Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

An electroglottographic study of phonation types of tones in Suzhou Wu Chinese

Chunyu Ge[1], *Wenwei Xu*[1], *Wentao Gu*[2], *Peggy Mok*[1]

[1]The Chinese University of Hong Kong, [2]Nanjing Normal University

chunyuge@cuhk.edu.hk, wenweixu@cuhk.edu.hk, wtgu@njnu.edu.cn, peggymok@cuhk.edu.hk

Abstract

Tonal systems often involve cues other than F0, such as phonation types. This paper investigated the phonation distinction of tones in Suzhou Wu Chinese. Simultaneous electroglottographic (EGG) and audio data of isolated syllables collected from eight speakers aged above 65 were analyzed. Closed quotient (CQ) and peak increase contact (PIC) were measured on the EGG signals. Generalized Additive Mixed Models (GAMM) were conducted to analyze the time course of CQ and PIC, with tone and speaker's gender as smoothing terms. CQ and PIC were lower for low register than high register tones, with smaller differences in female than male speakers. The time courses of CQ and PIC were also varied with genders. The correlations between EGG and acoustic measurements were also calculated. H1*-H2* and H1*-A1* were more strongly correlated with CQ, whereas the correlations between them and PIC were weak. This paper showed that low register tones in Suzhou Wu were pronounced with breathy voice, which was more prominent at the onset of vowel, while the degree of breathiness and its time course differed between females and males. The EGG measurements and their correlations with acoustic measurements provide evidence to explicate the evolution of tones in Suzhou Wu.

Index Terms: tone, breathy voice, electroglottography, GAMM, Wu Chinese

1. Introduction

Tone as a phonological feature can involve a bundle of cues, including phonation types and vowel quality, in addition to pitch [1, 2]. The parallel between "tonal complex" and "register complex" which both involve multiple dimensions of cues has inspired a laryngeally-based account of tonogenesis and tone split [2, 3]. This account proposed that the pitch distinction as transphonologized from the previous voice contrast in initial consonants was mediated by a phonation stage, where voiced consonants induce breathy voice. Wu Chinese is a group of Chinese dialects which has a tonal system of two registers that evolved from the voice contrast of initial consonants. This paper aims to investigate the phonation distinctions of tones in Suzhou Wu Chinese, a representative dialect of Wu Chinese.

Wu Chinese has a tone system with two registers with strict restrictions on the co-occurence of onset consonant, tone, and phonation type. The syllables with originally voiced consonants are phonetically voiceless (i.e., with positive VOT) but with low tones and breathy voice in the following vowel, while those with voiceless consonants are modal voiced with high tones. Take Suzhou Wu as an example, the pitch contours of its tone system are shown in Figure 1. It shows the averaged pitch contours in Hertz, and the register of the tones by line type. T7 and T8 are checked tones with a glottal stop coda, and other tones are unchecked tones of open syllables. T1, T3, T5, and T7 are high register tones which co-occur with voiceless consonants and modal voiced, and T2, T6, and T8 are low register tones (historical T4 has merged with T6) that only co-occur with voiced consonants and breathy voiced. It is clear from the figure that low register tones are lower than high register tones in pitch.

