Journal of Phonetics 105 (2024) 101330

Contents lists available at ScienceDirect

Journal of Phonetics

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

The effect of breathy voice on tone identification by listeners of different ages in Suzhou Wu Chinese

Chunyu Ge, Peggy Mok *

The Chinese University of Hong Kong, Shatin, Hong Kong

A R T I C L E I N F O

Article history:
Received 19 December 2022
Received in revised form 12 March 2024
Accepted 3 May 2024

Keywords:
Breathy voice
Tone perception
Talker voice
Suzhou Wu Chinese

A B S T R A C T

Suzhou Wu Chinese has undergone a transphonologization of a voicing contrast in initial consonants to a tone contrast. In consequence, the tone system has split into two registers, in which the high register tones are higher in pitch and modal voiced, whilst the low register tones are lower in pitch and breathy voiced. Our previous studies have found that breathy voice in the low register tones is disappearing in younger speakers' production. This finding motivated us to investigate the effect of breathy voice on tone identification across age groups. Participants from three age groups completed a tone identification experiment. Stimuli were constructed based on natural tokens produced by a middle-aged female speaker and an older female speaker. The manipulation of phonation was accomplished by using the base syllables of both high and low register tones, for both unchecked (T1 vs. T2) and checked (T7 vs. T8) tone pairs. The results showed that breathy voice is still used by younger listeners in their perception and its effect on their tone identification is similar to that for older and middle-aged listeners. Moreover, the effect of breathy voice is modulated by social indexical factors (i.e., talker voice). The implications of the results for the origin of the loss of breathy voice in Suzhou Wu and the mechanism of sound change are discussed.
© 2024 Elsevier Ltd. All rights reserved.

1. Introduction

Suzhou Wu Chinese is a Chinese dialect which used to have a voicing contrast in initial consonants. The voicing contrast has transphonologized into tone contrasts and caused a two-way tone split, in that the syllables with historically voiceless initial consonants developed into high register tones with a higher pitch and modal voice, whereas those with historically voiced consonants developed into low register tones with a lower pitch and breathy voice. Previous studies showed that there is an ongoing sound change in Suzhou Wu, in which the breathy voice in the low register tones is decreasing in the production of younger speakers (Ge et al., 2023). In this study, we aim to investigate how breathy voice influences the tone perception of younger listeners compared to older and middle-aged listeners, and the effect of talker voice is also investigated.

1.1. Tone split and the loss of breathy voice in Suzhou Wu

Tone split is the transphonologization of the voicing contrast in onset consonants into tone contrasts, in which the previously voiceless consonants give rise to high tones and the historically voiced consonants induce low tones (Michaud & Sands, 2020; Thurgood, 2020). In some languages with tone split, the previously voiced obstruents also induce breathy voice on the vowel. Although it is still not clear whether breathy voice is an obligatory stage in tone split (see counter evidence in Brunelle, Brown and Phạm, 2022 and Kirby, Pittayaporn and Brunelle, 2023), breathy voice is known to play a critical role in tone split in at least some languages (Thurgood, 2002, 2007; Chen, 2015). Breathy voice induced by voiced initial consonants in tone split is attested in Wu Chinese (Cao & Maddieson, 1992), Cham (Kingston, 2011), and Tamang-Gurung-Thakali-Manangke (TGTM) languages (Mazaudon & Michaud, 2008; Mazaudon, 2012), to name just a few. When tone split is completed, the voicing contrast transphonologized into pure tone contrasts based solely on f0, and breathy voice disappears. For example, in Cantonese, syllables with previously voiceless obstruent onsets have high tones, while those

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

https://doi.org/10.1016/j.wocn.2024.101330
0095-4470/© 2024 Elsevier Ltd. All rights reserved.

with previously voiced obstruent onsets have low tones with no breathy voice.

