Prosodic word as the domain of breathiness: Evidence from disyllabic words in Kunshan Wu

Wenwei Xu[1], Peggy Mok[2]

[1]Leiden University, [2]The Chinese University of Hong Kong
w.xu@hum.leidenuniv.nl, peggymok@cuhk.edu.hk

ABSTRACT

This study provides new evidence for the claim that the domain of breathiness in Wu Chinese is a prosodic word instead of merely a syllable. Acoustic analysis of disyllabic words in Kunshan Wu shows that breathiness not only can pervade the first syllable but can also continue into the second syllable, provided that F0 is kept low. While previous research on the domain of breathiness in Wu Chinese has mostly focused on Shanghai Wu and only initial syllables, the results from Kunshan Wu, a more conservative dialect, suggest that the duration of breathiness within a prosodic word largely depends on F0 regardless of syllable boundaries. This accords with the fact that the domain of tone sandhi in Wu Chinese is usually a prosodic word, and that breathiness in Wu Chinese is a redundant cue for tone rather than a feature of syllable onset.

Keywords: phonation, voice quality, breathy voice, tone sandhi, Wu Chinese.

1. INTRODUCTION

Cross-linguistic research on the timing of non-modal phonation (e.g., breathy voice, creaky voice) has found both similarities and differences that are of theoretical importance to speech production and perception. In languages where non-modal phonation is typically a result of coarticulation from adjacent segments, its timing is generally restricted to a small portion of the vowel. Conversely, non-modal phonation that is phonologically specified in some languages can often last longer, or even pervade the vowel [1, 2, 3]. For example, breathiness introduced by /h/ or aspirated onsets in English shows only in the first half of the vowel [3], while the phonemic breathiness in Jalapa Mazatec [2, 4] and White Hmong [3, 5] can last over the entire vowel.

In the case of Wu Chinese, the timing of the phonologically specified breathy voice has been long debated. Early impressionistic accounts [6, 7] and experimental studies [8, 9] often attributed the breathiness to syllable onsets, while phonological proposals [10, 11] mostly favoured syllables as the domain of breathiness. Yip [12, 13] and Duanmu [14] extensively discussed this issue and proposed a register contrast for explaining the unique alternation of voicing and voice quality in (Shanghai) Wu, which has been widely accepted.

In Wu dialects, a syllable with a phonologically voiceless onset carries an upper register tone that is presumably modal and higher-pitched, while a syllable with a phonologically voiced onset carries a lower register tone that is usually breathier and lower-pitched. However, in a prosodic word, the phonation difference only surfaces in the initial syllable, while phonetic voicing (VOT < 0) of the voiced onsets only surfaces in non-initial syllables [8, 13, 14].

Such a phonological account has led to a common assumption, although never implying it, that the timing of breathiness should be somewhere within the initial syllable. The consequence is that recent phonetic studies, mostly on Shanghai Wu, all focused on monosyllabic words [15, 16, 17] or the first syllable of disyllabic words [15, 18]. The results showed that breathiness lasted for at least half of the vowel in monosyllables and for the entire first vowel in disyllables in Shanghai Wu [15, 16, 17], while in our previous study on Kunshan Wu (a closely related but more conservative dialect), breathiness did not vanish until the vowel offset in monosyllables carrying certain lexical tones [19].

Nevertheless, auditory impression on disyllabic or polysyllabic words in some Wu dialects also indicates that breathiness may still be perceptible in non-initial syllables, and it is also reasonable to speculate that breathiness can continue from the initial syllable into the following ones. The reasons are two-fold.

On the one hand, phonation is underspecified for non-initial syllables [12, 13], which implies that the phonation is not blocked or ‘reset’ at the boundary between the initial syllable and the following ones. On the other hand, within a prosodic word, a left-dominant tone sandhi is obligatory in Northern Wu dialects (including Shanghai and Kunshan Wu) [20], and the nature of such kind of tone sandhi is believed to be a process of rightward tonal spreading [21]. If pitch can spread across syllables, voice quality as a suprasegmental feature may also be able to spread.

Therefore, this study will leverage the acoustic data of disyllabic words in Kunshan Wu to show that breathiness can continue across syllable boundaries in certain conditions, and further argue that the domain of breathiness should not be the initial syllable only.