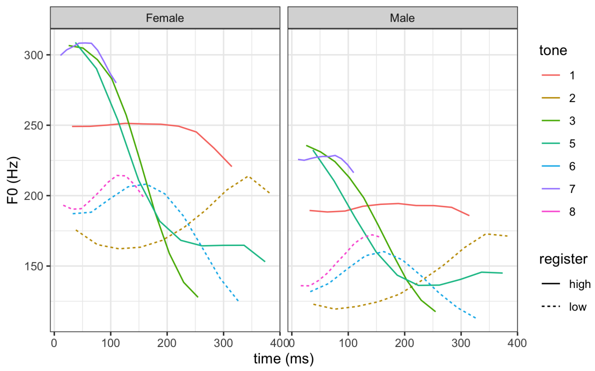

Figure 1: *Averaged pitch contours of Suzhou Wu tones*

Breathy voice as defined by [4] is articulated with the cartilaginous part of the vocal folds open, i.e., with post-glottal opening. Therefore, breathy voice is pronounced with a longer open phase within a glottal cycle (i.e., with higher open quotient (OQ), or shorter CQ, the inverse of OQ), and with more gradual abduction of the vocal folds. In acoustic terms, breathy voice has stronger amplitude in the first harmonic (H1) and steeper spectral slope, as well as stronger noise component [5] than modal voice. Some previous studies have investigated the breathy voice in Wu Chinese, including Suzhou and Shanghai Wu, a closely related dialect. In Suzhou Wu, low register tones were reported to have higher H1-H2 and H1-A1 [6] and stronger aspiration [7]. It is also found that the activity of cricothyroid (CT) was inhibited in syllables with voiced consonants, and they were pronounced with a lower laryngeal position [8]. Similar results were also reported for Shanghai Wu (e.g., [9, 10]). The low register tones in Shanghai Wu are with lower CQ with electroglottographic (EGG) data [9, 11]. Moreover, a gender difference was reported by [9] and [10] that the difference in CQ between high and low register tones was less for female than for male speakers. In addition, young speakers in Shanghai Wu are losing the breathy voice, as shown by [12] and [13]. To our knowledge, no EGG study has been done for Suzhou Wu. Suzhou Wu (seven tones) has a more complex tonal system than Shanghai Wu (five tones). This paper investigated the phonetic properties of phonation types of tones in Suzhou Wu Chinese with electroglottographic data from old speakers. Data from both genders were collected, in order to see whether these is also a gender difference in Suzhou Wu. Another aim of the paper is to explore the correlations of the EGG and acoustic measurements to see whether the breathy voice in Suzhou Wu is similar to Shanghai Wu and to investigate its implications on the phonetic explanation of tone split.

10.21437/SpeechProsody.2022-116

2. Method

Ten native speakers of Suzhou Wu participated in the experiment. The ages of the speakers were between 65 and 84 ($\mu = 71.3, \sigma = 6.15$) and there were five male and five female speakers.

A wordlist was compiled to cover all obstruent onsets. The test words were all CV (unchecked) or CVʔ (checked tones) syllables, where C = /p b t d k g f v s z ∅ (zero onset) ɦ/, and V = /ɑ ɛ ɪ/ for unchecked tones and /a ɪ ə/ for checked tones. The syllables covered all seven tones, and there were 103 words in total, notwithstanding accidental gaps.

Simultaneous EGG and audio recordings were made. EGG signals were recorded through the Glottal Enterprises EG2-PCX EGG equipment and audio signal through the Scarlett Solo Studio 2nd Gen audio interface. The EGG electrodes were attached to the speaker's neck, at the Adam's apple, where the thyroid cartilage locates. The sampling frequency was 44100 Hz for both audio and EGG recordings. The wordlist was randomized and printed on the paper. The speaker was instructed to read the word on the paper one by one. The words were read in isolation and self-paced. The wordlist was repeated three times by all speakers.

The sound files were manually segmented in Praat [14] according to audio waveforms and spectrograms. During the analysis, some tokens were found to be problematic and the signals of some speakers were too noisy. The data of two female speakers (F07 and F20) were discarded because the EGG signals were so noisy that the measurements were unreliable. The analysis of EGG signal was conducted in MATLAB. The signals were high-passed at 20 Hz and denoised. The peaks of dEGG signals and the point at which the EGG signal crosses 25% of the amplitude of the EGG peak were detected. The points were illustrated in Figure 2, where solid vertical lines represent the time at which dEGG peaks occur and dashed vertical lines represent that of the 25% threshold. Closed quotient (CQ) and Peak Increase Contact (PIC) were calculated with reference to these landmarks.

Closed quotient (CQ) is the ratio between the closed phase and the whole glottal cycle. Breathy voice is reported to have lower CQ than modal voice [15]. There are different methods to calculate CQ and this paper employed a hybrid method with dEGG peak as the closing peak and 25% threshold as the opening peak [16, 17]. The reason not to use negative dEGG peak as the opening peak is that this peak is usually less obvious than positive dEGG peak, as can be seen from Figure 2. The measurement of Peak Increase Contact (PIC) was straightforward as it is the amplitude of the positive peak of the dEGG signal [18]. It was called Derivative-EGG Closure Peak Amplitude (DECPA) by [19]. PIC is higher for breathy voice than modal voice [18, 20]. CQ and PIC were both measured at ten equidistant points during the vowel portion of the syllable.

