Journal of Phonetics 98 (2023) 101239

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

The change in breathy voice after tone split: A production study of Suzhou Wu Chinese

Chunyu Ge [a], Wenwei Xu [a], Wentao Gu [b], Peggy Pik Ki Mok [a,*]

[a] *The Chinese University of Hong Kong, Shatin, Hong Kong*
[b] *Nanjing Normal University, Nanjing, Jiangsu 210097, China*

ARTICLE INFO

Article history:
Received 13 March 2022
Received in revised form 13 March 2023
Accepted 1 April 2023

Keywords:
Breathy voice
Tone split
Wu Chinese

ABSTRACT

In some languages, breathy voice plays a pivotal role in tone split. After tone split, breathy voice can undergo further changes. Suzhou Wu Chinese used to have a voicing contrast in initial obstruents, which has transphonologized to a tone contrast and resulted in a two-way tone split, with breathy voice in the low register tones. This study investigates the change in breathy voice after the tone split in Suzhou Wu with apparent-time data from speakers from three age groups. Simultaneous audio and electroglottographic recordings were collected. Principal component analysis and linear discriminant analysis conducted on the acoustic measurements indicate that breathy voice is used less by younger speakers. Generalized Additive Mixed Models were conducted to reveal the changes in breathy voice during the time course of the vowel with regard to different low register tones. It is also found that T2 and T8 are undergoing a decrease in breathy voice with tone changes, but breathy voice is decreasing without tone change in T6. Younger female speakers are ahead of younger male speakers in the decrease in breathy voice. This paper provides a valuable investigation of the change in breathy voice after tone split and contributes to our understanding of the development of phonation types.

1. Introduction

In East and Southeast Asian languages, the devoicing of voiced initial consonants often results in the transphonologization of a voicing contrast to a tone contrast on the following vowel (Haudricourt, 1954, 1972; Maspero, 1912; Matisoff, 1973). This process of the emergence of tone contrast from earlier non-tone contrast is usually called tonogenesis (Michaud & Sands, 2020; Thurgood, 2020). Although tonogenesis can be induced by other factors, such as final laryngeal consonants and vowel intrinsic pitch (Michaud & Sands, 2020), this study focuses on tone split, which refers to the two-way tone contrast induced by the voicing contrast of the initial consonants. A growing body of evidence has raised the hypothesis that tone split may be mediated by a phonation stage, in which breathy voice induced by voiced consonants plays a pivotal role. When tone split is complete, breathy voice can undergo further development. The further change in breathy voice may interact with laryngeal features in consonants and tones. However, most of the time, we only observe the end result of the development of breathy voice, and few studies have been done on a change in progress. This paper presents a case study of Suzhou Wu Chinese, in which the breathy voice induced by formerly voiced consonants is changing. The rest of the Introduction will review the role of breathy voice in tone split and the possible outcomes of further change in breathy voice after tone split. Previous studies on breathy voice in Wu Chinese and the motivation for choosing Suzhou Wu will be summarized in the next section. The research question of this study is to investigate whether younger speakers of Suzhou Wu differ from older and middle-aged speakers in the production of breathy voice and whether they produce less breathy voice. This will be laid out in more detail in the final part of the Introduction.

1.1. The role of breathy voice in tone split

Tone split is often accompanied by a phonation stage, in which syllables with originally voiceless initial consonants retain modal voice and those with formerly voiced consonants develop breathy voice. This is especially true in East and

---

* Corresponding author.
E-mail address: peggymok@cuhk.edu.hk (P.P.K. Mok).

https://doi.org/10.1016/j.wocn.2023.101239
0095-4470/

Southeast Asian languages (Haudricourt, 1965). The process is represented schematically in (1). At first, the distinction is between voiceless (T) and voiced (D) obstruents. The voiced obstruents then devoice, becoming voiceless with breathy voice and a low tone on the following vowel (v̤̀), and the vowel following the originally voiceless obstruents has a high tone (V´) and with modal voice. The breathy voice can change further after tone split (there is more discussion of this in Section 1.2 below). There is some evidence of the role of breathy voice in tone split. For example, tone split caused by the merger of voiceless and voiced initials is in progress in Tamang-Gurung-Thakali-Manangke (TGTM) languages, a group of Tibeto-Burman dialects spoken in Nepal (Mazaudon, 2005, 2012). Tone 1 and Tone 2 are associated with voiceless initial consonants, while Tone 3 and Tone 4 are associated with the formerly voiced ones. A detailed acoustic analysis revealed that a bundle of features, including allophonic voicing, pitch and breathy voice, are used to characterize Tone 3 and Tone 4 in Risiangku Tamang (Mazaudon & Michaud, 2008). Similarly, the voicing contrast of the initial consonants has also been lost and resulted in tonogenesis in Kammu, a Mon-Khmer language spoken mostly in Laos (Kirby, Pittayaporn, & Brunelle, 2023; Svantesson, 1989; Svantesson & House, 2006). It is reported in some Kammu dialects that the voiced stops are devoiced, with accompanying breathy voice and low pitch (Premsrirat, 2003).

TV > TV́ > TV́ (1)

DV > Tv̤̀ > Tv̤̀/THV̀/TV̀

A two-way tone split caused by the devoicing of voiced initial consonants also occurred in Late Middle Chinese (8th-12th century CE) (Ferlus, 2009; Pulleyblank, 1978). Voiceless initial consonants gave rise to tones with higher pitch (Yin 阴), while voiced ones gave rise to tones with lower pitch (Yang 阳). The idea that the tone split was accompanied by an early stage of breathy voice was entertained as early as Pulleyblank (1978). He proposed that voiced initial consonants induced breathy voice on the following vowel, like in Wu Chinese. He also drew a parallel between breathy voice and the chest register in Mon-Khmer languages (Henderson, 1952; Shorto, 1966), which was also derived from voiced consonants. Impressionistic description of breathy voice in syllables with voiced initial consonants in Wu Chinese was first provided by Liu (1925) and Chao (1928), and experimental evidence was later provided by Cao and Maddieson (1992), which will be discussed below. The plausibility of this proposal is further confirmed by the presence of breathy voice in other Chinese dialects, such as Gan and Xiang dialects (see Zhu (2010), who used the term "slack voice" instead of breathy voice).

There is even a stage of breathy voice in the emergence of tone induced by contrasts other than onset voicing. In Phnom Penh Khmer, the trill /r/ in onset position is lost in colloquial speech, resulting in aspiration, lower fundamental frequency (henceforth f0), breathy voice, and diphthongization (Kirby, 2014). These cues all contribute to the contrast between /CrV/ and /CV/ in perception. Breathy voice and aspiration are employed as redundant cues, while f0 is the most salient cue (Kirby, 2014). Kirby accounted for the incipient tonogenesis with the hypothesis that the frication noise often found in trills extended to the vowel and caused breathy voice and low pitch.

It should be noted that breathy voice is also involved in registrogenesis (Brunelle & Tạ, 2021; Diffloth, 1982) in which the onset voicing is transphonologized to a register contrast in the following vowel, where the vowel following voiced onsets is more centralized and later causes a split of the vowel system. (For a thorough review of register and registrogenesis, see Brunelle and Tạ (2021).) For example, in Mon-Khmer languages, Huffman (1976) proposed a five-stage model of the evolution of register. In the second or "transitional" stage, voiced initial consonants induced slight aspiration and a lax breathy voice in the following vowels. This has been supported by experimental evidence. Although in most dialects of Khmer the vocalic split is complete, Wayland and Jongman (2003) found breathy voice in the low register in the Chanthaburi dialect of Khmer.

Based on the parallels of breathy voice in registrogenesis and tone split, some authors proposed a unified account for both: a phonation stage mediates the original voicing contrast and the split of the vowel system or tone split (Michaud, 2012; Thurgood, 2002). This "laryngeally-based" account is physiologically plausible in that the production of voiced initial consonants may have a lower laryngeal position and slacker vocal folds, which could lead to a larger glottal opening and thus breathy voice. Therefore, based on this account, the voiced initial consonants first induce breathy voice in the following vowel, which later develops into a low tone or a more centralized vowel. Multiple acoustic cues can co-exist to contribute to the contrast (Thurgood, 2002; Chen, 2015; Mazaudon, 2012). It should be pointed out that not all languages that have undergone tone split had a phonation stage in which breathy voice is involved. Evidence is also lacking of an automatic relationship between voicing and breathy voice (See Brunelle, Tạ, Kirby, & Giang, 2020; Brunelle, Brown, & Phạm, 2022 for more discussion). For example, the tone split in Africaans goes directly from the consonant voicing contrast to a tone contrast (Coetzee, Beddor, Shedden, Styler, & Wissing, 2018) and no phonation is involved in the incipient tone split in Seoul Korean (Bang, Sonderegger, Kang, Clayards, & Yoon, 2018), although it has been reported earlier that different types of stops in Korean are pronounced with different voice quality (Cho, Jun, & Ladefoged, 2002). More recently, Kirby et al. (2023) investigated the production of tone and voice quality in Northern and Eastern Kmhmu' (also referred to as Khmu) and found no evidence of phonation difference. Nevertheless, there is evidence in other non-tonal languages (e.g., Yerevan Armenian, Seyfarth & Garellek, 2018) that voiced stops are pronounced with breathy voice, which could be a precursor to a potential tonal contrast. To sum up, whether a phonation stage is obligatory for tonogenesis and tone split is still unclear.

1.2. Further development of breathy voice after tone split

After tone split, breathy voice can undergo different changes with regard to consonant and tone (Pittayaporn & Kirby, 2017). First of all, breathy voice can be retained for a long time as an enhancement feature of the tone contrast, as represented by Tv̤̀`̤ in (1). Such is the case in Wu Chinese. The breathy voice was reported impressionistically early on by Chao (1928), and it has been repeatedly confirmed by later instrumental studies, as will become clear in the next section. It has only very recently been reported that the breathy voice in Shanghai Wu (Gao, 2016; Zhang & Yan, 2018) and Lili Wu (Shi, Chen, & Mous, 2020) is disappearing, and only among younger speakers (aged below 30). Therefore, it is likely that the breathy voice has been well preserved for centuries.

On the other hand, breathy voice can be reanalyzed as aspiration of the initial consonant, thus changing the originally voiced consonants to voiceless aspirated consonants, represented by TH in (1). The process can be across-the-board, in which voiced consonants become voiceless aspirated consonants for all tones. This has been found in several Chinese dialects, for example, the Hakka dialect (Hashimoto, 1973, 1992), the Gan dialect (Sagart, 2002), and the Jin dialect (Sagart, 1990). In addition to Chinese dialects, this situation is also commonly found in Southeast Asian languages. In the She language (畲语) spoken in southern China, the previously voiced initial consonants have become voiceless aspirated consonants, while in other Hmong-Mien languages voiced initial consonants have become voiceless unaspirated consonants or voiceless consonants with breathy voice (Mao & Meng, 1986). The change of voiced initials to voiceless aspirated initials has also been reported for Central Thai (Abramson & Erickson, 1992) and Lao (Osatananda, 1997). This process can also be interrupted by other factors, which results in voiceless aspirated consonants, but only in certain tones. In Mandarin Chinese[1], only the originally voiced obstruents in Tone 2 became voiceless aspirated consonants, and no aspiration has been found in other tones. Pulleyblank (1978) proposed that breathy voice was lost in these other tones earlier than in Tone 2 due to dissimilation to the syllable-final /-h/. The reanalysis of breathy voice to aspiration in Tone 2 is prevalent in Northern Mandarin dialects of Chinese.

Tone change can also influence the development of breathy voice. In some languages, the loss of breathy voice can be accelerated by tone change. Mazaudon (2012) reported that Tone 3 and Tone 4 in Tamang are associated with formerly voiced initial consonants and both have breathy voice and a low pitch. However, in some dialects, Tone 3 or Tone 4 is no longer produced with breathy voice. For example, breathy voice is only retained in Tone 3 in Taglung Tamang and Marpha Thakali. Similarly, in Manang, there is only breathy voice in Tone 4 but not in Tone 3. Tone 4 in Taglung Tamang and Marpha has changed to a high falling tone [51] and Tone 3 in Manang has become [54]. Mazaudon (2012) proposed that when the pitch changes to a high pitch, breathy voice is no longer compatible with the articulation because breathy voice's articulatory settings of glottal opening and slacker vocal folds would inhibit the rate of vibration of the vocal folds. Therefore, as one of the features contributing to the tone contrast, a change in breathy voice and its possible loss may be influenced by tone change.

These findings show that breathy voice can have two possible outcomes, either to be preserved or to be lost. As a phonation type, breathy voice can interact with other voice source features. It can be transformed into aspiration of the initial consonant. The pitch of tones may also exert an influence on breathy voice and in some cases facilitate its loss.

1.3. The breathy voice of the low register tones in Wu Chinese

Wu Chinese is a dialect group spoken in eastern China, mainly in the Southern Jiangsu Province, Zhejiang Province and Shanghai. One of the defining features of Wu Chinese is that it retains the voicing contrast of obstruents in Middle Chinese (4th-12th century CE). There is a co-occurrence restriction between initial consonants and tones in Wu Chinese. Wu Chinese has a three-way contrast in plosives, with voiceless unaspirated plosives, voiceless aspirated plosives and voiced plosives. Tones in Wu Chinese can be divided into high and low registers, and voiceless unaspirated and aspirated initial consonants only co-occur with tones in the high register, whilst voiced consonants can only appear with the low register tones. Moreover, the previously voiced obstruents are now realized with no pre-voicing but with non-modal phonation on the following vowel. The non-modal phonation was initially termed as "voiceless sound with voiced aspiration" (清音浊流, Chao (1928, 1967), translation by Ren (1992, p. 9)). The non-modal phonation was also referred to as slack voice (Ladefoged & Maddieson, 1996, pp. 64–65), whispery voice (Rose (1989) for Zhenhai Wu, and Tian and Kuang (2021) for Shanghai Wu), and more often as breathy voice (see Tian and Kuang (2021) and references therein). This study uses the term breathy voice to refer to the non-modal phonation, yet with caution, as it may have some differences from typical breathy voice which is mainly caused by the larger glottal opening than modal voice, as found in Gujarati (Khan, 2012).