2. METHOD

2.1. Speakers

Sixteen native speakers of the downtown Kunshan Wu dialect were recruited, with age (middle-aged: 43–51; elderly: 67–73) and gender evenly distributed. All the speakers could also speak Mandarin, but none of them learnt it earlier than primary school.

2.2. Materials

The wordlist consisted of 128 disyllabic compounds and proper nouns, which were balanced for tone sandhi patterns (see Section 3.1) and voicing of the intervocalic consonants (voiceless unaspirated vs. voiced). While segments of the first syllable (S1) were not controlled, the second syllable (S2) was always composed of an obstruent onset (/p b t d k g f v s z tɕ dʑ/) and an oral vowel (/a ɛ ɪ ɔ/).

2.3. Procedure and analysis

Each speaker was recorded for three repetitions of the wordlist with the Behringer ECM 8000 Ultra-Linear Measurement Condenser Microphone at their home.

Various acoustic correlates to pitch and voice quality were measured using VoiceSauce [22]. F0 was estimated using the STRAIGHT algorithm [23]. Spectral tilts were then calculated as the differences between the amplitude of selected harmonics (H1-H2, H2-H4, H4-H2K, H2K-H5K, H1-H4, H1-H2K, H1-H5K, H1-A1, H1-A2, H1-A3), which were corrected for formant influence [24] (correction marked by * henceforth). Noise measures included the harmonics-to-noise ratios (HNRs) as well as the Cepstral Peak Prominence (CPP).

For a total of 6136 valid tokens, each acoustic correlates were measured on ten equidistant points in the vowels of both S1 and S2, and standardized into z-scores within speakers. Measures other than F0 were excluded for syllabic fricatives (e.g., /ɣ/) in S1.

To investigate the time course of the measures, general additive mixed models (GAMMs) were fitted with the R packages *mgcv* [25, 26] and *itsadug* [27], following [28, 29]. Significance testing was performed using the difference curve method with 99.5% confidence intervals to reduce Type I errors. For visualization of the time courses, fitted values in the GAMMs were plotted in normalized time with adjustment for mean duration of each tone.

3. RESULTS

3.1. Tone sandhi and F0

The disyllabic tone sandhi is, as mentioned earlier, largely left-dominant in Kunshan Wu, which means that the sandhi form of S1 depends solely on the lexical tone of S1, while the sandhi form of S2 depends on the lexical tones of both S1 and S2. The complex mapping between S1 and S2 will not be covered here, but as the focus of this study is on S2, the disyllabic sandhi patterns are grouped by S2 and summarized in Table 1. The F0 contours are plotted accordingly in Figure 1 (top).

For S1, the sandhi register (a = upper and modal, b = lower and breathy) corresponds to that of the underlying lexical tone. The sandhi contour is rising (TS1) or falling (TS2) on an unchecked rime, or short (TS3) on a checked rime. For S2, the sandhi tones do not depend on the underlying tone register, and the sandhi contours can be classified into four groups of similar pitch contours.

3.2. Voice quality

Principal component analysis (PCA) was first performed on a total of sixteen acoustic correlates, (averaged over S1). S1 tones were well separated as two registers along the first principal component (PC1, with 39.2% variance explained), and the useful acoustic correlates ($r > 0.25$ with PC1) were, apart from F0, H1*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, and all the HNRs. Two measures, H1*-H4* and HNR25 (measured over 0–2500Hz), which showed the largest correlations with PC1, were chosen for modelling phonation differences in GAMMs. The interpretation of HNR25 requires some caution because both breathier and creakier phonation can cause more noise (lower periodicity).

3.2.1. S1

The time courses of the acoustic correlates for S1 tones are plotted in Figure 1 (left). Ten words with a nasal onset were excluded for H1*-H4* in case the spectral tilt is biased by nasality.

TS1b (red dashed) showed a significantly larger H1*-H4* than the upper register tones except near the offset of the vowel (two time points). HNR25 was significantly smaller in TS1b than in TS1a except at the vowel onset (two time points), while there is no significant difference between TS1b and TS2a.