Northern Wu Chinese is a group of Chinese dialects spoken in Shanghai, Zhejiang and the southern part of Jiangsu province and it is in the final stage of tone split. As other languages with tone split, tone systems in Northern Wu Chinese can be divided into two registers. High register tones co-occur with voiceless initial obstruents and modal voice, while low register tones co-occur with voiced initial obstruents and breathy voice. For example, the tone system of Suzhou Wu and its co-occurrence restrictions with initial consonant voicing and phonation type can be seen in Table 1 and the pitch contours in Fig. 1. The phonetic realization of voiced obstruents in Wu Chinese has long been known as "voiceless sound with voiced aspiration" (清音浊流, Chao, 1928, 1967). Instrumental studies have confirmed that the voiced aspiration is indeed breathy voice on the following vowel (Cao & Maddieson, 1992; Ren, 1992; Tian & Kuang, 2021; Gao & Hallé, 2017). Vowels following voiced obstruents have larger spectral tilts, a stronger noise component and a larger glottal opening than those following voiceless obstruents (see thorough reviews of the acoustic and articulatory properties of breathy voice in Northern Wu Chinese in Tian & Kuang, 2021; Ge et al., 2023). However, breathy voice of the low register tones in Northern Wu Chinese is reported to be disappearing recently. In Shanghai Wu (Gao, 2016), Lili Wu (Shi et al., 2020) and Suzhou Wu (Ge et al., 2022, 2023), younger speakers make less use of breathy voice than older speakers do. This suggests that Northern Wu is in the final stage of tone split. As breathy voice disappears in Northern Wu, the tone systems are expected to rely solely on f0.

Our earlier production study revealed the change in breathy voice in Suzhou Wu after tone split (Ge et al., 2023), using simultaneous audio and electroglottographic data from 36 speakers in three age groups. First of all, the results showed that T2 and T8 in Suzhou Wu are changing. As shown in Fig. 1, the pitch contours of T2 and T8 are rising for older speakers, but for the middle-aged speakers, the rising starts later than the older speakers, and for younger speakers, the pitch contour becomes concave with a clear initial falling. This is especially obvious for T2. Moreover, creaky voice appears in the middle of T2 syllables for younger speakers. With regard to breathy voice, acoustic and articulatory measurements of phonation types were compared among older, middle-aged and younger speakers. Using Principal Component Analysis and Linear Discriminant Analysis, it is found that phonation cues are less successful in discriminating high and low register tones for younger speakers than for older and middle-aged speakers, which indicates that younger speakers make less use of breathy voice in low tones. The disappearance of breathy voice for low tones was also investigated using Generalized Additive Mixed Models. The results indicated that breathy voice is disappearing in all low register tones (T2, T6 and T8), but it is most advanced for T2. For T6 and T8, the magnitude of breathy voice is smaller for younger speakers than for older and middle-aged speakers, and there is almost no breathy voice in T2 produced by younger speakers.

Since breathy voice is disappearing in production in Suzhou Wu, it is natural to ask whether its loss is also mirrored in perception. This study aims to investigate the effect of breathy voice on tone identification by listeners of different age groups. Before presenting the detailed motivation and overview of the present study, the next section reviews the role of speech perception in sound change and focuses especially on the effect of listener age and talker voice. The effect of phonation type on tone perception is reviewed after that.

1.2. The effect of listener age and talker voice on speech perception of ongoing sound change

Speech perception is fundamental in Ohala's listener-based theory of sound change (Ohala, 1981, 1989). Listeners' misperceptions lead to failure in interpreting the acoustic signal as the one intended by the speaker, and thus initiates sound change. This idea is revised by Beddor (2009), in that listeners can be accurate perceivers yet they come up with different grammars. She investigated the production and perception of VN sequences and found that the nasal consonant (the coarticulatory source) and vowel nasalization (the coarticulatory effect) were perceptually equivalent. Listeners can be very accurate at perceiving the subtle phonetic patterns of coarticulation, but they can come up with grammars that are compatible with the patterns but different from those intended by the speakers.

As perception plays a fundamental role in sound change, sound change in progress can be manifested in speech perception in apparent time. Perceptual differences are found between different age groups in previous studies. On the one hand, as younger speakers are more advanced in the production of sound changes, they are also more advanced in perception change compared to older speakers. For example, there is a sound change in Standard British English, the lax /ʊ/-fronting, and younger listeners compensate less for lax /ʊ/-fronting in perception than older listeners do (Kleber et al., 2012). The phenomenon that younger speakers are more advanced than older speakers in perception change has been repeatedly reported in other studies on ongoing sound change, such as the centralization of /æ/ in Canadian English (De Decker, 2010), and the raising of /a/ in Copenhagen Danish (Pharao et al., 2015). On the other hand, however, a more complex picture appears as younger listeners may not show considerable difference in perception from older speakers. Kettig and Winter (2017) found that younger speakers in Jewish Montreal English have the retraction of /æ/ and the lowering and retraction of /ɛ/ as compared to older speakers, but in perception, younger and older listeners categorize the vowels similarly.