Suzhou Wu is a Northern Wu Chinese dialect spoken in Suzhou, a city to the west of Shanghai. There are seven tones in Suzhou Wu, including five unchecked tones, and two checked tones which occur only in syllables with a final glottal stop, as shown in Table 1. The tone names in the head row are inherited from Chinese historical phonology and their meanings (in italics) are thought to be indicative of their original tonal values. The tonal values of Suzhou Wu are transcribed using Chao's tone numerals (1930b; Chao, 1930a), where 5 represents the highest pitch level and 1 the lowest. As in other Wu dialects, Suzhou Wu also has a tone system where the voicing of onset consonant, pitch, and phonation type are interdependent. T1, T3, T5, and T7 are high register tones with modal voice and can only co-occur with voiceless unaspirated and aspirated initial consonants, while T2, T6, and T8 (historically T4 merged with T6) are low register tones with breathy voice which can only co-occur with voiced initial consonants. While T1, T2, T3, T5, and T6 are unchecked tones which appear in open syllables or syllables ending with a nasal coda, T7 and T8 are checked tones which only appear in syllables with a glottal stop coda and are shorter in duration. Although initial consonant voicing, phonation type and pitch are interdependent, f0, as the acoustic correlate of pitch, is the dominant cue for tones in Wu Chinese. For example, when listeners of different language backgrounds were asked to identify whether the initial consonant was voiced, the results showed that only a

---

[1] There are four tones in Mandarin Chinese. Pingsheng (平声, level) in Middle Chinese split into Tone 1 (high level) and Tone 2 (low rising) in Mandarin. Shangsheng (上声, rising) with voiced initials merged with those in Qusheng (去声, entering), which gave rise to Tone 4 (high falling) in Mandarin. Tone 3 (dipping) consists of Shangsheng (上声) with originally voiceless and sonorant initials only (Pulleyblank, 1978).

Table 1
Tones in Suzhou Wu, their pitch values and consonant voicing co-occurrence.

	Ping平声 *level*	Shang上声 *rising*	Qu去声 *departing*	Ru入声 *entering*
high register, voiceless	**T1** [44]	**T3** [51]	**T5** [412]	**T7** [$\underline{55}$ ˀ]
low register, voiced	**T2** [223]		**T6** [231]	**T8** [$\underline{23}$ ˀ]

low f0 can induce the response of voiced consonants, especially for listeners with a language background in Wu Chinese (Cao, 1987). This indicates that the tone split in Wu Chinese (especially for Northern Wu) is already complete.

Compared to Shanghai Wu, the most well studied Wu dialect, studies on Suzhou Wu have only investigated a few speakers and a comprehensive study is still lacking. More importantly, Suzhou Wu has been chosen for this study because it has a more conservative tone inventory than Shanghai Wu. Suzhou Wu has seven tones with three pairs (T1 vs. T2, T5 vs. T6, T7 vs. T8) in different registers. For Shanghai Wu, most of the studies focused on the unchecked tone pair (Tone 2 vs. Tone 3), while our study takes all the tone pairs (both unchecked and checked) in Suzhou Wu into consideration, thus providing a more comprehensive picture of the change in breathy voice. Moreover, Suzhou Wu tones are undergoing some changes in pitch, as reported by Bei (2011). The pitch of the low register tones is slightly higher for younger speakers than for older speakers, especially at the beginning of the vowel. This potential tone change might have an influence on the change in breathy voice, as demonstrated by the change in breathy voice in Tamang (Mazaudon, 2012), which is dependent on a change in pitch contours. Additionally, the checked tones have a glottal coda, which can often be implemented as creaky voice. Although it has been reported by Gao and Kuang (2022) that the glottal stop in the checked tones of Shanghai Wu has mostly been lost, close inspection during segmentation showed that the glottal stop in the checked tones of Suzhou Wu is largely intact. It can be realized as a glottal stop or creaky voice at the end of the vowel. With conflicting articulatory settings within a short duration, the breathy voice in T8 would also be expected to be difficult to maintain. Therefore, by including all the tone pairs in the Suzhou tone inventory, the study investigates whether the change in breathy voice is at the same pace for all the low register tones.

Thus, it is of interest to investigate the development of breathy voice in Wu Chinese after tone split and Suzhou Wu is ideal as a representative dialect of Wu Chinese, as it is more conservative than Shanghai Wu and is undergoing potential tone change. As reviewed above, breathy voice could be retained as an enhancement cue to the tone contrast, or it can be reanalyzed as consonant aspiration. It can also disappear, which may be influenced by tone change.

There have been some studies of the breathy voice in Wu Chinese, especially for Shanghai Wu. A summary of these studies is shown in Table 2. Apart from VOT and f0, commonly used acoustic measurements are H1-H2 (as f0-H2 by Shi, 1990, the difference between the amplitudes of the first and the second harmonics) and H1-A1 (as F1-f0 in Shi, 1990 and H1-F1 in Ren, 1992, the difference between the amplitude of the first harmonic and that of the harmonic near the first formant). Several articulatory investigations have also been made. Air pressure was measured using Air Flow/Air Pressure by Cao and Maddieson (1992) and intraoral pressure (P_o) by Iwata (1995). Electroglottography (EGG) was also used and Open Quotient (OQ, the ratio of the open phase to an entire vocal vibration cycle, Holmberg, Hillman, Perkell, Guiod, & Goldman, 1995), Contact/Closed Quotient (the reverse of OQ), Speed Quotient (the ratio of the contacting duration to the decontacting duration) and Peak Increase Contact (PIC) were measured. Some studies also used other articulatory technologies such as electromyography (EMG, Iwata, 1995) and electrophotoglottography (EPGG, Gao, Hallé, Honda, Maeda, & Toda, 2011). Although different measurements have been used, some consistent acoustic and articulatory properties have been reported. For the breathy voice following voiced consonants, there is more energy in the first harmonic and more noise than for the modal phonation following voiceless unaspirated consonants. In terms of articulation, the vowel following the voiced consonants is pronounced with a larger OQ than that of the vowel following the voiceless unaspirated consonants.

It is widely reported that the low register tones in Wu Chinese are pronounced with breathy voice with larger amplitude in the first harmonic (H1) relative to that of H2 and to the amplitude of the harmonics adjacent to the first three formants (A1, A2, A3). The earliest experimental study by Cao and Maddieson (1992) reported lower H2-H1 (i.e., higher H1-H2) and F1-H1 (i.e., higher H1-A1) for the low register tones at 30 ms after the initial stop release, using data from four Wu dialects, including both Northern (Shanghai, Changyinsha, Ningbo) and Southern (Wenzhou) Wu dialects. Consistent results were also reported by Ren (1992), and more recently by Gao and Hallé (2017) and Tian and Kuang (2021) for Shanghai Wu, and Xu and Mok (2021) for Kunshan Wu.

Although most of the early studies concentrated on spectral tilt measurements, noise measurements are also important in Wu breathy voice. The voiced stops in Suzhou Wu are produced with stronger aspiration than voiceless unaspirated stops, but smaller aspiration than voiceless aspirated stops, which was demonstrated by Shi (1983) with a simple experiment. In the experiment, an ignited match was put in front of the speaker's mouth. While the flame was extinguished when pronouncing the aspirated stop, it only swayed for the voiced stop and unaffected by the voiceless unaspirated stop. Acoustic measurements have also shown that there is a strong noise component for the voiced stops. It has been reported that there is a stronger noise component in the breathy voice of the low register tones in Jiashan Wu (Jiang & Kuang, 2016) and Shanghai Wu (Tian & Kuang, 2021; Zhang & Yan, 2018) with a lower cepstral prominence peak (CPP) and/or Harmonic-to-Noise Ratio (HNR). Tian and Kuang (2021) reported that the noise component is even more important than spectral measures for Shanghai Wu. They conducted a linear discriminant analysis (LDA) and found that CPP and

Table 2
A summary of previous studies on the phonation distinction in Wu Chinese.

source	Wu variety	speakers	method	acoustic measurements	articulatory measurements
Shi, 1983	Suzhou	2 M, 2F	acoustic	VOT, f0, amplitude	N/A
Shi, 1990	Suzhou	3 M, 1F	acoustic	f0-H2, F1-f0	N/A
Cao & Maddieson, 1992	Shanghai, Changyinsha, Ningbo, Wenzhou	2 M, 2F for Shanghai, 2 for each other dialects	acoustic, airflow	H2-H1, F1-H1	AF/AP (Air Flow/Air Pressure)
Ren, 1992	Shanghai	3 M, 1F	acoustic, airflow, fiberscope and transillumination	H1-H2, H1-F1	OQ, SQ, maximum and minimum airflow
Iwata, 1995	Suzhou	not mentioned	fiberscope, EMG	N/A	P_o, EMG
Chen, 2010	Shanghai	not mentioned	acoustic, EGG	H1-H2	OQ
Chen, 2011	Shanghai	2 M, 4F	acoustic	ACT (After closure time), f0, H1-H2	N/A
Gao et al., 2011	Shanghai	6 young (3 M), 4 old (3 M)	acoustic, EPGG (electrophotoglottography)	H1-H2, HNR	location and amplitude of glottal opening maximum
Gao, 2016	Shanghai	6 young (24–25, 3 M), 4 old (64–72, 3 M)	acoustic, EGG	H1-H2, H1-A1, H1-A2, CPP (Cepstral Prominence Peak), F1	OQ
Jiang & Kuang, 2016	Jiashan	6 M, 6F	acoustic	H1*–H2*[a], H1*-A1*, H1*-A2*, H1*-A3*, CPP	N/A
Gao & Hallé, 2017	Shanghai	12 young (6 M); 10 old (4 M)	acoustic, EGG	H1-H2, HNR,	OQ
Zhang & Yan, 2018	Shanghai	5 M, 5F	acoustic	VOT, f0, H1*–H2*, CPP	N/A
Kuang, Tian, & Jiang, 2019	Shaoxing	4 M, 4F	acoustic	H1*–H2*, H1*-A1*, H1*-A2*, H1*-A3*, CPP	N/A
Tian & Kuang, 2021	Shanghai	20 M, 32F	acoustic, EGG	H1*–H2*, H1*-A1*, H1*-A2*, H1*-A3*, CPP, HNR	CQ, SQ, PIC
Zhang, Xu & Mok, 2019	Wenzhou	4 young (2 M), 4 old (2 M)	EGG	N/A	CQ
Shi et al., 2020	Lili	20 young, 20 middle, 20 old	acoustic	H1*–H2*	N/A
Xu & Mok, 2021	Kunshan	6 young (3 M), 6 old (3F)	acoustic	H1*-A1*, H1*-A2*, H1*-A3*, H2*-H4*, H4*-H2K*, H2K*-H5K*, H1*-H4*, H1*-H2K*, H1*-H5K*, HNR, CPP	N/A

[a] The asterisk indicates formant corrected values of the spectral measurements.

HNR contribute more to the discrimination of breathy voice from modal voice.

Breathy voice in Wu Chinese also differs from modal voice in articulatory measurements. Iwata (1995) investigated the voiced consonants in Suzhou Wu using fiberscopy and electromyography (EMG) and found that the activities of the cricothyroid (CT) and vocalis (VOC) muscles are inhibited for voiced consonants, and they are characterized by an aryepiglottic constriction and lowering of the larynx. This is consistent with the tension hypothesis proposed by Halle and Stevens (1971) that the low pitch related to voiced consonants should be attributed to the lower CT activity and the lower larynx position. Using airflow and transillumination data measured at the initial and medial points of the vowel, Ren (1992) found that voiced stops in Shanghai Wu are pronounced with more abducted glottis than voiceless stops, but they are not as abducted as voiceless aspirated stops. Gao et al. (2011) reported earlier and lower glottal-opening peaks for syllables with breathy voice in Shanghai Wu using electrophotoglottography (EPGG). Larger OQ for breathy voice was reported by Ren (1992) using airflow, and also by Chen (2010), Tian and Kuang (2021) and Zhang, Xu, and Mok (2019) using EGG.

The time course of the breathy voice in Wu has also been investigated. It was reported by Shi (1983), Cao and Maddieson (1992), Ren (1992), and Gao and Hallé (2017) that breathy voice was not limited to the consonant release and continued at least until the midpoint of the vowel. Breathy voice can even continue throughout the syllable for checked tones in Kunshan Wu (Xu & Mok, 2021). Moreover, it has also been reported that the difference between breathy voice and modal voice decreases towards the end of the syllable (Cao & Maddieson, 1992; Ren, 1992).

It should be noted that not all studies have consistent findings, and the influence of gender and age should be highlighted. Shi (1990) found that although H1-H2 was higher for the low register tones for three male speakers in Suzhou Wu, it was lower than the high register tones for the one female speaker. Chen (2011) also showed that H1-H2 was higher for voiced consonants than voiceless consonants in syllables with /o/, but not for those with /ɪ/. More recent studies revealed less difference between the low and high register tones in phonation for younger speakers than older speakers. Zhang and Yan (2018) reported no significant difference in H1*–H2* and CPP with data from younger Shanghai Wu speakers (aged 19–30). Furthermore, by examining the production of older and younger speakers, Gao (2016) reported that the breathy voice in Shanghai Wu is disappearing. She also found that the loss of breathy voice is more advanced for female speakers and attributed it to the contact with Standard Chinese. The decrease in breathy voice was also reported by Shi et al. (2020) for Lili Wu. Their data also showed that the voicing ratios of voiced fricatives were decreasing but no significant difference was found between the VOTs of voiceless unaspirated and voiced stops. Therefore, the breathy voice in Shanghai and Lili Wu is gradually disappearing and it does not seem to be reanalyzed as aspiration of the initial consonant, as it is in the case of Hakka and Gan Chinese.

1.4. The present study

As pointed out in Section 1.1, breathy voice can interact with tone and the laryngeal features of initial consonants after tone split. The first aim of the present study is to document the current state of tone and initial consonant in Suzhou Wu and to investigate whether speakers of different age groups pronounce tone and initial consonant differently. Apart from that, this study will concentrate on the difference in phonation type among speakers of different age groups. Although a number of studies have been done on breathy voice in Wu Chinese, they had several limitations. First, although recent studies in Shanghai Wu are sophisticated in their measurements (e.g., Tian & Kuang, 2021), the spectral measurements in Suzhou Wu have not been corrected for formant frequencies. In Shi (1990), the uncorrected measurements were aggregated for vowels of different heights, which makes the results even more tentative. In other studies, inappropriate rhymes were used. For example, the rhymes used in Chen (2011) were /ɪŋ/ and /oŋ/, and the nasal final would introduce nasalization to the vowel, especially for /ɪŋ/, which was described as a nasalized vowel elsewhere. Nasalization would influence the measurement of H1-H2 as an indicator of breathiness (Garellek, Ritchart, & Kuang, 2016; Simpson, 2012). Apart from these, recent studies on the change in breathy voice may not be sufficiently fine-grained in the time course of vowels. For example, LDA was conducted on the averaged values of the acoustic measurements over the entire vowel in Gao (2016), and only three measurement points were used in Shi et al. (2020). These approaches may miss critical information as they cannot reveal how the time course of breathy voice changes within a syllable. Furthermore, as demonstrated by Kuang (2017), voice quality can covary with pitch. It becomes creakier when pitch lowers. This is especially pertinent to Suzhou Wu because in Suzhou Wu, tones in the high and low register tones have distinct pitch contours, unlike T1 and T3 in Shanghai Wu, which both have a rising pitch contour. Additionally, since we included checked tones in this study, the glottal stop coda can introduce glottalization, which may interact with the breathy voice in T8. Therefore, to overcome the limitations, the current study used CV syllables only and corrected the spectral measurements for vowel formant frequencies using VoiceSauce (Shue, Keating, Vicenik, & Yu, 2011). Due to the different pitch contours between the high and low register tones in Suzhou Wu, the phonation measurements were compared combining high and low register tones using Principal Component Analysis (PCA) and LDA. Also, in light of the fact that phonation can covary with pitch, and in order to demonstrate the interaction between breathy voice and the glottal stop coda in the checked tones, the time course of breathy voice was also investigated using Generalized Additive Mixed Models (GAMM, Wood (2017); Wieling (2018)).