As for TS2b (green dashed), H1*-H4* was significantly larger from that of the upper register tones throughout the vowel. HNR25 was significantly smaller in TS2b than in TS1a except at the vowel onset (two time points), while there is again no significant difference between TS2b and TS2a.

Given the extremely short duration of the checked tones (TS3a & TS3b), H1*-H4* and HNR25 were averaged over the entire vowel and plotted as triangular points in Figure 1. Both measures showed a significant difference ($p < 0.001$).

Possible S1	S2	Contour
TS1a/b, TS3a	A	high-falling
TS2a/b, TS3a	B	mid-level
TS2a	C	low-falling
TS2b, TS3b	D	low-dipping

Table 1: Disyllabic sandhi patterns in Kunshan Wu.

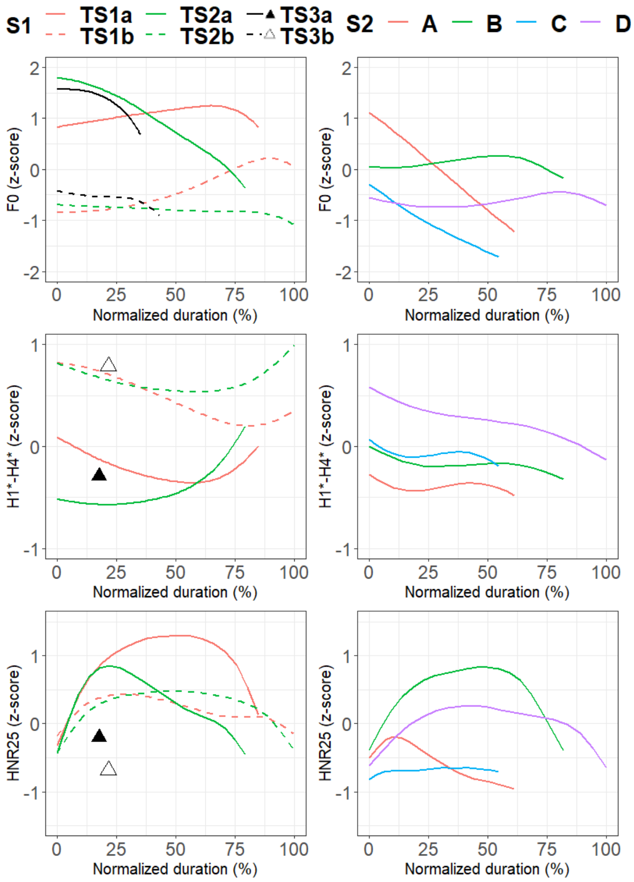

Figure 1: Time courses of F0, H1*-H4* and HNR25 (left = S1; right = S2).

3.2.2. S2

Once tone sandhi is applied, S2 loses its lexical tone identity, but the onset voicing still corresponds to the underlying register. However, no difference was found for H1*-H4* and HNR25 at all between S2 registers, which confirmed that S2 register could not initiate a phonation contrast as S1 register did.

Since the phonation of S2 did not depend on its register, the acoustic correlates were fitted by sandhi tones in S2, and the results are plotted in Figure 1 (right). Tone D (purple) showed a significantly larger H1*-H4* than the other tones for at least half of the vowel from the onset. HNR25 was significantly smaller in Tone D than in Tone B over the middle third of the vowel (discussion for Tone A and C in Section 4.1). Therefore, Tone D was breathier than the other tones.

3.2.3. S1 context for S2

As is shown in Table 1, the disyllabic sandhi is a many-to-many mapping between S1 and S2, which means the same S2 tone can be preceded by different S1 tones.

First, Tone A can be preceded by an upper register tone (TS1a or TS3a) or a lower register tone (TS1b). If breathiness in S1 can continue into S2, breathier phonation would be expected for Tone A when preceded by T1b. However, no acoustic difference was found. This is not surprising because the phonation difference between TS1a and TS1b already vanished before the offset of S1.

Similarly, Tone B can be preceded by an upper register tone (TS2a or TS3a) or a lower register tone (TS2b). Again, no acoustic difference was found between these two contexts, which contradicted our prediction because the phonation difference between TS2a and TS2b did not vanish before the offset of S1.