Table 1
Tones in Suzhou Wu with their registers, phonation type and initial consonant voicing (historical T4 has merged with T6, tones transcribed using Chao's letters (Chao, 1930). The transcriptions are based on Ye (1979) and Wang (1987)).

	Ping 平声	Shang 上声	Qu 去声	Ru 入声
high register, modal, voiceless	**T1** 44	**T3** 51	**T5** 412	**T7** 55
low register, breathy, voiced	**T2** 223		**T6** 231	**T8** 23

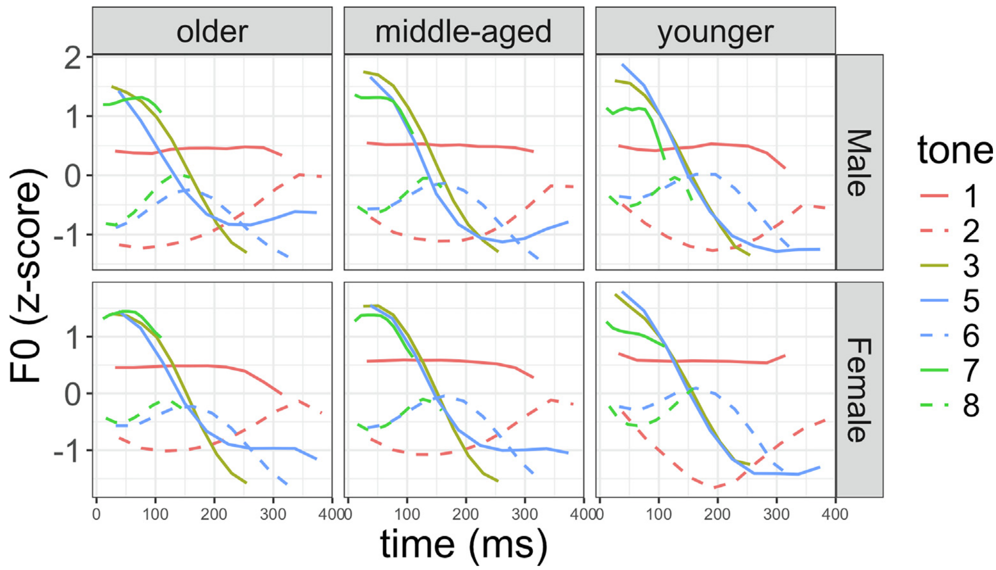

Fig. 1. The averaged pitch contours of Suzhou tones across three age groups.

More relevant to the present study, previous studies on the transphonologization of consonant voicing also show that younger listeners are still sensitive to the disappearing cue in perception, despite the fact that they produce it less than older speakers do. In Afrikaans, younger speakers produce voiced stops with less prevoicing than older speakers do, but with a larger difference in F0. In perception, however, both older and younger listeners use F0 as a perceptual cue, and prevoicing is still used by some younger listeners (Coetzee et al., 2018). Similar results were also found by Kuang and Cui (2018) in Southern Yi, where the register contrast is gradually changing from phonation-based to vowel quality-based. They found that the change occurs first in production but younger listeners' sensitivity to phonation in perception is retained for the front vowel syllable /be/. Nevertheless, the agreement in perception between younger and older listeners despite production differences is not unexpected, because during the progress of sound change, innovative and conservative variants are both found in the community and listeners need to accommodate both in their perception (Kettig & Winter, 2017).

In addition to the influence of listener age on the perception of sound change, social information, such as talker voice, can further modulate the perception process. For example, in New Zealand English, the NEAR /iə/ and SQUARE /eə/ vowels are merging and younger speakers do not make a distinction between these two diphthongs in their production, but this distinction can still be heard in the speech of older speakers. However, Warren et al. (2007) showed that younger speakers can employ the difference between NEAR and SQUARE vowels in their perception when hearing stimuli constructed using an older voice. Hay et al. (2006) further demonstrated that listeners' expectation of a speaker's age (i.e., the perceived age) also influences their perception. When the perceived age is older, the perception of identical stimuli is more distinct than when the perceived age is younger. These experiments suggest that listeners are sensitive to the cues that are no longer present in their own production but are present in the speech community, and the use of these cues is related to social indexical factors (e.g., age). Recently, Schertz et al. (2019) also investigated the effect of talker voice on the perception of the Korean three-way stop contrast. There is an incipient tonogenesis in Seoul Korean (Kang, 2014; Bang et al., 2018), in which the stop contrasts are gradually shifting into f0 contrasts. But this sound change is not found in all Korean dialects, such as Hunchun Korean. However, Schertz et al. (2019) found that younger listeners of Hunchun Korean are sensitive to the f0 difference, and more importantly, the younger listeners rely even more on f0 when hearing a Seoul talker as compared to a Hunchun talker. Although the effect of talker voice on the perception of ongoing sound change is less investigated than listener age, these studies indicate that listeners are indeed sensitive to talker voice during the progress of sound change.