To investigate changes in breathy voice in Suzhou Wu, this study aims to answer the following questions.

Table 3
Number of speakers in different genders and age groups.

Age group	Female	Male	Mean age (age range)
younger	5	5	24 (21–29)
middle-aged	10	6	49.8 (43–54)
older	5	5	71.3 (65–84)

1. Are there any differences between younger, middle-aged and older speakers in initial consonants and pitch contours of tones?
2. Is the breathy voice of the low register tones used less by younger speakers than by older and middle-aged speakers?
3. What are the differences in the time course of breathy voice between younger, middle-aged and older speakers? Do older and middle-aged speakers have more extensive breathy voice in low register tones than younger speakers?

To answer these questions, simultaneous audio and electroglottographic (EGG) data of monosyllabic words were collected from younger, middle-aged, and older speakers. By including speakers from three age groups and all the tones in the tone inventory, this study will contribute to our understanding of the change in breathy voice after tone split, and the change in phonation type in general.

2. Method

2.1. Speakers

Thirty-six native speakers of Suzhou Wu were recruited to participate in the production experiment. The ages of these speakers at the time of recording ranged from 21 to 84. The speakers were divided into three age groups, where older speakers were those aged over 60 (μ = 71.3, σ = 6.15), younger speakers were aged under 30 (μ = 24, σ = 2.71), and the middle-aged speakers were in-between (μ = 49.8, σ = 2.99). Table 3 summarizes the number of speakers in each gender and age group. The gender ratio was balanced for the older and younger speakers, while there were more females in the middle-aged group. All speakers were native speakers of Suzhou Wu and spoke Suzhou Wu to their family members. They also spoke Mandarin Chinese and all the younger speakers had also learned English in school from an early age.

2.2. Materials

Test words used in this study were monosyllabic words of CV (unchecked tones) or CVʔ (checked tones) structures. The onsets were stops, fricatives, or zero, that is /p b t d k g f v s z ø ɦ/ (here ø indicates zero onset), and the rhymes were /ɑ ɪ ɛ/ for unchecked tones and /aʔ ɪʔ əʔ/ for checked tones. Syllables of all seven tones were included. Therefore, there should ideally be 12 tokens for each voiceless onset, and 9 for each voiced onset. However, due to phonotactic constraints, some combinations are not possible and there were 103 tokens in total. Affricate and nasal onsets were not used because there are only a few syllables with the low register tones. Aspirated stops were also not used because the aspiration would induce some breathy voice in the following vowel and complicate the data. Note that we treated voiced glottal fricative /ɦ/ as the counterpart of zero onset in the low register. The zero onset is usually realized as a glottal stop. Thus, the zero onset and the voiced glottal fricative is parallel to voiceless and voiced stops[2]. Although the articulation of the glottal fricative /ɦ/ should have a spread-glottis configuration, it is likely that it is no stronger than that in voiced stops, as there is also a similar configuration in voiced stops due to stronger aspiration. The aspiration of voiced stops was demonstrated by Shi (1983) match experiment, as reviewed in Section 1.3. On the other hand, no creaky voice was found after glottal stops. Therefore, the inclusion of the glottal sounds should have little effect on the results. Table A1 in the Appendix shows the test words.

2.3. Procedure

The monosyllabic test words in Chinese characters were printed on a sheet of paper and were read in isolation by the speakers. The order of the words was randomized, and all speakers read the material in the same randomized order. All speakers read the word list three times. Therefore, there were 309 tokens for each speaker.

Simultaneous audio and electroglottographic (EGG) recordings were made for all the speakers. The recordings were in two channels, with the first as audio signals and the second as EGG. EGG signals were recorded through the Glottal Enterprises EG2-PCX EGG equipment and audio signals using a Behringer ECM8000 microphone 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 is located. Both audio and EGG signals were sampled at 44,100 Hz with a 16-bit precision.

2.4. Measurements and analyses

The recordings were manually segmented in Praat (Boersma, 2020). Each syllable was segmented into a consonant interval and a vowel interval. For stops, VOT was marked between the release burst and the zero crossing of the first periodic pulse. For fricatives, the duration was marked as visible frication until the first periodic pulse of the vowel. The vowel interval was defined as the portion between the zero crossing of the first periodic pulse and the end of F2. During segmentation, all voiced initial consonants were found to be phonetically voiceless. The voiced portion during the voiced fricative was also segmented but it turned out that there were only a few voiced pulses before the vowel, which is not different from a voiceless fricative (but see Gao and Hallé, 2017 for different patterns in Shanghai Wu, which have longer voicing ratios for voiced fricatives). Another point worth noting is that the initial /ɦ/ has a very short duration of voiced aspiration (Schertz and Khan, 2020; Garellek, Chai, Huang, & Doren, 2021). Some tokens had a short interval of voiceless frication while some other tokens (especially those preceding the high front vowel /i/) had no frication at all. For the former, the

---

[2] Although the zero onset is sometimes realized as a glottal stop, it is weak in articulation and does not induce creaky voice in the following vowel. The voiced glottal fricative is used to indicate the presence of breathy voice. The zero onset and the voiced glottal fricative are usually treated as voiceless and voiced counterparts in previous literature on Wu Chinese (see Li, 1966 on Wenling Wu), parallel to other voiceless and voiced obstruent onsets.

frication was excluded from the vowel interval, and for the latter the whole syllable was taken as the vowel interval. The acoustic and EGG measurements were made during the vowel interval.

During segmentation, some of the tokens were found to have been mispronounced and these tokens were excluded. The mispronunciation can be classified into the following categories. In the first category, the speaker's pronunciation of the syllable was not the same as intended by the experimental design due to literary or colloquial readings[3]. Another category is that some speakers merged some tokens of T5 to T3, and these merged T5 tokens were also excluded. Some other tokens were simply that the speakers wrongly pronounced the character, either due to misidentification of the character or slips of the tongue. The judgements were made by the first and second authors, who are both native Wu speakers. The exclusion of these tokens resulted in 6.3% of data loss and there were 10,427 valid tokens in total.

Acoustic measurements were made using VoiceSauce (Shue et al., 2011). The acoustic measurements were averaged over ten intervals during the vowel duration. F0 was measured using the STRAIGHT algorithm, and formants using the SNACK Toolkit. Spectral tilt measurements (H1*–H2*, H1*-A1*, H1*-A2*, H1*-A3*) and noise measurements (CPP, HNR in 0–500, 0–1500, 0–2500, 0–3500 Hz bands) were measured. All spectral tilt measurements were corrected for vowel formants and their bandwidths using the algorithm in Iseli, Shue, and Alwan (2007) and marked with an asterisk (*). We included various acoustic measurements to capture different aspects of breathy voice. Breathy voice has higher H1*–H2* than modal voice, which is correlated with a larger opening of the glottis (Kreiman, Iseli, Neubauer, Shue, Gerratt, & Alwan, 2008) and decreased medial surface thickness (Zhang, 2016a). Higher H1*-A1* reflects larger posterior opening related to the membranous part of the vocal folds (Hanson, Stevens, Kuo, Chen, & Slifka, 2001), while H1*-A2* and H1*-A3* are thought to reflect abruptness of the vocal fold adduction (Hanson et al., 2001; Holmberg et al., 1995). Breathy voice also has a stronger noise component than modal voice, which can be quantified using Cepstral Peak Prominence (CPP, Hillenbrand, Cleveland, and Erickson (1994)) and Harmonic-to-Noise Ratio (HNR, de Krom (1993)).

EGG measurements were made using a MATLAB program. EGG signals were high-passed at 20 Hz. However, even after high-pass filtering, the EGG signals of some speakers were still too noisy. The EGG data of six speakers were removed. These included two older female speakers (F07, F20), one middle-aged female speaker (F13), and three younger speakers (F15, F16, M01). Thus, the EGG analyses consisted of the data of 15 female speakers and 15 male speakers. Note that the acoustic data of these speakers were not excluded. Each glottal cycle was identified in the signal using the EGG peak. The highest point of dEGG signal during each glottal cycle was taken as the dEGG peak. CQ was calculated based on the dEGG peak and 25% threshold of the current EGG cycle's amplitude, i.e., using the hybrid method (Henrich, d'Alessandro, Doval, & Castellengo, 2004; Herbst & Ternström, 2006; Howard, 1995). An example of EGG and dEGG signals of /pɑ1/ (top) and /bɑ2/ (bottom) produced by the same speaker is shown in Fig. 1. The solid blue lines are EGG signals and the orange dashed lines are dEGG signals. The solid vertical line represents the position of the dEGG peak and the vertical dashed line the position of the 25% threshold of the amplitude of the EGG peak. The interval between the previous dEGG peak and the 25% threshold is the closed phase, while that between the 25% threshold and the following dEGG peak is the open phase. The figure is representative in that /bɑ2/ has a longer open phase than /pɑ1/. The reason for not using the negative dEGG peak to determine the beginning of glottal opening (as done by Mazaudon and Michaud (2008), DiCanio (2009) and Tian and Kuang (2021)) is that the negative peak is usually not clearly defined, as can be seen from the dEGG signal in Fig. 1, especially in the bottom panel. Breathy voice has a smaller CQ than modal voice, indicating shorter glottal constriction (DiCanio, 2009; Mazaudon & Michaud, 2008). Like the acoustic measurements, the EGG measurements were averaged over ten intervals during the vowel interval.

2.5. Statistical analyses

To investigate whether the breathy voice is used less by the younger speakers, principal component analysis (PCA) was conducted using all the phonation measurements (both spectral and noise measurements) averaged over the duration of the whole vowel. All measurements were centered around the mean and scaled by one standard deviation within speaker. Data beyond 2.5 standard deviations were discarded and not included in the analyses. PCA was done separately for unchecked and checked tones. Linear discriminant analysis (LDA) was also conducted using all the averaged acoustic measurements as predictors, following Gao (2016) and Tian and Kuang (2021), in order to compare the results with those for Shanghai Wu.

Generalized Additive Mixed Models (GAMM, Wood (2017); Wieling (2018)) were fitted to acoustic and EGG measurements to compare the time course of breathy voice between two register tones and across age groups. The acoustic and EGG measurements used in GAMM were normalized within speaker. GAMMs were fitted using the `bam()` function in the `mgcv` (Wood, 2017) package in R (R Core Team, 2021). Visualization was realized with the help of the `itsadug` package (van Rij, Wieling, Baayen, & van Rijn, 2020). To fit GAMM on the entire data would result in many interactions and interpretation would be difficult. Moreover, to fit the data using the smooth over the year of birth as in Sóskuthy, Hay, and Brand (2019) would also not be appropriate. The reason is that the age distribution is sparse and modeling the smooth over such a sparse distribution of year of birth would result in overfitting. Therefore, separate GAMM was fitted for each age group and for unchecked and checked tones. Tone and Gender were combined to construct an ordered factor `toneGender`, to separate intercept and non-linear differences (Sóskuthy, 2021; Wieling, 2018). For each model, the factor toneGender was included as the intercept. The factor smooth over measure-

---

[3] In Chinese dialects, some characters have more than one pronunciation, in which the colloquial reading is used in an informal context and the literary reading is borrowed from the standard dialect and used in more formal contexts.

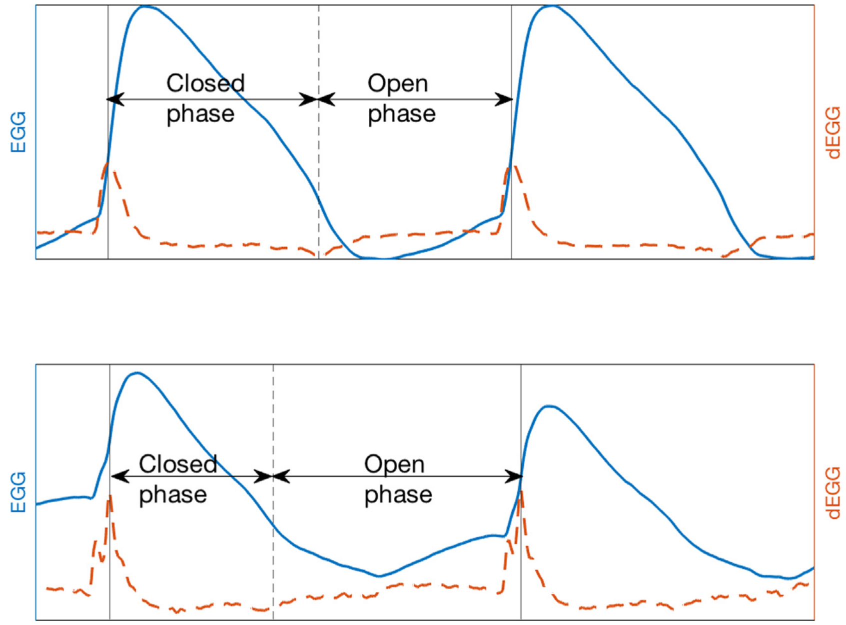


Fig. 1. Examples of EGG (solid blue) and dEGG (orange dashed) signals of /pɑ1/ (top) and /bɑ2/ (bottom).

ment point by toneGender was also included in the model. Gender was included because previous studies have shown that the disappearance of breathy voice in Shanghai Wu is more advanced for females (Gao, 2016). The model also included the random smooth over measurement point by word item and Repetition. The by-speaker random smooth was not included because gender and subject are covariates and the inclusion of subject resulted in failure to converge. The number of basis functions specifies how wiggly the curve is. Since the number of measurement points is 10, we set the number of basis functions `k` as 9. T1 and T7 were treated as reference levels for unchecked and checked tones respectively, so that the low register tones were compared to T1 (for unchecked tones) or T7 (for checked tones). T1 was chosen as the reference level because it is a high level tone and it has the most stable modal voice. For checked tones, T7 was treated as the reference level. For Gender, male was treated as the reference level. Therefore, the reference level is T1.Male or T7. Male for the ordered factor toneGender. The measurement point was from 0 to 9. The autocorrelated error was handled by specifying the AR1 correlation parameter rho as the error at lag 1 (the correlation between the values of immediately adjacent points), following the practice of Baayen, Vasishth, Kliegl, and Bates (2017). To illustrate the time course of phonation type, the fitted values and the confidence intervals were plotted using plot_smooth(). The difference curves between T1-T2, T1-T6 and T7-T8 pairs were also plotted to compare the time course of the differences between T1 and T2, T1 and T6, as well as T7 and T8 across age groups with the function `plot_diff()` in `itsadug`. As this study is exploratory in nature, p values will not be reported in the text because the interpretation of p values of multiple-level factors in GAMM is not straightforward (see Sóskuthy (2021) for a discussion of significance testing in GAMM), and only visualization is presented.

```
measurement ~ toneGender +
# factor smooth by tone and gender
s(point, by = toneGender, k = 9) +
# random smooth by word item and repetition
s(point, item, bs = "fs", m = 1) +
s(point, repetition, bs = "fs", m = 1)
```