CQ and PIC were both z-transformed within speakers. The reason is that the values of CQ and PIC are influenced by the physiology and articulation of individuals. For example, [9] suggested that there is a trading relation between OQ (the inverse of CQ) and F0 range. Simply aggregating the absolute values would jeopardize the relative differences between tones and registers and lead to misleading results. Unchecked and checked tones were analyzed separately because checked tones end with glottal stops and are thus shorter than unchecked tones. To model the time course of EGG measurements, Generalized Additive Mixed Models (GAMM) were fitted on CQ and PIC for unchecked and checked tones separately, using the `bam` function available in the `mgcv` [21] package of `R` [22]. GAMM is better than linear

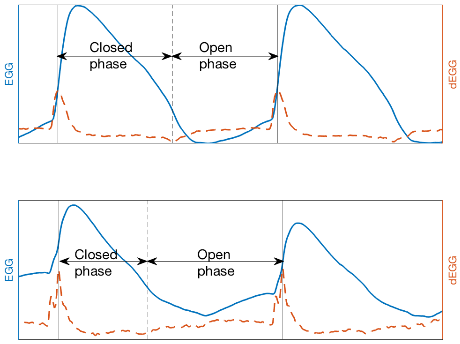

Figure 2: *Examples of EGG (solid blue) and dEGG (dashed orange) signals of /pɑ1/ (top) and /bɑ2/ (bottom)*

models in that it can capture the non-linear patterns in the data while relatively immune to outliers [23]. Another advantage of GAMM is that it can alleviate the influence of autocorrelated errors commonly found in time series data [23]. Tone and gender were entered as both intercepts and smoothing terms. The factor tone was transformed to an ordered factor, so that the coefficients capture the differences between tones. Word item was included as random intercepts. The random smooth of speaker was not included because gender was a covariate of speaker. As there were ten points in each vowel, the value of smoothing knots `k` was set to 9. Autocorrelated errors were handled by specifying the value of `rho` in the `bam` function as the error at lag 1 (i.e, the correlation between the values of the point and the immediately preceding point). Visualizations were conducted with the help of the `itsadug` package [24]. The correlations between EGG and acoustic measurements were also calculated. The acoustic measurements including H1*-H2*, H1*-A1*, H1*-A2*, and H1*-A3*. These measurements were calculated based on the audio signal using VoiceSauce [25]. The spectral measurements were all corrected for formants. As CQ and PIC, these measurements were also made at ten points during the vowel. Due to the EGG data loss of noisy signal, EGG and acoustic measurements were trimmed to include the same number of speakers (five male and three female speakers) and tokens. The Pearson's correlation r was calculated with values of all points.

3. Results

3.1. Closed quotient (CQ)

Figure 3 showed the fitted values of CQ of unchecked tones for male (top) and female (bottom) speakers. Tones were represented with different colours, while solid lines represented high register tones and dashed lines low register tones. The shaded ribbons represented the 95% confidence interval of the fitted values. Low register tones were lower in CQ than high register tones, especially at the first half of the vowel. The difference between the two registers were smaller for females than for males. Also, there were some overlap between T1 and T6 for female speakers. For the time course of CQ, it was relatively constant for high register tones for the most part of the vowel and was falling at the end for male but rising for female. Low register tones shared the same falling (for male) or rising (for female) towards the end of the vowel, but were slightly rising until 70% of the vowel. Therefore, there were effects of both tone and gender in CQ of unchecked tones.