Finally, different from Tone A and B, Tone C can only be preceded by an upper register tone (TS2a), and Tone D is preceded exclusively by the lower register tones (TS2b or TS3b). Therefore, no further comparison was made.

4. DISCUSSION

4.1. Pitch-dependent voice quality

Voice quality is continuous and multidimensional in both production and perception, which means there is no precise boundary for deciding whether a tone is breathy or not. In this study, a breathy tone is defined in relative terms to a modal tone, for which TS1a is the best reference because as an upper register tone, it also showed the largest HNR25. In comparison with TS1a, the disyllabic sandhi tones are identified with distinct F0 and phonation patterns in Figure 1.

The first sandhi group is {TS1a/b + A} (red lines), which starts with an upper or a lower register tone and ends with a high-falling tone. Breathiness in TS1b is shown by a positive H1*-H4* and an HNR25 smaller than that of TS1a, which are both neutralized before the offset of S1. The phonation in S2 then becomes increasingly creakier as it approaches the pitch floor, which is clearly shown by a relatively low H1*-H4* and a remarkable fall in HNR25.

The second sandhi group is {TS2a/b + B} (green lines). It also starts with an upper or a lower register tone but ends with a mid-level tone. Breathiness in TS2b is shown by a positive H1*-H4* and a HNR25 smaller than that of TS1a, which are maintained at the offset of S1, but there is no phonation difference in S2 between the two S1 contexts. As Tone B generally shows a medial H1*-H4* (around z-score = 0) and relatively high HNR25 (near z-score = 1), it can be considered modal.

The third group only contains one pattern, {TS2a + C} (green solid + blue), which starts with an upper register tone falling from the pitch ceiling and ends with a low-falling tone towards the pitch floor. The onset of TS2a features a lower H1*-H4* than TS1a, while HNR25 starts high but falls abruptly later. This means TS2a is initiated in regular high pitch with larger glottal constriction, which is typical of tense voice. As pitch drops and becomes more irregular, it turns into (prototypical) creaky voice [30, 31]. In S2, Tone C exhibits a medial H1*-H4* and a negative HNR25, implying considerable noise without too much glottal constriction. This corresponds to the properties of 'unconstricted creaky voice', a form of phrase-final creak found in other languages [30, 31].

The last sandhi group also contains one pattern, {TS2b + D} (green dashed + purple), which starts with a lower register tone (slightly falling) and ends with a low-dipping tone. Breathiness in TS2b, as mentioned above, is maintained at the offset of S1. Compared with TS1a and Tone B, Tone D shows a positive H1*-H4* and a lower HNR, so there is still breathiness in S2.

In summary, voice quality is closely dependent on pitch in Kunshan Wu. Tones are generally realized with tense voice when pitch is extremely high, and with breathy or creaky voice when pitch is low. What is interesting is that the speakers tend to use different strategies to maintain or reach a low pitch. The phonologically relevant low tones are produced with breathy voice, while for the low pitch in falling tones, they switch to creaky voice, similar to where creaky voice is found in Mandarin Chinese [32].

4.2. The domain of breathiness

As discussed above, although the lower register tones are all initiated with breathiness, it vanishes before the offset of TS1b, so only TS2b and TS3b has the potential for carrying over breathiness to S2. Since Tone D is still breathy for at least half of the vowel, such carry-over effect seems to exist. Meanwhile, Tone B is not breathier when preceded by TS2b than TS2a, so there is no carry-over of breathiness from TS2b to Tone B. This raises the question on conditions that permit such kind of continuation.

Comparing the time courses of F0 and H1*-H4* between TS1a and TS1b, and between Tone B and Tone D, it is evident that breathiness tends to vanish as soon as F0 is raised to somewhere around the speakers' mean F0 (z-score = 0). The consequence is that breathiness seems to only continue from TS2b into Tone D as the F0 is kept low, but not into Tone B as the F0 is raised. This pattern is consistent with what was previously found in monosyllabic words in Kunshan Wu [19], where breathiness also vanished earlier in the low-rising tone than in the low-dipping tone. Such covariation is believed to be a result of conflicting perceptual demands for both pitch and voice quality [1, 2, 19].