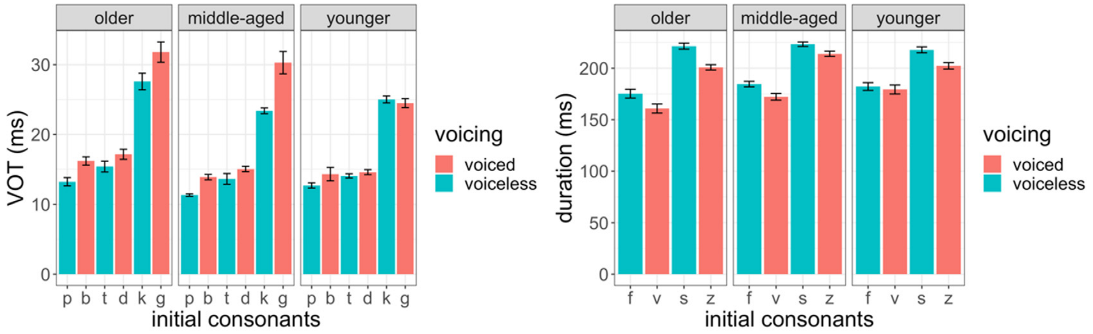
3. Results

3.1. The initial consonants and tones across age groups

First, we shall look at the initial consonants and the pitch contours of tones in Suzhou Wu across age groups, as these may interact with the change in breathy voice. Voiced initial consonants in Suzhou Wu are all phonetically voiceless and there is no voicing during closure for stops or in frication for fricatives[4]. For initial consonants, if the breathy voice has been reanalyzed as aspiration, there would be a remarkably longer interval between the stop release and voice onset (i.e., a longer VOT for stops or a longer duration of frication for fricatives) for voiced obstruents than voiceless obstruents. Also, if the voiced obstruents are changing to voiceless aspirated, the difference between voiced and voiceless obstruents would be larger for younger speakers than older and middle-aged speakers. Fig. 2 shows the VOT of voiceless and voiced stops and the duration of voiceless and voiced fricatives for the three age groups. The voiced and voiceless obstruents are represented by red and blue bars, respectively. For stops, linear mixed effects models were fitted on VOT, with voicing, place, age group, and the interaction between voicing and place as fixed effects, and subject and word item as random intercepts. The inclusion of voicing improved the model significantly

---

[4] The initial obstruents occurring in the low register tones are referred to in this study as voiced initial consonants, because (1) it is indicative of their historical origin and (2) they are phonetically voiced in intervocalic positions, though phonetically voiceless in isolation.



a. Stops

b. Fricatives

Fig. 2. Bar plot of VOT (in ms) of stop (a) and duration (in ms) of fricative (b) initial consonants for the different age groups. The error bar shows the standard error.

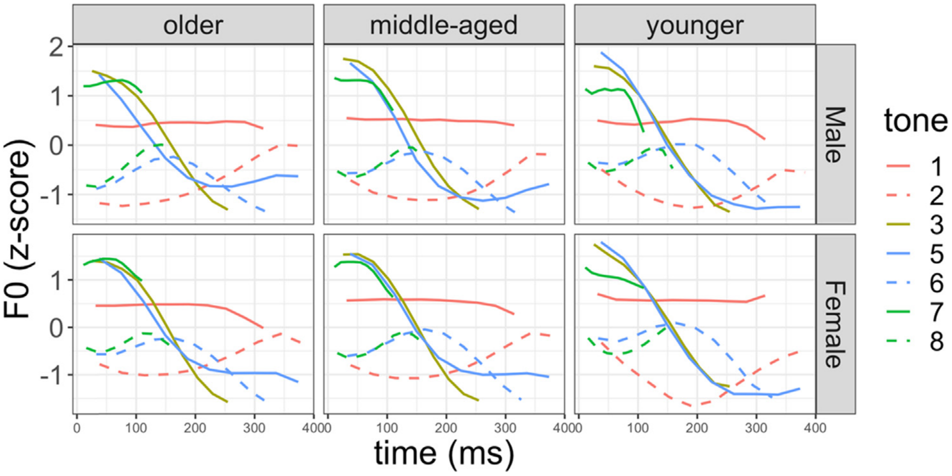


Fig. 3. Pitch contours of Suzhou Wu tones by gender and age groups.

($\chi^2(1) = 12.63$, $p < 0.01$), while the inclusion of age group ($\chi^2(1) = 2.95$, $p = 0.23$) and the interaction between voicing and place did not ($\chi^2(1) = 2.41$, $p = 0.29$). Nevertheless, the VOT of voiced stops is only 2.3 ms longer than for voiceless stops and the difference is too small to be called aspiration. As a matter of fact, slightly longer VOT for voiced stops was also reported by Shi (1983). The same was also found for the duration of fricatives[5]. Only voicing significantly improved the model ($\chi^2(1) = 5.40$, $p = 0.02$) and the duration of voiced fricatives is 13 ms shorter than voiceless fricatives. Moreover, although there are some small differences in duration between voiceless and voiced obstruents, age group does not have an influence. That is to say, the duration difference between voiced and voiceless obstruents is not changing across age groups. Therefore, we can conclude that the breathy voice has not been reanalyzed as voiceless aspiration.

For the pitch contours, there were also some differences between younger and older speakers, similar to that reported by Bei (2011). The averaged f0 contours for the three age groups are shown in Fig. 3. F0 was normalized by speaker. The general patterns across age groups are similar, although with notable differences, especially for the younger speakers. T1 and T6 are not changing in that T1 is level and T6 is rising-falling for all three age groups. There is a tendency that T3 and T5 are merging for the younger speakers, in that their pitch contours are becoming very similar. This was noted during segmentation, and some speakers confused T3 and T5 in their pronunciation. T5 tokens which were merged to T3 were excluded from analysis, as explained in Section 2.3. Nevertheless, both T3 and T5 are high register tones, and the potential merge will not influence our analysis of breathy voice in the low register tones. More importantly, there is another difference across age groups, which involves the low register tones (T2 and T8). For the two low rising tones T2 and T8, the point at which rising starts is gradually shifting towards the middle or even the end of the syllable from the older speakers to the middle-aged and younger speakers. The pattern is most remarkable for the younger speakers' T2, which is in a dipping shape. T2 seems to change from a rising tone to a falling-rising tone, and T8 is also showing such a trend, which is less dramatic. Moreover, inspection of the waveforms and spectrograms shows that the younger speakers pronounced T2 with heavy creaky voice in the middle of the syllable, especially female speakers, who almost always pronounced T2 with creaky voice. Fig. 4 shows a representative token of T2 produced by a younger female speaker. A large portion in the mid-

---

[5] Note that the glottal fricative and zero onset (glottal stop) were not included here. The reason is that it is often hard to identify separate intervals of glottal fricatives (the "visible aspiration", as in Shertz and Khan 2020:19). The realization of the voiced glottal fricative /ɦ/ and its voiceless counterpart /h/ in Suzhou Wu is different from that reported in Garellek et al. (2021) and it is presumably due to the convention that /ɦ/ is used to transcribe the breathy voice in Wu Chinese rather than a clearly defined segment (Chao, 1930b:217).

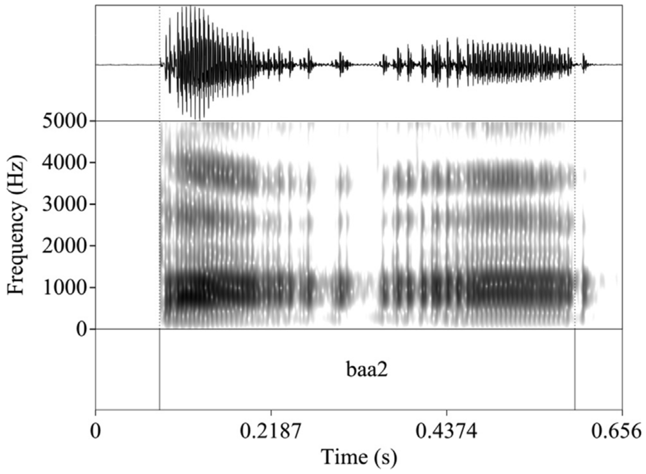


Fig. 4. The waveform and spectrogram of an example of T2 produced by a younger female speaker.

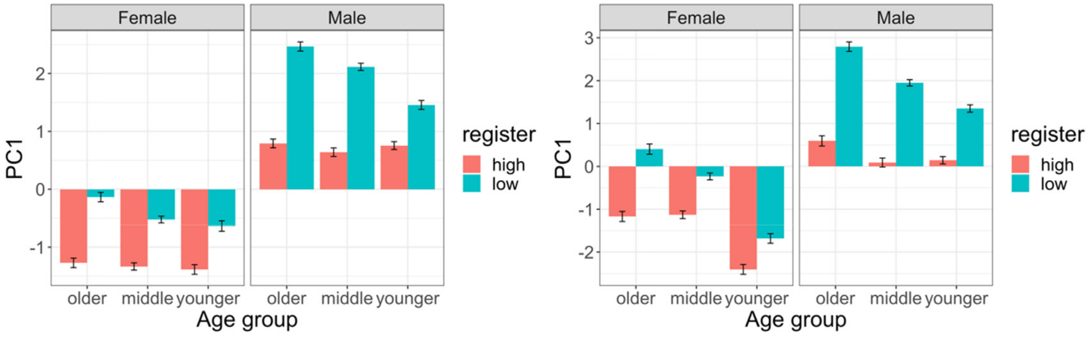
dle of the vowel shows highly irregular pulses, and only the start and end of the vowel show a regular waveform. What is also clear from the spectrogram is that at both ends of the vowel, the formant structure is clear and no aperiodic voice typical of breathy voice is visible. Thus, in this token, T2 is pronounced without breathy voice, but rather with creaky voice. On the other hand, although T8 has a similar change in the pitch contour as T2, there is no creaky voice, even for younger female speakers. The tone changes in T2 and T8 may influence the breathy voice in these two tones, especially T2, which is pronounced as creaky voice by younger females. Nevertheless, another low register tone T6 has similar pitch contours for the younger, middle-aged and older speakers. A change in breathy voice in T6, if any, would not be the consequence of a tone change in T6.

3.2. Breathy voice in the three age groups

3.2.1. The change in breathy voice across age groups

To compare the breathy voice across age groups, PCA was conducted on all phonation measurements. All measurements were averaged over the entire vowel. As the pitch contours of the high and low register tones in Suzhou Wu have quite different shapes, it is expected that f0 would play a larger role in distinguishing the two registers. Because our primary concern is on phonation, f0 and VOT were thus excluded. PCA was done separately for unchecked and checked tones. For unchecked tones, the first principal component (PC1) accounted for 50% of the variance and the second principal component (PC2) accounted for 18%, while for checked tones, PC1 accounted for 46% and PC2 accounted for 26%. For both the unchecked and checked tones, PC1 was highly correlated with noise measurements (HNR15: −0.91, HNR25: −0.91, HNR35: −0.87 for unchecked tones; HNR15: −0.91, HNR25: −0.90, HNR35: −0.85 for checked tones), and also with spectral tilts (H1*-A1*: 0.72, H1*-A2*: 0.73, H1*-A3*: 0.56, for unchecked tones; H1*-A1*: 0.61, H1*-A2*: 0.64, H1*-A3*: 0.56, for checked tones). The highest correlations were found between PC2 and H1*-A1* for both the unchecked (0.60) and checked tones (0.71). The PCA results show that noise measurements are the most important cues for breathy voice, followed by the amplitude difference between the first harmonic and those near the first three formants. The phonation contrast between the two registers is mainly captured by PC1, and the difference across age groups also varies along PC1. The detailed results are included in the supplementary materials and only PC1 is reported. Fig. 5 shows the scores for PC1 of the high and low register tones by gender and age groups for unchecked and checked tones, respectively. For unchecked tones, all three age groups have similar scores in the high register tones for both male and female speakers, and the scores of the low register tones decrease from the older and middle-aged speakers to the younger speakers. The difference between the high and low register tones was smaller for the younger speakers than for the older and middle-aged speakers. Thus, the difference between the high and low register tones in phonation is decreasing gradually from the older and middle-aged to the younger speakers. The same is also true for checked tone in that the difference between T7 (high register) and T8 (low register) is also decreasing.

The data were also submitted to LDA, to compare the results with those for Shanghai Wu (Gao, 2016). Data used in LDA were also all the phonation measurements (both spectral tilt and noise measurements) averaged over the entire vowel. LDA was done separately for unchecked and checked tones, and also separately for each age group and each gender. Wilk's Lambda was calculated for each LDA model. The results are shown in Table 4. A greater value means a larger unexplained variance. Wilk's Lambda is greatest for younger speakers in both unchecked and checked tones, indicating that high and low register tones are produced less discriminately by the younger speakers than by the older and



a. Unchecked tones

b. Checked tones

Fig. 5. The scores of the first principal component (PC1) of unchecked (a) and checked (b) tones by gender and age groups.

Table 4
Wilk's Lambda of the LDA model for each age group and gender (higher value means greater variance unexplained).

tone	age group	Female	Male
unchecked	older	0.60	0.57
	middle-aged	0.67	0.55
	younger	0.75	0.73
checked	older	0.43	0.40
	middle-aged	0.48	0.36
	younger	0.57	0.56

middle-aged speakers. Wilk's Lambda is also increasing gradually for female speakers from the older and middle-aged to the younger speakers. However, Wilk's Lambda of middle-aged male speakers is comparable to that of older male speakers for unchecked tones and even smaller for checked tones. This suggests that the difference between breathy voice and modal voice is decreasing gradually for female speakers, but is retained well by middle-aged male speakers and is only decreasing for younger male speakers. Compared with the results of Shanghai Wu (Gao, 2016), the values for younger speakers of Suzhou Wu are smaller for both unchecked and checked tones. For example, the LDA model of unchecked tones in Shanghai Wu has Wilk's Lambda of 0.88 for female and 0.74 for male younger speakers. That is to say, the acoustic measurements are less efficient for the younger speakers in Shanghai Wu than Suzhou Wu in discriminating the low register tones from the high register tones. To sum up, both PCA and LDA confirmed that the breathy voice of the low register tones in Suzhou Wu is decreasing for the younger speakers.