In the same manner as Figure 3, Figure 4 plotted the fitted

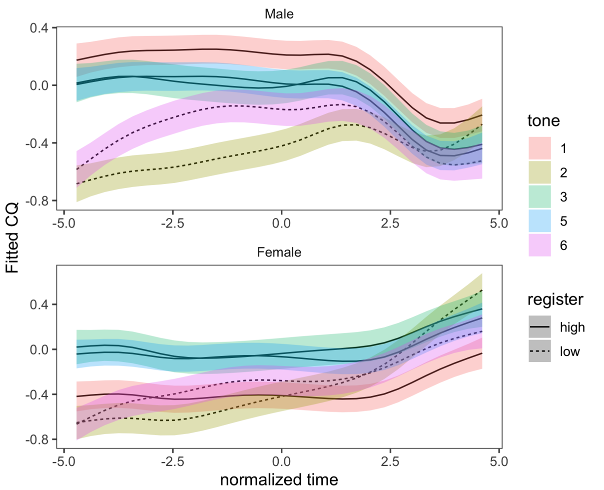

Figure 3: *CQ of unchecked tones for male (top) and female (bottom) speakers*

CQ of checked tones for male and female speakers. CQ was also lower for the low register checked tone, T8 than T7, but the difference was smaller than that of unchecked tones in that there were some overlap between T7 and T8 even at the beginning of the vowel. For male speakers, CQ was relatively constant for T7, but was gradually increasing for T8. Although the time course of CQ was similar for female speakers at the first half of the vowel, it was again rising at the end of the vowel. Compared to unchecked tones, the curves of CQ were more smooth for checked tones and it was almost linear for male speakers.

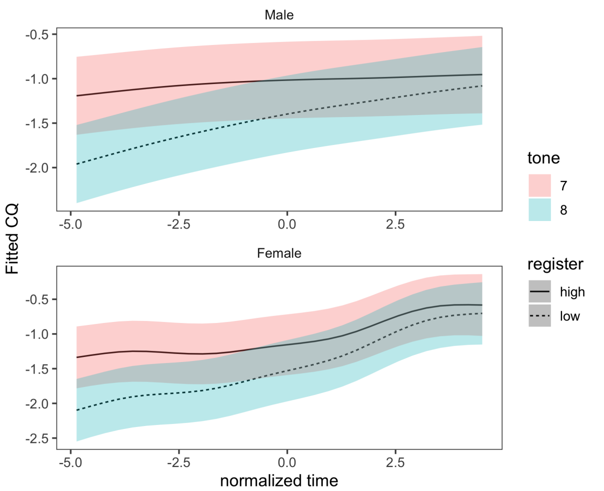

Figure 4: *CQ of checked tones for male (top) and female (bottom) speakers*

3.2. Peak Increase Contact (PIC)

PIC was also fitted with GAMM models. The fitted values of PIC were plotted for unchecked and checked tones separately as in Figures 5 and 6. PIC was lower for T2 and T6 at the beginning, but female speakers showed smaller differences, as also observed for CQ. For female speakers, PIC of T2 and T6 had a large overlap with high register tones. In general, unchecked tones shared a pattern of relatively stable PIC at the first half of the vowel and falling during the second half. The difference between high and low register tones was also gradually decreasing towards the end of the vowel, except for T2 of female speakers. PIC of T2 had a more complex curve than other tones in that it was gradually increasing until 70% of the vowel, and was then falling. PIC of T2 even exceeded T1 at the end of the vowel for females.

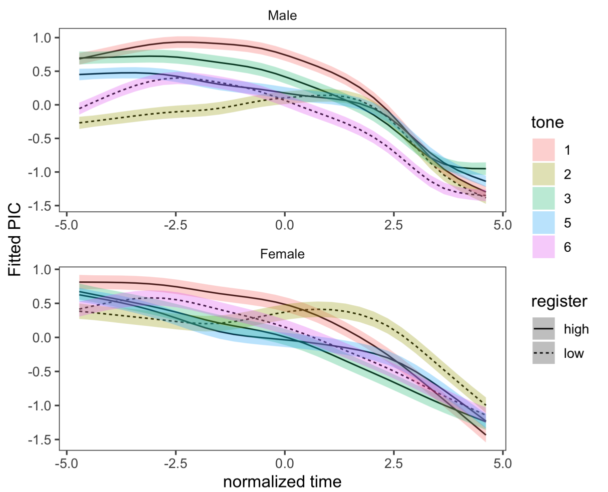

Figure 5: *PIC of unchecked tones for male (top) and female (bottom) speakers*

Similar patterns can also be observed for checked tones as in Figure 6. The low register checked tone T8 was lower in PIC than the high register tone T7 for both genders, and the difference was greater than unchecked tones. T8 resembled T2 in having a rising falling curve of PIC. T8 was lower than T7 over the whole part of the vowel for male speakers, but they overlapped during the late half for females.