One may also consider the possibility that a low pitch can automatically cause breathiness in Kunshan Wu, but it is clearly not true because the low pitch in Tone A and C, sequenced after various phonation types, is realized with creaky voice. Therefore, breathiness found in Tone D should only be explained as the continuation of the phonologically motivated breathy voice in the preceding syllable.

In other words, the continuity of breathiness in a disyllabic word is not blocked by syllable boundaries, but constrained by the continuity of a low pitch. This accords with perception studies that have found breathiness to be a redundant cue for low tones in Wu dialects [33, 34]. Given the evidence that breathiness not only pervades the first syllable but also continues into the second syllable in a low-pitched disyllabic word, the domain of breathiness in Kunshan Wu can no longer be just the initial syllable, but should be a larger domain where the alternation of voicing and breathiness is applied. A common notion of such a domain is a prosodic word, where the left-dominant tone sandhi is also obligatory in Kunshan Wu. The domain for variation in pitch and voice quality is thus unified, in which the carry-over of breathiness is not striking, but nicely aligned with the rightward tonal spreading trend in this left-dominant sandhi system. A new question is thus raised whether breathiness also continues across syllable boundaries in Southern Wu dialects, where tone sandhi is mostly right-dominant instead, and further research is called for.

5. CONCLUSION

Acoustic analysis of disyllabic words in Kunshan Wu showed that breathiness not only can pervade the first syllable, but can also continue into the second syllable in some conditions, while the duration of such continuation is believed to be contingent on a continuous low pitch. In addition, the spreading of voice quality across syllable boundaries echoes the spreading of tone across syllables in the left-dominant sandhi, which constitutes evidence for the proposal that the domain of breathiness should be, for example, a prosodic word that is larger than merely a syllable, and unified with the domain of tone sandhi.

Acknowledgements: This study was supported by the RGC General Research Fund 2019/20 Project Reference 14607619 awarded to the second author.

6. REFERENCES

[1] D. Silverman, “Phasing and recoverability,” Ph.D. dissertation, UCLA, CA, 1995.

[2] B. Blankenship, “The timing of nonmodal phonation in vowels,” *J. Phon.*, vol. 30, pp. 163–191, 2002.

[3] M. Garellek, “The timing and sequencing of coarticulated non-modal phonation in English and White Hmong,” *J. Phon.*, vol. 40, no. 1, pp. 152–161, 2012.

[4] M. Garellek and P. Keating, “The acoustic consequences of phonation and tone interactions in Jalapa Mazatec,” *J. Int. Phon. Assoc.*, vol. 41, no. 2, pp. 185–205, 2011.

[5] C. M. Esposito, “An acoustic and electroglottographic study of White Hmong tone and phonation,” *J. Phon.*, vol. 40, no. 3, pp. 466–476, 2012.

[6] Y. R. Chao, *Studies in the modern Wu-dialects*. Peking: Tsinghua College Research Institute, 1928.

[7] F. Liu, *Étude expérimentale sur les tons du chinois.* Paris: Soicété Les Belles Lettres, 1925.

[8] J. Cao and I. Maddieson, “An exploration of phonation types in Wu dialects of Chinese,” *J. Phon.*, vol. 20, pp. 77–92, 1992.

[9] N. Ren, “An acoustic study of Shanghai stops,” Ph.D. dissertation, Univ. of Connecticut, Storrs, CT, 1992.

[10] M. Sherard. “Shanghai phonology,” Ph.D. dissertation, Cornell Univ., Ithaca, NY, 1972.

[11] P. Rose, “Phonetics and phonology of Yang tone: phonation types in Zhenhai,” *Cahiers de Linguistique-Asie Orientale*, vol. 18, no. 2, pp. 229–245, 1989.

[12] M. Yip, “The tonal phonology of Chinese,” Ph.D. dissertation, MIT, Cambridge, MA, 1980.

[13] M. Yip, "Tonal register in East Asian languages," in *The Phonology of Tone: The Representation of Tonal Register*, H. van der Hulst and K. L. Snider, Eds., Berlin: Mouton de Gruyter, pp. 245–268, 1993.