3.2.2. The time course of the breathy voice across age groups

As the averaged values of the acoustic measurements were used in PCA and LDA, much information may be hidden in the time course of the measurements. The case is even subtler for T2 as it is creaky voiced in the middle of the vowel for younger speakers. Therefore, GAMMs were carried out on the acoustic and EGG measurements to investigate the differences in the time course of breathy voice across age groups. The acoustic and EGG measurements were normalized within speaker. GAMMs use splines to fit non-linear patterns in time series data and can deal with autocorrelated errors (Baayen et al., 2017, 2018; Wood, 2017). Furthermore, the itsadug package provides the plot_diff function, which plots the difference curve between levels of factors (van Rij et al., 2020). This would be very informative for the comparison of the time course of breathy voice in the low register tones with reference to the high register tones across age groups in this study. The estimated difference curves were reported in the text and the estimated smooth curves were included in the supplementary materials. Because some of the acoustic measurements were highly correlated, only the results for H1*–H2*, H1*-A3* and HNR15 are reported here. H1*-A3* was highly correlated with H1*-A1* (0.54) and H1*-A2* (0.52), and HNR15 was highly correlated with CPP (0.52), HNR05 (0.72), HNR25 (0.96) and HNR35 (0.92). H1*–H2* is also reported not only because it is one of the most commonly used spectral tilt measurements, but also because it can illustrate the creaky voice in younger female speakers' T2. Furthermore, as reported in the EMG study by Iwata (1995), breathy voice in Suzhou Wu is differentiated from modal voice by the inhibition of the activity of CT and VOC, and thus the slackness of the vocal folds. It is possible that the main articulatory mechanism underlying the breathy voice in Suzhou Wu is not the larger vocal fold opening, but the slackness of the vocal folds. Therefore, we also chose these measurements because they capture different aspects of phonation and we would like to show a fuller picture of the change in breathy voice.

3.2.2.1. H1–H2*.* The estimated difference curves of H1*–H2* are plotted in Fig. 6. The values of T1 and T7 are used as the references since T1 has stable modal voice and T7 is the only high register checked tone. The top panel shows the estimated difference curves of the unchecked tones, where T1-T2 is represented by a solid line and T1-T6 by a dashed line. The estimated difference curves of the checked tones are shown in the bottom panel, showing the difference between T7 and T8. The ribbons represent the 95% confidence interval of the estimated difference. The zero line is marked by a horizontal dashed line. If the ribbon includes the zero line, it indicates that there is no significant difference between the tone pair[6]. If the ribbon is below the zero line, it means that the high register tones have a lower H1*–H2* than the low register tone, and vice versa if the ribbon is above the zero line. Therefore, the ribbons would be below the zero line if the low register tones are breathier than T1 or T7.

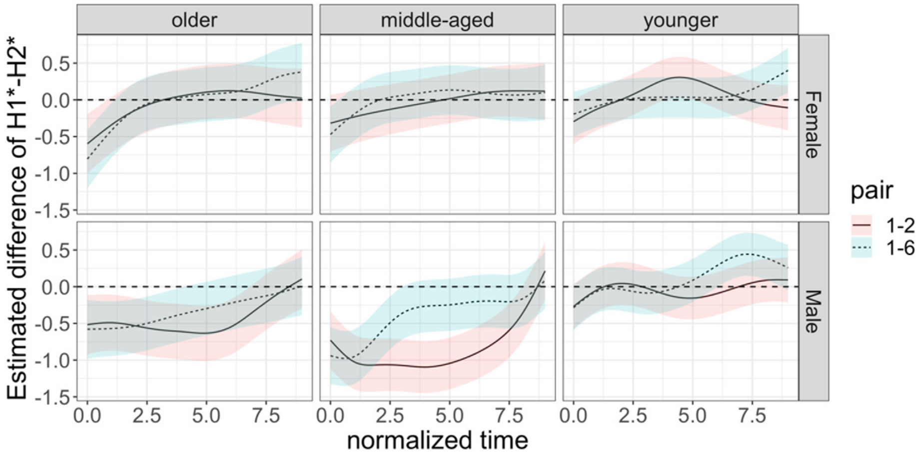
It is clear from Fig. 6 that there are differences across age groups in the differences in H1*–H2* between the high and low register tones. For T1-T2 at the top panel, most of the difference curve includes the zero line for females, and only a small portion at the start of the vowel is below the zero line. The same is also true for T1-T6. For males, the difference curves have a large part below the zero line for older and middle-aged speakers for both T1-T2 and T1-T6 pairs, but the ribbons of both T1-T2 and T1-T6 overlap with the zero line throughout for younger males, and the ribbon is even above the zero line at the end for T1-T6. These patterns indicate that, unlike for the older and middle-aged speakers, H1*–H2* of T2 and T6 is no longer higher than that of T1, suggesting smaller differences in glottal opening between T2, T6 and T1 for the younger speakers than for the older and middle-aged speakers. At the end of T6 and at the middle of T2 for females, H1*–H2* is even lower than T1, meaning that T6 is creakier at the end and T2 is creakier at the middle of the vowel than T1.

The difference curves of the checked tones are shown in the bottom panel of Fig. 6. Although a considerable part of the difference curve between T7 and T8 is below the zero line for the older and middle-aged speakers, it is mostly above the zero line for younger females. For younger male speakers, however, the difference curve is still below the zero line, indicating higher H1*–H2* for T8 than T7. Thus, there is a gender difference for the checked tones. Although the younger female speakers' T8 is not breathier than T7, the breathy voice is preserved to some extent by the male speakers.

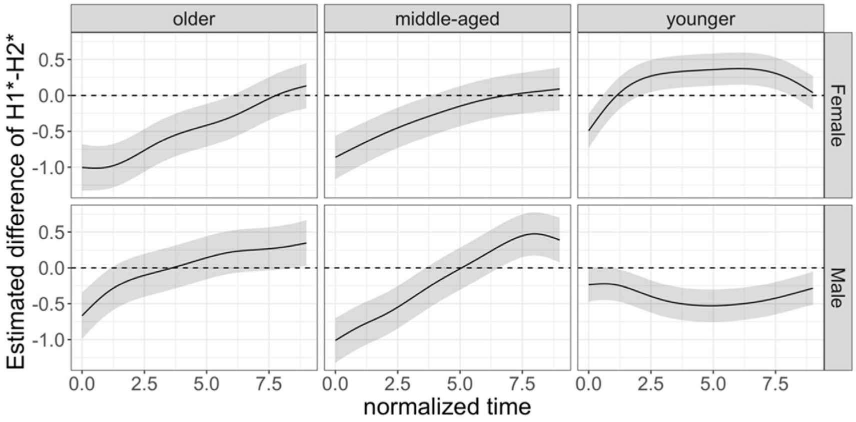
Taken together, H1*–H2* of the low register tones of the younger speakers is no longer higher than that of the high register tones, except for T8 of younger males. The low register

---

[6] Note that the difference as represented by the difference curve is in a point-by-point manner (Sóskuthy, 2021).



a. Unchecked tones



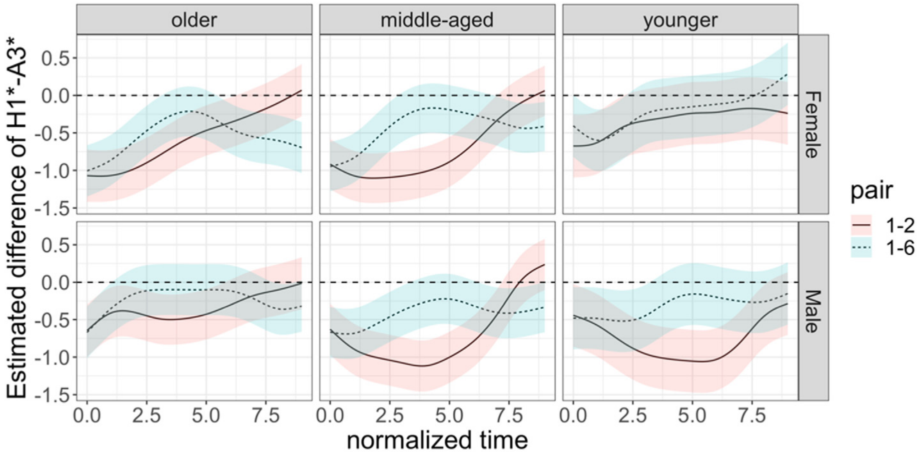
b. Checked tones

Fig. 6. Estimated difference curve of H1*–H2* for T1-T2 (solid line), T1-T6 (dashed line) pairs (a) and T7-T8 pair (b) across genders and age groups.

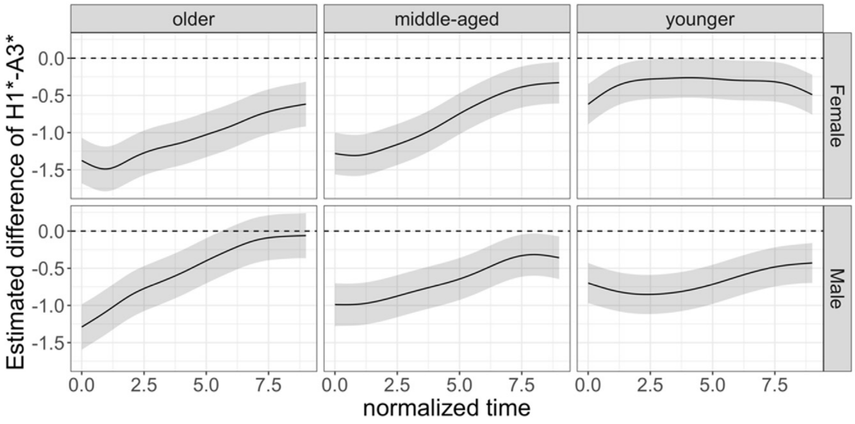
tones are no longer breathy voiced in terms of H1*–H2* and they are even creaky voiced in the middle of T2 and the end of T6. Female speakers are more advanced in the change in H1*–H2* of the low register tones for both unchecked and checked tones. Younger male speakers still maintain higher H1*–H2* of T8 than T7.

3.2.2.2. H1-A3*.* While H1*–H2* is usually thought to correlate with how open the glottis is (Kreiman et al., 2008; Zhang, 2016b), H1*-A3* correlates with the gradualness of the glottal closure, i.e., how fast the vocal folds adduct (Hanson et al., 2001; Holmberg et al., 1995). It was suggested by DiCanio (2009) that the opening of the glottis and the gradualness of glottal closure can be controlled independently. It is possible that the difference is maintained in H1*-A3* while it is no longer for H1*–H2*. The GAMM-predicted difference curves of H1*-A3* for the unchecked and checked tones are shown in the top and bottom panels of Fig. 7. On the top panel, both the difference curves of the T1-T2 and T1-T6 pairs overlap with the zero line for the younger female speakers, while they are mostly below the zero line for the older and middle-aged speakers. The difference curves are below the zero line for only a brief portion at the onset of the vowel. Therefore, T2 and T6 are also less breathy for younger female speakers than older and middle-aged female speakers. H1*-A3* of the T1-T6 pair for the younger male speakers also has a large part of overlap with the zero line, but that of T1-T2 is mainly below the zero line. T6 of the younger male speakers is less breathy than that of the older and middle-aged males, but T2 is similarly breathy voiced, as measured by H1*-A3*.

The difference between T7 and T8 in H1*-A3* is also smaller for the younger speakers than for the older and middle-aged speakers, but H1*-A3* of T8 is still higher than T7 for the younger speakers. As shown by the difference curves in Fig. 7 (b), although the curves are still below the zero line, the difference between T7 and T8 is smaller for the younger speakers. For the female speakers, the ribbon is very close to the zero line. For H1*-A3*, even if both female and male younger speakers have higher values for T8, the difference is decreasing compared to the older and middle-aged speakers.



a. *Unchecked tones*



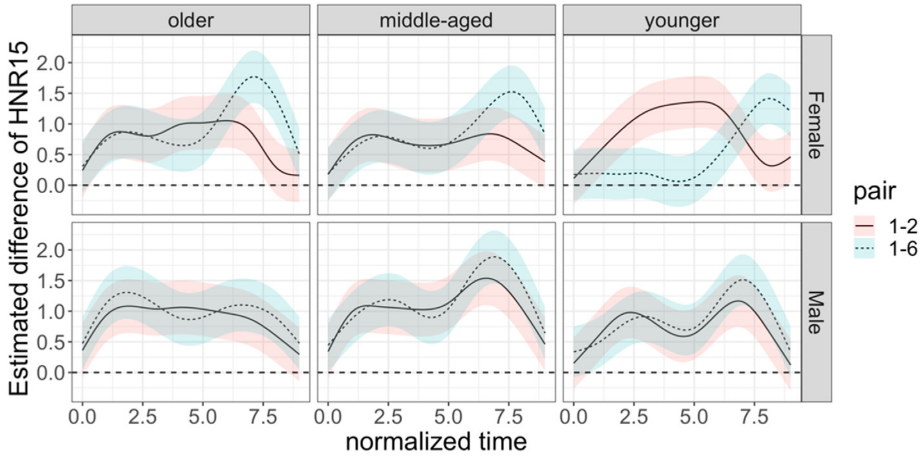
b. *Checked tones*

Fig. 7. Estimated difference curve of H1*-A3* for T1-T2 (solid line) and T1-T6 (dashed line) pairs (a) T7-T8 pair (b) across genders and age groups.

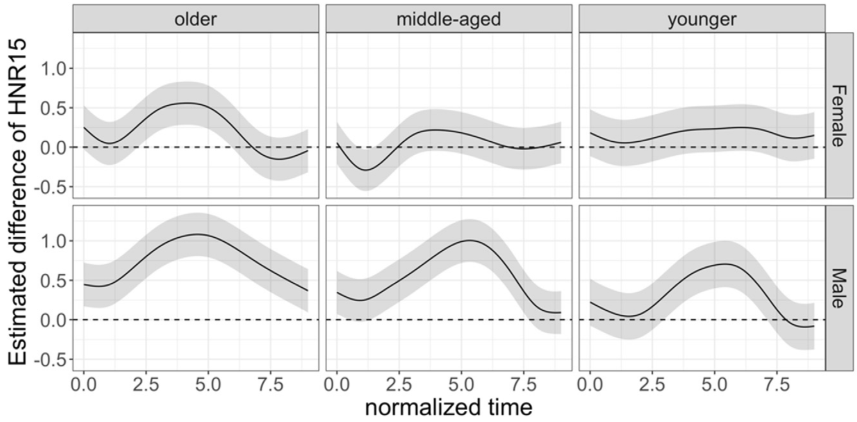
Compared to H1*–H2*, H1*-A3* is better maintained by the younger speakers, although the difference between the high and low register tones is also decreasing. Similar to H1*–H2*, the difference in H1*-A3* is better maintained by the younger male speakers than by the younger female speakers.