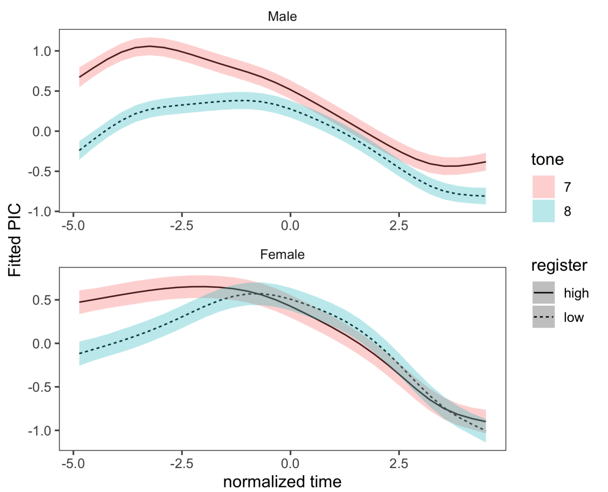

Figure 6: *PIC of checked tones for male (top) and female (bottom) speakers*

3.3. Correlations between EGG and acoustic measurements

Although the space limit prevents a full presentation of the acoustic data, correlations between EGG and some spectral measurements were also calculated and presented in Table 1. Correlations were calculated for male and female speakers separately, and bold texts indicated significant correlations ($p < 0.01$). Most correlations were significant except for that between PIC and H1*-A2* for males and H1*-H2* for females. CQ was negatively correlated with all spectral measurements. An unexpected observation is that the correlation between CQ and H1*-H2* was less strong than H1*-A1*, H1*-A2* and H1*-A3*. For male, the correlation between CQ and H1*-A3* was stronger than that between H1*-H2*, and the correlations between CQ and H1*-H2* was the smallest among the four spectral measures for females. Compared to CQ, the correlations between PIC and the spectral measures were weak. The correlations between PIC and H1*-A2* for male and H1*-H2* for female were almost zero, suggesting that there was not a relationship between PIC and the spectral measurements.

Table 1: *Correlations between CQ, PIC and H1*-H2*, H1*-A1*, H1-A2*, and H1*-A3**

	Male		Female	
	CQ	PIC	CQ	PIC
H1*-H2*	**-0.22**	**0.07**	**-0.15**	0.00
H1*-A1*	**-0.20**	**0.03**	**-0.25**	**-0.05**
H1*-A2*	**-0.17**	0.00	**-0.28**	**-0.06**
H1*-A3*	**-0.26**	**-0.04**	**-0.19**	**-0.17**

4. Discussion and Conclusions

The study investigated the phonation distinction of tones in Suzhou Wu using two EGG measurements, CQ and PIC. Both unchecked and checked tones showed low CQ and PIC for the low register tones. The lower CQ and PIC for the low register tones indicated that they were pronounced with breathy voice, which is consistent with previous acoustic results in Suzhou Wu [6]. Gender also has an influence on the phonation distinction. Female speakers were less breathy in low register tones than male speakers and thus the phonation distinction was reduced for female speakers. The effect of gender could have two sources. The first is that EGG measurements may be less accurate for female than for male [26]. The Adam's apple is less distinct and there is more adipose tissue in the neck for women. Another source is that there may be an ongoing disappearing of the breathy voice, and female speakers were more advanced than male speakers in this process. As reviewed in the Introduction, [13] demonstrated the ongoing loss of breathy voice in low tones of Shanghai Wu. The old male speakers were the only group which had higher OQ (i.e., lower CQ) for low tones, among old female and young speakers. She attributed the disappearing of breathy voice to the intense contact with Standard Chinese. The less breathy voice in the low register tones for female speakers found in Suzhou Wu therefore resembles that in Shanghai Wu. Suzhou Wu is in a comparable position with Shanghai Wu in having intense contact with Standard Chinese. It is reasonable to expect that Suzhou Wu is also in a process of ongoing loss of the breathy voice in the low register tones.

