 7–10.

[20] X. Zhu, 'F0 normalization: how to deal with between-speaker tonal variations', *Lang. Sci.*, vol. 9, pp. 3–19, 2004.

[21] C. Genolini, B. Falissard, and P. Kiener, *kml: K-Means for Longitudinal Data*. 2023. [Online]. Available: https://CRAN.R-project.org/package=kml

[22] R Core Team, 'R: A language and environment for statistical computing'. Vienna, Austria, 2018. [Online]. Available: https://www.r-project.org/

[23] C. Genolini, X. Alacoque, M. Sentenac, and C. Arnaud, 'kml and kml3d: R Packages to Cluster Longitudinal Data', *J. Stat. Softw.*, vol. 65, no. 4, pp. 1–34, 2015, doi: 10.18637/jss.v065.i04.

[24] S. Ray and R. H. Turi, 'Determination of number of clusters in k-means clustering and application in colour image segmentation', in *Proceedings of the 4th international conference on advances in pattern recognition and digital techniques*, Calcutta, India:, 1999, p. 143.

[25] G. R. Guy, 'Linking usage and grammar: Generative phonology, exemplar theory, and variable rules', *Lingua*, vol. 142, pp. 57–65, 2014.