3.2.2.3. HNR15. HNR15 is the Harmonic-to-Noise Ratio in the 0–1500 Hz band. Breathy voice has lower HNR15 values than modal voice, indicating a stronger noise component. Therefore, if the difference curve is below the zero line, it means the low register tones are noisier than the high register tones, and vice versa. The difference curves of the unchecked and checked tones estimated by GAMM are plotted in Fig. 8. Unlike the spectral measurements, the largest difference between the high and low register tones appears in the middle of the vowel. It is perhaps due to the effect of the initial consonant at the beginning and the boundary effect at the end. Lower HNR at both ends was also reported for White Hmong (Garellek, 2012). In the same manner as spectral measurements, HNR15 of T2 and T6 was also compared to that of T1, the tone with the most stable modal voice. But the difference curves show dramatically different patterns from the spectral tilt measurements. Only in the first half of the T1-T6 pair for the younger females does the ribbon include the zero line. The younger male speakers show only slightly lower curves than the older and middle-aged speakers. The difference between the younger females' T1 and T2 is even larger than that of the older and middle-aged females, especially in the middle of the vowel. This is not unexpected, though. As has been shown above, the younger females' T2 is creaky voiced in the middle of the vowel and creaky voice also has lower HNR than modal voice, as reported in Mazatec (Garellek & Keating, 2011), and Hmong (Garellek, 2012).

The largest difference in HNR15 between T7 and T8 is also in the middle of the vowel. Nonetheless, the younger females show overlap of the ribbons with the zero line, whilst the middle-aged females have only a brief portion that does not include the zero line, as shown in Fig. 8. The younger male speakers, on the other hand, also show a smaller difference than the older and middle-aged males. Moreover, at both ends of the vowel, there is no significant difference between T7 and T8 for the younger males.



a. *Unchecked tones*



b. *Checked tones*

Fig. 8. Estimated difference curve of HNR15 for T1-T2 (solid line) and T1-T6 (dashed line) pairs (a) T7-T8 pair (b) across genders and age groups.

The noise measurements show some different, yet consistent, patterns as compared to the spectral tilt measurements. The younger females are ahead of the males in the change in breathy voice in the low register tones, although they have more noise in the middle of T2, probably due to creaky voice. The younger males still maintain the difference in HNR15 in T1-T2, T1-T6 and T7-T8, yet to a smaller extent than the older and middle-aged male speakers.

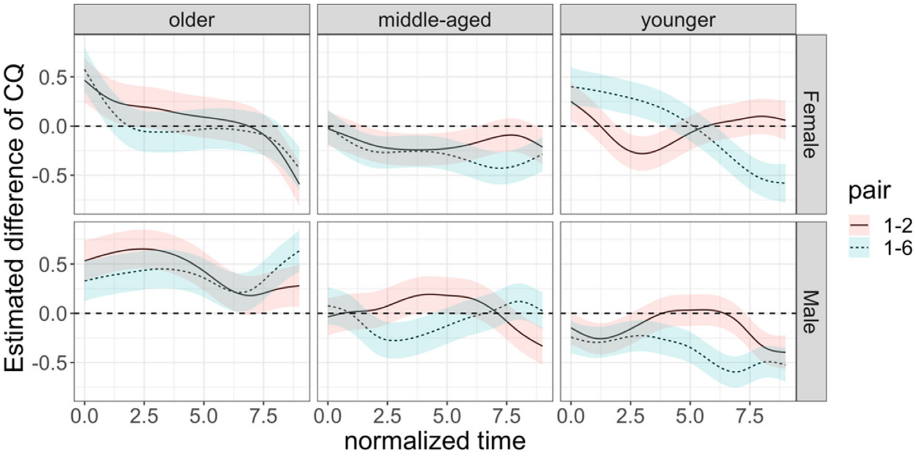
3.2.2.4. CQ. CQ was measured from the EGG signals, and due to the noisy signals, the data of several speakers (mostly female) were removed (see Section 2.4). Breathy voice has a shorter closed phase and thus a smaller CQ. The GAMM-estimated difference curves are shown in Fig. 9. For the unchecked tones, among all three age groups, only the older speakers show a considerable part of the curve above the zero line, i.e., a smaller CQ for the low register tones than T1, except for the T1-T6 of younger females. In all other cases, T2 and T6 have a CQ similar to or greater than T1. The dipping in T2 of the younger females is reminiscent of the pattern in H1*–H2*, indicating the presence of creaky voice. However, the first half of T6 for younger females is smaller than T1, which contradicts the pattern in all acoustic measurements.

Clear patterns can be seen for the checked tones, and they are highly consistent with H1*–H2*. The middle-aged and younger female speakers show no significant difference in CQ between T7 and T8, with only a lower CQ of T8 than that of T7 at the end of the vowel for the younger females. Due to the glottal stop coda, it indicates less creaky voice at the end of T8. The younger males also show only a slight difference in the latter half of the vowel.

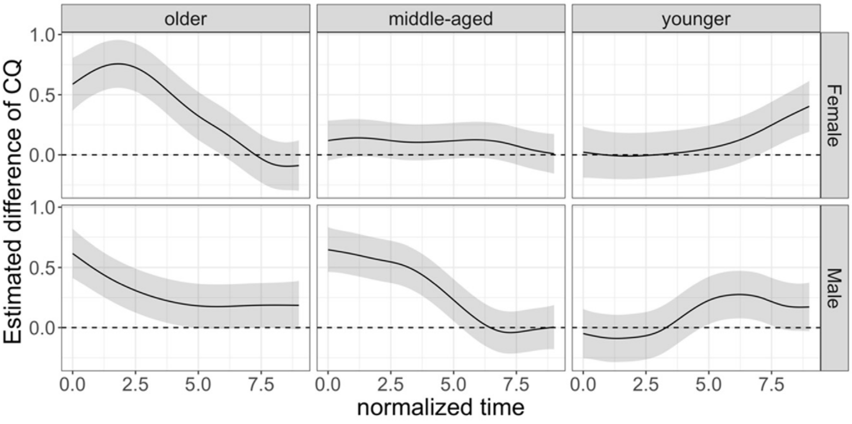
CQ shows generally similar patterns with H1*–H2*, except for the younger females' T6. For T2 and T8, the differences in CQ from T1 or T7 are becoming smaller for both the younger females and males. But for younger females T6 is different in CQ from the acoustic measurements.

4. Discussion

This study aimed to investigate the difference in breathy voice of the low register tones in Suzhou Wu across age groups and whether breathy voice is used less by younger speakers than by older and middle-aged speakers. Acoustic



a. Unchecked tones



b. Checked tones

Fig. 9. Estimated difference curve of CQ for the T1-T2 (solid line) and T1-T6 (dashed line) pairs (a) T7-T8 pair (b) across genders and age groups.

and articulatory data from three age groups were analyzed. PCA and LDA were used to compare the differences in breathy voice across age groups and GAMM was used to explore the differences in time course. In the following discussion in Section 4.1, we first summarize the main findings of this study and answer the research questions raised. In Section 4.2, based on the results, we will discuss the decrease in breathy voice in Suzhou Wu and speculate about the possible reasons for this change. In Section 4.3, the tone changes of T2 and T8 in Suzhou Wu will be discussed. The implications of this study for our understanding of the diachronic change in phonation type are elaborated in Section 4.4.

4.1. The change in breathy voice in Suzhou Wu

This study investigated the change in breathy voice in Suzhou Wu and asked three research questions. First of all, as reviewed in the Introduction (Section 1.2), the change in breathy voice is related to the initial consonants and tone pitch contours. Comparisons of VOT and the duration of fricatives across age groups indicate that the younger speakers have similar initial stops and fricatives to those of the older and middle-aged speakers. That is to say, initial consonants are not becoming aspirated. However, two of the low register tones, T2 and T8, are undergoing tone change. Their pitch contours are changing such that the point at which rising starts is shifting towards the middle of the vowel. Most remarkably, younger female speakers pronounce T2 with creaky voice in the middle of the vowel. Secondly, this study used PCA and LDA to compare the differences in breathy voice across age groups. The scores of PC1 and Wilk's Lambda both show that breathy voice is used less by the younger speakers than by the older and middle-aged speakers. Compared with Shanghai Wu, the acoustic measurements in Suzhou Wu are more successful at discriminating the low register tones from the high register tones. Methodological decisions may be the main reason for this difference. This study included the data within 2.5 standard deviations, while Gao (2016), in a study of Shanghai Wu, only included the data within two standard deviations. With less variance in the data, the results of Shanghai Wu would be more extreme.

Furthermore, fine-grained patterns of the changes in breathy voice are revealed using GAMM on the acoustic and EGG measurements. The breathy voice of Suzhou Wu is gradually decreasing for all three low register tones. T2 is changing from rising to falling-rising and has become creaky voiced for younger female speakers, who have almost entirely lost the breathy voice. T8 is also undergoing a similar tone change to

T2 but without creaky voice. Breathy voice in T2 and T8 is decreasing. On the other hand, T6 is not influenced by tone change and its pitch contour is not changing across age groups, but its breathy voice is also decreasing.

In terms of time course, the older and middle-aged speakers show the largest difference in spectral tilt measurements and CQ at the onset of the vowel between the high and low register tones, which disappears towards the end of the vowel. This is consistent with previous studies that found that breathy voice in Wu Chinese is most obvious at the beginning of the vowel and gradually decreases towards the end of the vowel. But for younger speakers, even the large difference at the onset of the vowel is disappearing. One of the reviewers raised the possibility that the difference in phonation may be simply due to the fact that the measurements covary with f0 (Kuang, 2017). However, if that is the case, the covariation would be uniform across all age groups. Given the similarity of the pitch contour across age groups as in Fig. 3 (except for T2 and T8), we would expect that there would also be a large difference at the onset of the vowel for the younger speakers. Take T6 as an example, which has no pitch contour change across age groups. The differences between T6 and T1 in H1*–H2* and H1*-A3* are also smaller for the younger than for the older and middle-aged speakers. Therefore, there is indeed distinction in phonation in Suzhou, which should not be attributed to covariation with f0 alone. Nevertheless, T2 and T8 exhibit a more remarkable decrease in breathy voice than T6, which suggests that the covariation with f0 does have some contribution to the phonation contrast.

To better understand the change in breathy voice, the checked tones were also included in this study. The glottal stop coda may have some influence on phonation, and close inspection of our data show that while most speakers still have the glottal stop coda in the checked tones, one of the younger female speakers has lost the glottal stop coda (F15), and another female had only a few tokens with a glottal stop (F17). The loss of the glottal stop coda is also reported in Deqing Wu by Zhu and Jiao (2011) and in Shanghai Wu by Gao and Kuang (2022). Inspection of the spectrograms also shows that when the glottal stop disappears, it takes with it the glottalization and creaky voice in the vowel. CQ shows some signs of this change. In Fig. 9, towards the end of the syllable in the checked tones, there is no difference between T7 and T8 for the older and middle-aged speakers, but the younger speakers' T7 has greater CQ than T8. That is to say, the loss of glottal stop is more frequent in T8 than in T7. The results show no difference in HNR15 and CQ between T7 and T8 at the onset of the vowel, and less difference in H1*–H2* and H1*-A3* for the younger than the older and middle-aged speakers. Therefore, it is likely that T8 is changing from a tone with complex phonation of both breathy and creaky voice to a more modal-voiced tone.

4.2. The decrease in breathy voice

The results of this study show that breathy voice in Suzhou Wu is decreasing, similar to that reported in Shanghai Wu (Gao, 2016; Zhang & Yan, 2018) and Lili Wu (Shi et al., 2020). Although the reason for this decrease still requires further investigation, the analyses in this study can shed some light on its cause.

It is possible that breathy voice is a secondary cue in tone split. As f0 becomes dominant, breathy voice becomes redundant and its importance diminishes. This is similar to the shifting of the primary cue of the register contrast in Southern Yi from phonation to formant (Kuang & Cui, 2018). In Suzhou Wu, breathy voice seems to have also been a secondary cue for a long time, since in Suzhou Wu tones of the two registers have quite distinct pitch contours and pitch alone can provide sufficient information for the discrimination of the tones. Therefore, it is possible that the tone system of Suzhou Wu is changing from one that is based on both pitch and phonation to one that is based on pitch alone.

Another factor that may be at play is the change in tone pitch. The results show that T2 and T8 are more advanced in the decrease in breathy voice than T6, which is not undergoing tone change. It is reminiscent of the observation that breathy voice disappears in some tones but remains intact in other tones in the Tamang-Gurung-Thakali-Manangke (TGTM) languages (Mazaudon, 2012). But it is still doubtful that tone change and the decrease in breathy voice may be simply occurring simultaneously, with no causal relationship between the two. Further study is needed to investigate whether there is a causal relationship between tone change and the change in phonation type.

4.3. Tone changes in T2 and T8 in Suzhou Wu

Apart from the decrease in breathy voice, this study has another interesting finding. There are ongoing tone changes in T2 and T8. The pitch contour of T2 is changing from rising to falling-rising, which is accompanied by creaky voice, especially for the younger female speakers. The rise in T8 is also shifting gradually towards the middle of the vowel. T2 and T8, as rising tones, show a similar trend of tone change, in that the time at which the rise starts is delayed for the younger speakers. The delay of the rise in a rising tone is not uncommon cross-linguistically. As a matter of fact, it is the most frequent tone change found by Yang and Xu (2019), reviewing data from Sinitic, Tai-Kadai, Hmong-Mien, and Tibeto-Burman languages. For example, Bangkok Thai has acoustic recordings starting at the dawn of experimental phonetics (the early 1900s) through until quite recently and a comparison of these recordings showed that Tone 5 has changed from mid rising to low rising, with rising only starting in the second half of the syllable (Pittayaporn, 2018; Zsiga, 2008). It can also be seen in the speech of King Rama IX of Thailand between 1959 and 1997 (Yang, Pittayaporn, Kirby, & Jitwiriyapont, 2021). The inflection point of Tone 5 has been gradually shifting from the beginning to near the middle, and its shape was more typically rising during recordings from 1959 to 1969, but was falling-rising during recordings from 1991 to 1997. In Chinese dialects, the change of low rising tone to falling-rising was also reported for Taiwan Mandarin (Tone 2, Sanders, 2008), Taiwanese Min (Tone 2, Chiung, 2003), Shanghai and Wuxi Wu (Zhang, 2014), and the Chongqing dialect (Tone 4, Ming & Zhang, 2015).

In addition to the change in pitch contour, T2 is also becoming creaky voiced for the younger female speakers. One of the reasons is that creaky voice is related to low pitch in Suzhou Wu. As a matter of fact, creaky voice is also found with low pitch targets in other tones. For example, creaky voice also occurs at the end of T3, T6 and the second half of T5. The same is also reported for Mandarin, where creaky voice is also found in Tone 2 and Tone 4, in addition to Tone 3 (Kuang, 2017). It is possible that creaky voice in Suzhou Wu is driven by the low pitch target. In this respect, creaky voice in Suzhou Wu is similar to creaky voice in Tone 3 of Mandarin and Tone 4 of Cantonese (Yu & Lam, 2014) in that it is pitch-dependent (Kuang, 2017).