In summary, both listener age and talker voice can influence the perception of ongoing sound change. This study also investigates the effect of both factors on the perception of breathy voice in Suzhou Wu, and the next section reviews previous studies on the perception of phonation type in tone identification.

1.3. The effect of phonation type on tone perception

Phonation type can form lexical contrasts together with tone/pitch and it can have different relationships with tone (Esposito & Khan, 2020; Garellek, 2022). Phonation and tone can be fully orthogonal, i.e., all possible combinations of phonation and tone contrasts are attested, as in Jalapa Mazatec (Garellek & Keating, 2011). Phonation can also be an allophonic cue for a certain tone, as in the creaky voice of Tone 3 in Mandarin Chinese (Kuang, 2017) and Tone 4 in Cantonese (Yu & Lam, 2014). It can also be fused with tone to form register systems, which is widely found in Southeast Asian languages (Brunelle & Tạ, 2021).

The effect of phonation on tone perception has been explored in a number of languages. Non-modal phonation can be classified according to glottal aperture, with breathy voice at the most open end and creaky voice at the most closed end (Gordon & Ladefoged, 2001). Despite the mixed findings on the effect of creaky voice on tone perception (see Yang, 2015 on Mandarin; Yu & Lam, 2014; Zhang & Kirby, 2020 on Cantonese; Garelle et al., 2013 on White Hmong), breathy voice has an important and consistent influence on tone perception. The language investigated by the present study, Wu Chinese, has breathy voice in the low register tones, which is associated with the historically voiced initial consonants. Shanghai Wu is a representative dialect of Northern Wu Chinese and some perception experiments have been conducted on it. Ren (1992) investigated the perception of breathy voice in Shanghai Wu with synthesized stimuli using a formant synthesizer with nine steps of open quotient (OQ) and three steps of F0 onset. The participants were asked to make a two-alternative forced choice between high (with modal voice) and low (with breathy voice) register tones and the results showed that breathy voice (larger OQ) shifted the response towards the low register tones. Zhang and Yan (2018) employed a different approach and the manipulation of the phonation of stimuli was accomplished by cross-splicing the initial consonant (either voiceless or voiced) and vowel (either modal or breathy voiced). Breathy voiced vowels were more likely to be identified as having a low tone. The role of breathy voice in tone perception in Shanghai Wu was also investigated by Jiang et al. (2020), together with another Wu Chinese dialect, Jiashan Wu. They found that although both Shanghai and Jiashan listeners use breathy voice as a secondary cue for tone contrasts, breathy voice is more important in the perception of Jiashan listeners than it is for Shanghai listeners. They took it as evidence reflecting the decrease in breathy voice in the production of younger Shanghainese speakers reported by Gao (2016) and Zhang and Yan (2018). Moreover, the perception of breathy voice by younger speakers of Shanghai Wu was also investigated by Gao et al. (2020), using both natural and synthesized stimuli. They found that breathy voice can facilitate the identification of the low register Tone 3 even if breathy voice is produced less by younger speakers than older speakers.

It should be highlighted that the influence of phonation on pitch perception can be relatively independent of production. That is, native speakers of languages that do not employ phonation to construct lexical contrasts can also use phonation to facilitate their perception of pitch. For example, Kuang and Liberman (2018) tested English speakers' perception of pitch and voice quality by constructing stimuli with two pitch peaks and shifted the spectral slope of each peak so that one of the peaks was either breathier or tenser than the other. Native English speakers were asked to choose which peak had a higher pitch. Both speech and non-speech stimuli showed that breathier voice quality can shift pitch perception to be lower than that of tenser voice quality. Therefore, it can be concluded that phonation cue can influence tone perception and it is an important cue for pitch perception cross-linguistically.