The correlations between the EGG and acoustic measurements offered valuable evidence on the phonetic properties of the breathy voice in Suzhou Wu. Although the correlations between PIC and the spectral measurements were weak and hard to interpret, which reflects the fact that the role of PIC was not yet clear as to which aspects of glottal activity it captures [10, 19], the correlations between CQ and the spectral measurements are more informative. The lower CQ indicates that the low register tones were pronounced with longer open quotient. However, the correlations between CQ and H1*-H2* were comparable or even lower than those between H1*-A1*, H1*-A2* and H1*-A3*, unlike that found in other languages with breathy voice, for example, White Hmong [20]. As H1*-H2* is related to glottal aperture, and H1* minus the amplitude of the harmonics near the first three formants is related to the gradualness of glottal opening [27], it is likely that the breathy voice in Suzhou Wu may be more related to the slackness of the vocal folds than to the larger opening of vocal folds. Previous studies on Wu Chinese supported this view. [8] reported lower activity of CT for voiced stops than voiceless stops in Suzhou Wu, which means less tension in vocal folds. Moreover, [10] reported that H1* minus the amplitude of the harmonics near the first three formants was more important than H1*-H2* in contributing to the breathy voice in Shanghai Wu. The evidence suggests that breathy voice in Wu Chinese is different from the typical breathy voice pronounced with post-glottal opening of the glottis as defined by [4]. Rather, it is more related to the articulatory properties of the voiced consonants. The laryngeally-based model can explicate the exact process of tone split in Suzhou Wu. The voiced consonant was articulated with lower larynx and slack vocal folds, which caused concomitantly low pitch and breathy voice in the following vowel. Later, low pitch and breathy voice were phonologized and the tonal system was split into two registers [2]. Looking further into the future, this and other studies [12, 13] suggested that the breathy voice would eventually disappear.

EGG measurements and their correlations with acoustic measurements of tones in Suzhou Wu Chinese were examined in this paper. Low register tones were lower in CQ and PIC than the high register tones, indicating breathy voice in low register tones. The less difference between these two groups of tones for female speakers may be a sign of the ongoing loss of the breathy voice. The characteristics of breathy voice as manifested by the correlations of EGG and acoustic measurements supported the laryngeally-based model of tone split. This paper provides inspiring evidence on the evolution of tone systems of Suzhou Wu. Further investigation is underway to compare the phonation distinctions in old speakers with that in middle-aged and young speakers, in order to confirm whether these is an ongoing loss of the breathy voice in Suzhou Wu and to draw a fuller picture of tonal evolution in Wu Chinese.

5. Acknowledgements

This study was supported by the RGC General Research Fund 2019/20 Project Reference 14607619 awarded to the last author and the CUHK Research Fellowship Scheme 2019-20, 2020-21 awarded to the first author, and also partially supported by the Major Program of the National Social Science Fund of China (13&ZD189) and a project from Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences (2017STD006) awarded to the third author.

6. References

[1] A. S. Abramson and T. Luangthongkum, "A fuzzy boundary between tone langauges and voice register languages," in *Frontiers*

in Phonetics and Speech Science, G. Fant, H. Fujisaki, and J. Shen, Eds. Beijing, China: The Commercial Press, 2009, pp. 149–155.

[2] M. Mazaudon, “Paths to tone in the Tamang branch of Tibeto-Burman (Nepal),” in *The Dialect Laboratory: Dialects as a Testing Ground for Theories of Language Change*, G. De Vogelaer and G. Seiler, Eds. Amsterdam / Philadelphia: John Benjamins Publishing Company, 2012, pp. 139–177.

[3] G. Thurgood, “Vietnamese and tonogenesis: Revising the model and the analysis,” *Diachronica*, vol. 19, no. 2, pp. 333–363, Dec. 2002.

[4] J. Laver, *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press, 1980, no. 31.

[5] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, “Acoustic Correlates of Breathy Vocal Quality,” *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 4, pp. 769–778, Aug. 1994.