However, creaky voice does not always appear in the younger male speakers' T2, although they share the same pitch contour as the younger females. Other factors must be involved, apart from the correlation between creaky voice and a low pitch. A possible source of this discrepancy may be the contact with Mandarin Chinese. With the change in pitch contour, T2 in Suzhou Wu becomes similar to the low-dipping Tone 3 [214] in Mandarin Chinese. Now that the younger speakers are bilingual in Suzhou Wu and Mandarin Chinese, they may have employed the same articulatory strategy as Mandarin Tone 3 to realize a similar pitch contour, which is usually creaky in the middle of the vowel. It has been demonstrated by Yao and Chang (2016) that Mandarin Chinese as L2 can reverse the ongoing vowel merger in Shanghai Wu as L1. A perceptual link between L1 and L2 can be formed when the two pronunciations in L1 and L2 are phonetically similar, which would give rise to the convergence of these pronunciations. This also seems to be the case for T2 in Suzhou Wu, which has a very similar pitch contour to Tone 3 in Mandarin. Furthermore, the effect of gender is also expected in this scenario. It has long been noted that young women take the lead in sound change (Labov, 2001). Similarly, young women are also found to be ahead of young men in the loss of breathy voice in Shanghai Wu (Gao, 2016) and the tone change in Shanghai Wu (Zhang, 2014). Therefore, creaky voice of T2 in Suzhou Wu may appear due to its correlation with low pitch as well as the contact with Mandarin.

4.4. The diachronic change in phonation

In addition to contributing to our understanding of how breathy voice changes after tone split, this study also has important implications for our understanding of diachronic change in phonation. This study demonstrates differences between breathy voice and creaky voice in how they can appear and disappear in languages. Both breathy voice and creaky voice are related to low pitch, and it is proposed that they are alternative means of achieving a low pitch target (Kuang, 2017). However, the conditions under which breathy voice and creaky voice can appear are different. It is proposed by Kuang (2013, 2017) that non-modal phonation can be divided into pitch-dependent and pitch-independent phonations, which are the result of different physiological mechanisms. Extremely low pitch is sufficient to induce creaky voice, but less able to induce breathy voice. For example, Tone 3 in Mandarin Chinese is a low tone and usually creaky voiced. Although Zheng (2006) observed breathy voice with Tone 3, Kuang (2017) did not find any tokens with breathy voice. In contrast, breathy voice is usually induced by aspiration or a strong airflow. Voiced stops have slack vocal folds and a low laryngeal position, which can induce breathy voice. The strong airflow required by trills can also lead to breathy voice, as in Phnom Penh Khmer (Kirby, 2014). On the other hand, there is another difference between breathy voice and creaky voice in relation to their loss. As shown in this study and the previous studies reviewed in the Introduction, breathy voice can be lost even when pitch remains low. However, this seems not to be the case for creaky voice. As long as the pitch reaches the bottom of the pitch range, creaky voice consistently occurs (Kuang, 2013, 2017). These two aspects together show that compared with breathy voice, creaky voice is more closely related to pitch. There is also evidence in perception that breathy voice is less correlated with a low pitch than creaky voice. For example, in White Hmong, the listeners mainly employed pitch to identify the creaky voiced tone, while breathy voice was essential for the identification of the breathy-voiced tone (Garellek, Keating, Esposito, & Kreiman, 2013). It is possible that the characteristics of breathy voice and creaky voice in production and perception are key factors influencing their diachronic change.

Much effort has been made to understand the physical properties of phonation, its relationship with pitch, and its role in linguistic contrasts. However, to get a complete understanding of phonation, a further step ought to be taken towards understanding how it can change in languages and this study is a valuable attempt towards this direction.

5. Conclusion

This study investigated the change in breathy voice after tone split in Suzhou Wu with apparent-time data from three age groups. PCA and LDA conducted on acoustic measurements indicated that breathy voice is decreasing in all the low register tones among the younger speakers. More detail was revealed by GAMM. Different measurements show different degrees of change. H1*–H2* is the most susceptible to change, followed by H1*-A3* and HNR15. It is likely that breathy voice will be eventually lost after tone split. We propose that as the tone system of Suzhou Wu is changing to be more pitch-based, breathy voice as a secondary cue will diminish. It is also noted that T2 and T8 are undergoing tone change. The change in T2 possibly results from a low pitch as well as from language contact with Mandarin. This study presents a valuable endeavor in understanding changes in breathy voice and contributes to our understanding of diachronic change in phonation type. The finding of the decrease in breathy voice raises a further question. Now that the younger speakers use breathy voice less as the older speakers, does breathy voice still have a role to play in their lexical tone perception? Further investigation of the perception of breathy voice is underway.

CRediT authorship contribution statement

Chunyu Ge: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Wenwei Xu:** Data curation, Investigation, Writing – review & editing. **Wentao Gu:** Investigation, Writing – review & editing. **Peggy Pik Ki Mok:** Conceptualization, Investigation, Methodology, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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. Partial support was received from the Major Program of the National Social Science Fund of China (13&ZD189) awarded to the their author. We would like to thank Feifan Wang for his assistance in the early stage of the project. We are also grateful to the three anonymous reviewers for their valuable comments and suggestions.

Appendix A. . Test words

See Table A1.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2023.101239.

Table A1
Test words used in this study.

Tone	IPA	character	gloss	IPA	character	gloss	IPA	character	gloss
T1	pɑ	巴	‘bar’	-	-	-	kɑ	家	‘home’
T2	bɑ	牌	‘card’	-	-	-	gɑ	茄	‘eggplant’
T3	pɑ	摆	‘put’	-	-	-	kɑ	假	‘fake’
T5	pɑ	拜	‘bow’	tɑ	带	‘lace’	kɑ	嫁	‘marry’
T6	bɑ	败	‘fail’	dɑ	汏	‘wash’	gɑ	解	‘unfasten’
T7	paʔ	百	‘hundred’	taʔ	搭	‘take’	kaʔ	夹	‘clip’
T8	baʔ	拔	‘pull’	daʔ	达	‘arrive’	gaʔ	轧	‘squeeze’
T1	-	-	-	sɑ	筛	‘sieve’	ɑ	挨	‘near’
T2	-	-	-	zɑ	柴	‘firewood’	ɦɑ	鞋	‘shoes’
T3	-	-	-	sɑ	洒	‘sprinkle’	ɑ	矮	‘short’
T5	-	-	-	sɑ	啥	‘what’	-	-	-
T6	-	-	-	zɑ	惹	‘provoke’	-	-	-
T7	faʔ	发	‘distribute’	saʔ	杀	‘kill’	aʔ	鸭	‘duck’
T8	vaʔ	罚	‘fine’	zaʔ	闸	‘brake’	ɦaʔ	盒	‘box’
T1	pɪ	边	‘class’	tɪ	颠	‘top’	-	-	-
T2	bɪ	便	‘cheap’	dɪ	田	‘farm’	-	-	-
T3	pɪ	扁	‘flat’	tɪ	点	‘point’	-	-	-
T5	pɪ	变	‘change’	tɪ	店	‘shop’	-	-	-
T6	bɪ	便	‘convenient’	dɪ	电	‘electricity’	-	-	-
T7	pɪʔ	笔	‘pen’	tɪʔ	跌	‘fall’	-	-	-
T8	bɪʔ	别	‘separate’	dɪʔ	敌	‘enemy’	-	-	-
T1	fɪ	飞	‘fly’	sɪ	鲜	‘fresh’	ɪ	烟	‘sorrow’
T2	vɪ	肥	‘fat’	zɪ	齐	‘uniform’	ɦɪ	盐	‘salt’
T3	-	-	-	sɪ	选	‘select’	ɪ	演	‘perform’
T5	fɪ	肺	‘lung’	sɪ	线	‘string’	ɪ	咽	‘pharynx’
T6	vɪ	未	‘not yet’	zɪ	序	‘order’	ɦɪ	盐	‘pickle’
T7	-	-	-	sɪʔ	雪	‘snow’	ɪʔ	一	‘one’
T8	-	-	-	zɪʔ	席	‘mat’	ɦɪʔ	叶	‘leaf’
T1	pɛ	班	‘class’	tɛ	堆	‘heap’	kɛ	该	‘should’
T2	bɛ	赔	‘lose’	dɛ	台	‘table’	-	-	-
T3	pɛ	板	‘board’	tɛ	胆	‘gallbladder’	kɛ	改	‘change’
T5	pɛ	扮	‘disguise’	tɛ	对	‘correct’	kɛ	盖	‘cover’
T6	bɛ	办	‘deal’	dɛ	蛋	‘egg’	gɛ	隑	‘lean’
T7	pəʔ	拨	‘stir’	təʔ	德	‘virtue’	kəʔ	鸽	‘pigeon’
T8	bəʔ	荸	‘’	dəʔ	掉	‘lost’	gəʔ	搿	‘that’
T1	fɛ	翻	‘overturn’	sɛ	三	‘three’	ɛ	哀	‘sorrow’
T2	vɛ	烦	‘annoying’	zɛ	馋	‘gluttonous’	ɦɛ	咸	‘salty’
T3	fɛ	反	‘inverted’	-	-	-	-	-	-
T5	fɛ	贩	‘peddle’	sɛ	散	‘scattered’	ɛ	晏	‘late’
T6	vɛ	饭	‘meal’	zɛ	赚	‘earn’	ɦɛ	害	‘harm’
T7	fəʔ	弗	‘not’	səʔ	色	‘color’	əʔ	遏	‘prevent’
T8	vəʔ	佛	‘buddha’	zəʔ	石	‘stone’	ɦəʔ	合	‘together’

References

Abramson, A. S., & Erickson, D. M. (1992). Tone Splits and Voicing Shifts in Thai: Phonetic Plausibility. *Status Report on Speech Research, Haskins Laboratories, 53* (2), 255–262.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. (2018). Autocorrelated Errors in Experimental Data in the Language Sciences: Some Solutions Offered by Generalized Additive Mixed Models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed-Effects Regression Models in Linguistics* (pp. 49–69). Springer International Publishing. https://doi.org/10.1007/978-3-319-69830-4_4.

Baayen, H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language, 94*, 206–234. https://doi.org/10.1016/j.jml.2016.11.006.

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics, 66*, 120–144. https://doi.org/10.1016/j.wocn.2017.09.005.

Bei, X. (2011). The between-group statistical difference of the high and low register tones in Suzhou Chinese (苏州话阴阳调类的分组统计差异). *Journal of Huzhou Vocational and Technological College (湖州职业技术学院学报), 9*(02), 44–48.

Boersma, P. (2020). *Praat: Doing phonetics with computers*.

Brunelle, M., Brown, J., & Phạm, T. T. H. (2022). Northern Raglai voicing and its relation to Southern Raglai register: Evidence for early stages of registrogenesis. *Phonetica*. https://doi.org/10.1515/phon-2022-2019.

Brunelle, M., & Tạ, T. T. (2021). Register in languages of Mainland Southeast Asia: The state of the art. In *The Languages and Linguistics of Mainland Southeast Asia* (pp. 683–706). De Gruyter Mouton. https://doi.org/10.1515/9783110558142-029.

Brunelle, M., Tạ, T. T., Kirby, J., & Giang, ã. L. (2020). Transphonologization of voicing in Chru: Studies in production and perception. *Laboratory Phonology: Journal of the Association for Laboratory*. *Phonology, 11(1), Article 1*. https://doi.org/10.5334/labphon.278.

Cao, J., & Maddieson, I. (1992). An Exploration of Phonation Types in Wu Dialects of Chinese. *Journal of Phonetics, 20*, 77–92.

Cao, J. (1987). The ancient initial voiced consonants in Modern Wu dialects. *Proceedings of 11th ICPhS*.

Chao, Y. R. (1930a). A system of Tone-Letters. *Le Maître Phonétique, 30*, 24–27.

Chao, Y. R. (1930b). Transcribing reversed English. *Bulletin of Institute of History and Philology, 2*(2), 205–223.

Chao, Y. R. (1967). Contrastive Aspects of the Wu Dialects. *Language, 43*(1), 92–101. https://doi.org/10.2307/411386.

Chao, Y. R. (1928). *Studies of Modern Wu Dialects (现代吴语的研究)*. The Tsinghua Academy of Chinese Learning (清华学校研究院).

Chen, Z. (2010). An acoustic study of voiceless onset followed by breathiness of Wu dialects: Based on the Shanghai dialect (吴语清音浊流的声学特征及鉴定标志以上海话为例). *Studies in Language and Linguistics (语言研究), 30*(03), 20–34.

Chen, Y. (2011). How does phonology guide phonetics in segmentF0 interaction? *Journal of Phonetics, 39*(4), 612–625. https://doi.org/10.1016/j.wocn.2011.04.001.

Chen, Z. (2015). 气嗓音与低调 / Breathy voice and low tone. *Journal of Chinese Linguistics, 43*(1A), 90–118.

Chiung, W.-V.-T. (2003). Tone change in Taiwanese: Age and geographic factors. *University of Pennsylvania Working Papers. Linguistics, 8*(1), 43–55.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics, 30*(2), 193–228. https://doi.org/10.1006/jpho.2001.0153.

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics, 66*, 185–216. https://doi.org/10.1016/j.wocn.2017.09.009.

de Krom, G. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech and Hearing Research, 36*(2), 254–266. https://doi.org/10.1044/jshr.3602.254.

DiCanio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association, 39*(2), 162–188. https://doi.org/10.1017/S0025100309003879.

Diffloth, G. (1982). *Registres, dévoisement, timbres vocaliques: Leur histoire en katouique* (pp. 47–82). XI: Mon-Khmer Studies.

Ferlus, M. (2009). What were the four divisions of Middle Chinese? *Diachronica, 26*(2), 184–213. https://doi.org/10.1075/dia.26.2.02fer.

Gao, J. (2016). Sociolinguistic motivations in sound change: On-going loss of low tone breathy voice in Shanghai Chinese. *Papers in Historical Phonology, 1*, 166–186. https://doi.org/10.2218/pihph.1.2016.1698.

Gao, J., & Hallé, P. (2017). Phonetic and phonological properties of tones in Shanghai Chinese. *Cahiers de Linguistique Asie Orientale, 46*(1), 1–31. https://doi.org/10.1163/19606028-04601001.

Gao, J., Hallé, P., Honda, K., Maeda, S., & Toda, M. (2011). Shanghai Slack Voice: Acoustic and EPGG Data. *Proceedings of ICPhS XVII*, 719–722.

Gao, X., & Kuang, J. (2022). Phonation Variation as a Function of Checked Syllables and Prosodic Boundaries. *Languages, 7*(3), 171. https://doi.org/10.3390/languages7030171.

Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics, 40*(1), 152–161. https://doi.org/10.1016/j.wocn.2011.10.003.

Garellek, M., Chai, Y., Huang, Y., & Doren, M. V. (2021). Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association, 1–28*. https://doi.org/10.1017/S0025100321000116.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association, 41* (2), 185–205. https://doi.org/10.1017/S0025100311000193.

Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America, 133* (2), 1078–1089. https://doi.org/10.1121/1.4773259.