[14] S. Duanmu, “A formal study of syllable, tone, stress and domain in Chinese languages,” Ph.D. dissertation, MIT, Cambridge, MA, 1990.

[15] J. Gao, “Interdependence between tones, segments and phonation types in Shanghai Chinese,” Ph.D. dissertation, Univ. Sorbonne Nouvelle–Paris 3, 2015.

[16] J. Gao and P. Hallé, “Phonetic and phonological properties of tones in Shanghai Chinese,” *Cahiers De Linguistique-Asie Orientale*, vol. 46, no. 1, pp. 1–31, 2017.

[17] J. Tian and J. Kuang, “The phonetic properties of the non-modal phonation in Shanghainese,” *J. Int. Phon. Assoc.*, pp. 1–27, 2019.

[18] Y. Chen, “How does phonology guide phonetics in segment– f0 interaction?” *J. Phon.*, vol. 39, no. 4, pp. 612–625, 2011.

[19] W. Xu and P. Mok, “The acoustic correlates and time span of the non-modal phonation in Kunshan Wu Chinese,” in *Proc. 12th ISCSLP*, 2021.

[20] E. Selkirk. and T. Shen, “Prosodic domains in Shanghai Chinese,” In *The phonology-syntax connection*, S. Inkelas and D. Zec, Eds., Chicago and London: Univ. Chicago Press, pp. 313–337, 1990.

[21] J. Zhang, “Tonal processes defined as tone sandhi,” in *The Cambridge Handbook of Chinese Linguistics*, C.-R. Huang, Y.-H. Lin, I.-H. Chen, and Y.-Y. Hsu, Eds. Cambridge: Cambridge Univ. Press, 2022, pp. 291–312.

[22] Y. Shue, P. Keating, C. Vicenik and K. Yu, “VoiceSauce: a program for voice analysis,” in *Proc. 17th ICPhS*, 2011, pp. 1846–1849.

[23] H. Kawahara, I. Masuda-Katsuse and A. de Cheveigne, “Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency based F0 extraction,” *Speech Commun.*, vol. 27, pp. 187–207, 1999.

[24] M. Iseli, Y. Shue and A. Alwan, "Age, Sex, and Vowel Dependencies of Acoustic Measures Related to the Voice Source," *J. Acoust. Soc. Am.*, vol. 121, no. 4, pp. 2283–2295, 2007.

[25] S. N. Wood, “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models,” *J. Royal Stat. Soc.: Series B*, vol. 73, no. 1, pp. 3–36, 2011.

[26] S. N. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed. Chapman and Hall/CRC, 2017.

[27] J. van Rij, M. Wieling, R. H. Baayen and H. van Rijn. *itsadug: Interpreting time series and autocorrelated data using GAMMs*. (2022). [Online]. Available: https://cran.r-project.org/web/packages/itsadug/itsadug.pdf

[28] M. Wieling, “Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English,” *J. Phon.*, vol. 70, pp. 86–116, 2018.

[29] M. Sóskuthy, “Evaluating generalised additive mixed modelling strategies for dynamic speech analysis,” *J. Phon.*, vol. 84, pp. 101017, 2021.

[30] P. Keating, M. Garellek and J. Kreiman, “Acoustic properties of different kinds of creaky voice,” in *Proc. 18th ICPhS*, 2015.

[31] M. Garellek, “The phonetics of voice,” in *The Routledge Handbook of Phonetics*, W. F. Katz and P. F. Assmann, Eds. Abingdon, UK, pp. 75–106, 2019.

[32] J. Kuang, “Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice,” *J. Acoust. Soc. Am.*, vol. 142, no. 3, pp. 1693, 2017.

[33] J. Gao, P. Hallé, and C. Draxler, “Breathy voice and low-register: A case of trading relation in Shanghai Chinese tone perception?” *Lang. and Speech*, vol. 63, no. 3, pp. 582–607, 2020

[34] B. Jiang, M. Clayards and M. Sonderegger, “Individual and dialect differences in perceiving multiple cues: A tonal register contrast in two Chinese Wu dialects”, *Lab. Phon.*, vol. 11, no. 1, pp. 11, 2020.