[6] F. Shi, “Reanalysis of voiced initials in Suzhou Chinese [in Chinese],” in *Explorations in Phonetics*. Beijing: Beijing University Press, 1990, pp. 164–176.

[7] ——, “Acoustic characteristics of voiced obstruents in Suzhou Chinese [in Chinese],” *Studies in Language and Linguistics (Yuyan Yanjiu)*, no. 01, pp. 49–83, 1983.

[8] R. Iwata, “The physiological characteristics of voiced stops in Suzhou dialect [in Chinese],” in *Study on Chinese Abroad*, F. Shi, Ed. Beijing: Beijing Language College Press, 1995, pp. 48–58.

[9] 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, Jan. 2017.

[10] J. Tian and J. Kuang, “The phonetic properties of the non-modal phonation in Shanghainese,” *Journal of the International Phonetic Association*, pp. 1–27, 2019.

[11] Z. Chen, “An Acoustic Study of Voiceless Onset Followed by Breathiness of Wu Dialects: Based on the Shanghai Dialect [in Chinese],” *Studies in Language and Linguistics (Yuyan Yanjiu)*, vol. 30, no. 03, pp. 20–34, 2010.

[12] J. Zhang and H. Yan, “Contextually dependent cue realization and cue weighting for a laryngeal contrast in Shanghai Wu,” *The Journal of the Acoustical Society of America*, vol. 144, no. 3, pp. 1293–1308, Sep. 2018.

[13] J. Gao, “Sociolinguistic motivations in sound change: On-going loss of low tone breathy voice in Shanghai Chinese,” *Papers in Historical Phonology*, vol. 1, pp. 166–186, Nov. 2016.

[14] P. Boersma, “Praat: Doing phonetics with computers,” 2020.

[15] S. D. Khan, “The phonetics of contrastive phonation in Gujarati,” *Journal of Phonetics*, vol. 40, no. 6, pp. 780–795, Nov. 2012.

[16] N. Henrich, C. d’Alessandro, B. Doval, and M. Castellengo, “On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation,” *The Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1321–1332, Feb. 2004.

[17] C. Herbst and S. Ternström, “A comparison of different methods to measure the EGG contact quotient,” *Logopedics Phoniatrics Vocology*, vol. 31, no. 3, pp. 126–138, Jan. 2006.

[18] P. Keating, C. Esposito, M. Garellek, S. D. Khan, and J. Kuang, “Phonation Contrasts Across Languages,” in *Proceedings of ICPhS*, Hong Kong, China, 2011, pp. 1046–1049.

[19] A. Michaud, “A measurement from electroglottography: DECPA, and its application in prosody,” in *Speech Prosody 2004*. Nara, Japan: International Speech Communication Association (ISCA), 2004, pp. pp. 633–636.

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

[21] S. N. Wood, *Generalized Additive Models: An Introduction with R*, second edition ed. Boca Raton: CRC Press/Taylor & Francis Group, 2017.

[22] R Core Team, “R: A language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria, 2021.

[23] R. H. Baayen, J. van Rij, C. de Cat, and S. Wood, “Autocorrelated Errors in Experimental Data in the Language Sciences: Some Solutions Offered by Generalized Additive Mixed Models,” in *Mixed-Effects Regression Models in Linguistics*, D. Speelman, K. Heylen, and D. Geeraerts, Eds. Cham: Springer International Publishing, 2018, pp. 49–69.

[24] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, “Itsadug: Interpreting time series and autocorrelated data using GAMMs,” 2020.

[25] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, “VoiceSauce: A program for voice analysis,” in *Proceedings of ICPhS XVII*, Hong Kong, China, 2011, pp. 1846–1849.

[26] R. H. Colton and E. G. Conture, “Problems and pitfalls of electroglottography,” *Journal of Voice*, vol. 4, no. 1, pp. 10–24, Jan. 1990.

[27] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. C. Guiod, and S. L. Goldman, “Comparisons Among Aerodynamic, Electroglottographic, and Acoustic Spectral Measures of Female Voice,” *Journal of Speech, Language, and Hearing Research*, vol. 38, no. 6, pp. 1212–1223, Dec. 1995.