Garellek, M., Ritchart, A., & Kuang, J. (2016). Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics, 59*, 110–121. https://doi.org/10.1016/j.wocn.2016.09.001.

Halle, M., & Stevens, K. N. (1971). A Note on Laryngeal Features. *MITResearch Laboratory of Electronics Quarterly Progress Report, 198–213*. https://doi.org/10.1515/9783110871258.45.

Hanson, H. M., Stevens, K. N., Kuo, H.-K.-J., Chen, M. Y., & Slifka, J. (2001). Towards models of phonation. *Journal of Phonetics, 29*(4), 451–480. https://doi.org/10.1006/jpho.2001.0146.

Hashimoto, M. (1973). *The Hakka Dialect: A linguistic study of its phonology, syntax and lexicon*. Cambridge University Press.

Hashimoto, M. (1992). Hakka in Wellentheorie perspective / 客家方言的波浪性发展. *Journal of Chinese Linguistics, 20*(1), 1–49.

Haudricourt, A.-G. (1954). De l'origine des tons en vietnamien. *Journal Asiatique, 242*, 69–82.

Haudricourt, A.-G. (1965). Les mutations consonantiques des occlusives initiales en môn-khmer. *Bulletin de La Société de Linguistique de Paris, 60*(1), 160–172.

Haudricourt, A.-G. (1972). Two-way and three-way splitting of tonal systems in some Far-Eastern languages. In J. G. Harris & R. B. Noss (Eds.), & C. Court (Trans.), *A Conference on Tai Phonetics and Phonology* (pp. 58–86). Mahidol University.

Henderson, E. J. A. (1952). The Main Features of Cambodian Pronunciation. *Bulletin of the School of Oriental and African Studies, University of London, 14* (1), 149–174.

Henrich, N., d'Alessandro, C., Doval, B., & Castellengo, M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *The Journal of the Acoustical Society of America, 115*(3), 1321–1332. https://doi.org/10.1121/1.1646401.

Herbst, C., & Ternström, S. (2006). A comparison of different methods to measure the EGG contact quotient. *Logopedics Phoniatrics Vocology, 31*(3), 126–138. https://doi.org/10.1080/14015430500376580.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic Correlates of Breathy Vocal Quality. *Journal of Speech, Language, and Hearing Research, 37*(4), 769–778. https://doi.org/10.1044/jshr.3704.769.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P. C., & Goldman, S. L. (1995). Comparisons Among Aerodynamic, Electroglottographic, and Acoustic Spectral Measures of Female Voice. *Journal of Speech, Language, and Hearing Research, 38*(6), 1212–1223. https://doi.org/10.1044/jshr.3806.1212.

Howard, D. M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice: Official Journal of the Voice Foundation, 9*(2), 163–172. https://doi.org/10.1016/s0892-1997(05)80250-4.

Huffman, F. E. (1976). The Register Problem in Fifteen Mon-Khmer Languages. *Oceanic Linguistics Special Publications, 13*, 575–589.

Iseli, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America, 121*(4), 2283–2295. https://doi.org/10.1121/1.2697522.

Iwata, R. (1995). The physiological characteristics of the voiced initial consonants in Suzhou Wu Chinese (苏州方言浊音发声的生理特性). In . 石锋, (Ed.), & 石鋒 & 李智強 *(Trans.), Studies of Chinese Abroad (汉语研究在海外)* (pp. 48–58). Beijing Language and Culture University Press (北京语言学院出版社).

Jiang, B., & Kuang, J. (2016). Consonant effects on tonal registers in Jiashan Wu. *Proceedings of the Linguistic Society of America*, *1*(0), 30-1-13. https://doi.org/10.3765/plsa.v1i0.3730.

Khan, S. D. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics, 40*(6), 780–795. https://doi.org/10.1016/j.wocn.2012.07.001.

Kirby, J. P. (2014). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics, 43*, 69–85. https://doi.org/10.1016/j.wocn.2014.02.001.

Kirby, J., Pittayaporn, P., & Brunelle, M. (2023). Transphonologization of onset voicing: Revisiting Northern and Eastern Kmhmu'. *Phonetica*. https://doi.org/10.1515/phon-2022-0029.

Kreiman, J., Iseli, M., Neubauer, J., Shue, Y.-L., Gerratt, B. R., & Alwan, A. (2008). The relationship between open quotient and H1*-H2*. *The Journal of the Acoustical Society of America, 124*(4), 2495–2495. https://doi.org/10.1121/1.4782813.

Kuang, J. (2013). The Tonal Space of Contrastive Five Level Tones. *Phonetica, 70*(1–2), 1–23. https://doi.org/10.1159/000353853.

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America, 142*(3), 1693–1706. https://doi.org/10.1121/1.5003649.

Kuang, J., Tian, J., & Jiang, B. (2019). The effect of vocal effort on contrastive voice quality in Shaoxing Wu. *The Journal of the Acoustical Society of America*, *146*(3), EL272–EL278. https://doi.org/10.1121/1.5126120.

Labov, W. (2001). *Principles of linguistic change. Volume 2: Social factors*. Blackwell publishers.

Kuang, J., & Cui, A. (2018). Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics, 71*, 194–214. https://doi.org/10.1016/j.wocn.2018.09.002.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Blackwell.

Li, R. (1966). A phonetic analysis of the Wenling dialect (温岭方言语音分析). *Studies of the Chinese Language (中国语文), 1*, 1–9.

Liu, F. (1925). *Étude expérimentale sur les tons du chinois*. Société Les Belles Lettres.

Mao, Z., & Meng, Z. (1986). *Sheyu Jianzhi*. Minzu Chubanshe.

Maspero, H. (1912). Étude sur la phonétique historique de la langue annamite: Les initiales. *Bulletin de l'École française d'Extrême-Orient, 12*, 1–126.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In L. Hyman (Ed.), *Consonant Types and Tone* (pp. 71–96).

Mazaudon, M. (2012). Paths to tone in the Tamang branch of Tibeto-Burman (Nepal). In G. De Vogelaer & G. Seiler (Eds.), *The Dialect Laboratory: Dialects as a testing ground for theories of language change* (pp. 139–177). John Benjamins Publishing Company.

Mazaudon, M., & Michaud, A. (2008). Tonal Contrasts and Initial Consonants: A Case Study of Tamang, a 'Missing Link' in Tonogenesis. *Phonetica, 65*(4), 231–256. https://doi.org/10.1159/000192794.

Mazaudon, M. (2005). On tone in Tamang and neighbouring languages: Synchrony and diachrony. In S. Kaji (Ed.), *Proceedings of the Symposium Cross-linguistic Studies of Tonal Phenomena: Historical Development, Tone-Syntax Interface, and Descriptive Studies* (pp. 79–96). Research Institute for Languages and Cultures of Asia and Africa (ILCAA), Tokyo University of Foreign Stud.

Michaud, A. (2012). Monosyllabicization: Patterns of evolution in Asian languages. In T. Stolz, N. Nau, & C. Stroh (Eds.), *Monosyllables: From Phonology to Typology* (pp. 115–130). Akademie Verlag. https://doi.org/10.1524/9783050060354.115.

Michaud, A., & Sands, B. (2020). Tonogenesis. In I. A. Michaud & B. Sands (Eds.), *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.748.

Ming, M., & Zhang, X. (2015). The patterns and evolution mechanism of tone value of Chinese dialects (试论汉语方言的调值格局及其演变机制). *Journal of Southwest University (Social Sciences Edition) (西南大学学报(社会科学版)), 41*(04), 145–155 +192.

Osatananda, V. (1997). *Tone in Vientiane Lao*. University of Hawai'i at Manoa [PhD thesis]..

Pittayaporn, P. (2018). Phonetic and Systemic Biases in Tonal Contour Changes in Bangkok Thai. In H. Kubozono & M. Giriko (Eds.), *Tonal Change and Neutralization* (pp. 249–278). De Gruyter. https://doi.org/10.1515/9783110567502-010.

Pittayaporn, P., & Kirby, J. (2017). Laryngeal contrasts in the Tai dialect of Cao Baň˘g. *Journal of the International Phonetic Association, 47*(1), 65–85. https://doi.org/10.1017/S0025100316000293.

Premsrirat, S. (2003). Khmu dialects: A case of register complex and tonogenesis. In S. Kaji (Ed.), *Ross-linguistic studies of tonal phenomena: Historical development, phonetics of tone, and descriptive studies* (pp. 13–28). ILCAA.

Pulleyblank, E. G. (1978). The nature of the Middle Chinese tones and their development to Early Mandarin / 中古汉语声调的本质和到早期官话的演变. *Journal of Chinese Linguistics, 6*(2), 173–203.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ren, N. (1992). *Phonation types and stop consonant distinctions-Shanghai Chinese*. The University of Connecticut [PhD thesis]..

Rose, P. (1989). Phonetics and Phonology of Yang Tone Phonation Types in Zhenhai. *Cahiers de Linguistique Asie Orientale, 18*(2), 229–245. https://doi.org/10.1163/19606028-90000316.

Sagart, L. (1990). Review of The History and Development of Tonal Systems and Tone Alternations in South China. *Language, 66*(3), 614–618. https://doi.org/10.2307/414625.

Sagart, L. (2002). Gan, Hakka and the formation of Chinese dialects. *Dialect Variations in Chinese*, 129–154.

Sanders, R. (2008). Tonetic sound change in Taiwan Mandarin: The case of Tone 2 and Tone 3 citation contours. *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, 87–107.

Schertz, J., & Khan, S. (2020). Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. *Journal of Phonetics, 81*. https://doi.org/10.1016/j.wocn.2020.100979 100979.

Seyfarth, S., & Garellek, M. (2018). Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics, 71*, 425–450. https://doi.org/10.1016/j.wocn.2018.09.001.

Shi, F. (1983). Acoustic characteristics of voiced obstruents in Suzhou Chinese [in Chinese]. *Studies in Language and Linguistics (Yuyan Yanjiu), 01*, 49–83.

Shi, M., Chen, Y., & Mous, M. (2020). Tonal split and laryngeal contrast of onset consonant in Lili Wu Chinese. *The Journal of the Acoustical Society of America, 147*(4), 2901–2916. https://doi.org/10.1121/10.0001000.

Shi, F. (1990). The Reanalysis of the voiced initial consonants in Suzhou Chinese (苏州话浊音声母的再分析). In *Explorations in phonetics (*语音学探微*)* (pp. 164–176). Peking University Press (北京大学出版社).

Shorto, H. L. (1966). Mon vowel systems: A problem in phonological statement. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 398–409). Longmans.

Shue, Y.-L., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. *Proceedings of ICPhS XVII*, 1846–1849.

Simpson, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics, 40*(3), 477–490. https://doi.org/10.1016/j.wocn.2012.02.001.

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics, 84*. https://doi.org/10.1016/j.wocn.2020.101017 101017.

Sóskuthy, M., Hay, J., & Brand, J. (2019). Horizontal diphthong shift in New Zealand English. *Proceedings of the 19th International Congress of Phonetic Sciences*, 597–601.

Svantesson, J.-O. (1989). Tonogenetic Mechanisms in Northern Mon-Khmer. *Phonetica, 46*(1–3), 60–79. https://doi.org/10.1159/000261829.

Svantesson, J.-O., & House, D. (2006). Tone Production, Tone Perception and Kammu Tonogenesis. *Phonology, 23*(2), 309–333.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica, 19*(2), 333–363. https://doi.org/10.1075/dia.19.2.04thu.

Thurgood, G. (2020). Tonogenesis. In R. D. Janda, B. D. Joseph, & B. S. Vance (Eds.), *The Handbook of Historical Linguistics* (pp. 47–62). John Wiley & Sons, Ltd.. https://doi.org/10.1002/9781118732168.ch3.

Tian, J., & Kuang, J. (2021). The phonetic properties of the non-modal phonation in Shanghainese. *Journal of the International Phonetic Association, 51*(2), 202–228. https://doi.org/10.1017/S0025100319000148.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2020). *Itsadug: Interpreting time series and autocorrelated data using GAMMs*.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics, 31*(2), 181–201. https://doi.org/10.1016/S0095-4470(02)00086-4.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* ((Second edition).). CRC Press/Taylor & Francis Group.

Xu, W., & Mok, P. (2021). The Acoustic Correlates and Time Span of the Non-modal Phonation in Kunshan Wu Chinese. *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, 1–5. https://doi.org/10.1109/ISCSLP49672.2021.9362083.

Yang, C., Pittayaporn, P., Kirby, J., & Jitwiriyapont, S. (2021). Change and stability in the tonal contours of King Rama IX of Thailand, 1959-1997. *1st International Conference of Tone and Intonation*.

Yang, C., & Xu, Y. (2019). Cross-linguistic trends in tone change: A review of tone change studies in East and Southeast Asia. *Diachronica, 36*(3), 417–459. https://doi.org/10.1075/dia.18002.yan.

Yao, Y., & Chang, C. B. (2016). On the cognitive basis of contact-induced sound change: Vowel merger reversal in Shanghainese. *Language, 92*(2), 433–467. https://doi.org/10.1353/lan.2016.0031.

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America, 136*(3), 1320–1333. https://doi.org/10.1121/1.4887462.

Zhang, J. (2014). *A sociophonetic study on tonal variation of the Wuxi and Shanghai dialects*. Netherlands Graduate School: LOT.

Zhang, Z. (2016a). Cause-effect relationship between vocal fold physiology and voice production in a three-dimensional phonation model. *The Journal of the Acoustical Society of America, 139*(4), 1493–1507. https://doi.org/10.1121/1.4944754.

Zhang, Z. (2016b). Mechanics of human voice production and control. *The Journal of the Acoustical Society of America, 140*(4), 2614–2635. https://doi.org/10.1121/1.4964509.

Zhang, W., Xu, W., & Mok, P. (2019). A Preliminary study on the age variation of the voicing contrast in Wenzhou Wu Chinese. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)*.

Zhang, J., & Yan, H. (2018). Contextually dependent cue realization and cue weighting for a laryngeal contrast in Shanghai Wu. *The Journal of the Acoustical Society of America, 144*(3), 1293–1308. https://doi.org/10.1121/1.5054014.

Zheng, X. (2006). Voice Quality Variation with Tone and Focus in Mandarin. *Proceedings TAL, 2006*, 132–136.

Zhu, X. (2010). Slack Voice: The Phonetic Nature of the Voiced Consonants in Ancient Chinese (全浊弛声论兼论全浊清化(消弛)低送高不送). *Studies in Language and Linguistics (语言研究), 30*(03), 1–19.

Zhu, X., & Jiao, L. (2011). Short tones without a plosive coda: A new type of the entering tone syllable (短调无塞尾——报告一例特殊的入声类型). *Bulletin of Linguistic Studies, 语言研究集刊), 8 (*第八辑, 142–147.

Zsiga, E. (2008). Modeling diachronic change in the Thai tonal space. *University of Pennsylvania Working Papers*. *Linguistics, 14*(1), 395–408.