1.4. The present study

Based on these backgrounds, the present study aims to investigate whether listener age and talker voice would influence the effect of breathy voice on tone identification in Suzhou Wu, in which breathy voice is disappearing in the production of younger speakers. As reviewed in the section above, there have been a few studies on the perception of breathy voice in Wu Chinese (Ren, 1992; Zhang & Yan, 2018; Gao et al., 2020; Jiang et al., 2020), with Gao et al. (2020) being the only one that intended to study the perception of breathy voice in the scenario of its disappearance in Shanghai Wu. Nevertheless, the participants they recruited included younger listeners only. To better understand the loss of breathy voice and its transmission in the community, other age groups such as older and middle-aged listeners should be included and compared with younger listeners. Moreover, because talkers' age can also influence speech perception, this study also included two talker voices (older vs. middle-aged female voices) to determine whether the listeners are sensitive to talker voices. As talker voice has been rarely investigated in the perception of phonation types, this study is exploratory in this direction and only used naturally produced tokens in the experiment. In addition, previous studies seldom included the checked tone pairs in perception, with Jiang et al. (2020) being the only one which used checked tone pairs. Unchecked tones in Wu Chinese are syllables ending with a vowel or a nasal, while checked tones are syllables ending with a glottal stop. Our study also included checked tone pairs together with unchecked tone pairs in the perception experiment. Therefore, although Suzhou Wu is not unique in the loss of breathy voice as compared to other Northern Wu Chinese dialects, our study is unique in investigating the relationship between age effects (both listener age and talker voice) and the effect of breathy voice on tone identification.

Based on the previous studies, some general predictions can be made on the effect of listener age and talker voice. First, phonation is an important cue for pitch perception cross-linguistically and even listeners who do not use it in their own language (as in English, Kuang and Liberman (2018)) can be sensitive to it. As breathy voice is still being produced by older and middle-aged speakers in Suzhou Wu, it should still have a role to play in younger listeners' perception. But since breathy voice is disappearing in younger speakers' production, it is not clear whether the effect of breathy voice in perception is the same for younger and older/middle-aged listeners. Finally, as revealed by Warren et al. (2007) and Hay et al. (2006), a talker's age can influence how listeners perceive the stimuli. Therefore, it is reasonable to expect that stimuli of an older voice would induce a larger effect of breathy voice on tone perception than a middle-aged voice.

By examining the effect of listener age and talker voice on the perception of breathy voice, this study also hopes to shed light on the origins of the loss of breathy voice in Suzhou Wu and also contribute to our understanding of the relationship between perception and production in sound change (cf. Beddor et al. (2018)). Although the loss of one cue in production does not immediately lead to listener's insensitivity to it, it would be possible that different listeners, especially listeners of different age groups would perceive it differently. In this study, if the loss of breathy voice originates in production, then even if it is no longer present in younger speakers' production, they should still be able to perceive it. On the contrary, if this sound change is perceptually driven, in that the younger speakers produce less breathy voice because they are not sensitive to it in perception, the effect of breathy voice on tone perception would be much smaller for younger listeners than for older/middle-aged listeners.

2. Method

2.1. Participants

Data from 57 participants were collected. All of the participants were native speakers of Suzhou Wu. They all spoke Mandarin Chinese as a second language, and 40 of them (all the younger and middle-aged speakers and one older speaker) had learnt at least one foreign language, mostly English. The participants were divided into three age groups. Table 2 shows the valid number of participants in each gender and age group. Younger participants were those aged under 32 (μ = 24.5, σ = 2.68), while older participants were those aged over 58 (μ = 61.7, σ = 4.08) and middle-aged participants were those aged 37–51 (μ = 44.4, σ = 4.28). One of the younger male participants seemed to respond randomly in the perception experiment, with accuracy at near chance level for all pitch steps and with extremely small response time. Therefore, the data of this participant were discarded and there were 56 valid participants in total.

2.2. Stimuli

Four sets of T1–T2 and T7–T8 pairs were selected as the base syllables for resynthesis. The pair T3–T6 was not included because the pitch contours of T3 and T6 differ greatly. While T3 is a high falling tone, T6 is a rising falling tone. For each tone pair, the base syllables with stop (/p, b/) and fricative (/s, z/) onsets produced by a middle-aged (49) and an older (67) female speaker were selected. Mean spectral measurements of the base syllables were made during the vowel interval and the data of the common spectral tilt measurements (H1*–H2*, H1*–A1*, H1*–A2* and H1*–A3*) were in Tables A3–A6 in Appendix A. It is clear that for each minimal pair, the base syllables of the low register tones have larger values than those of the high register tones, indicating breathier voice in the low register tones. The older voice is comparable to the middle-aged voice in some measurements, but there is a tendency that the older voice is breathier than the middle-aged voice, indicated in particular by the larger values of H1*–H2* of the older voice than the middle-aged voice for both the high and low register tones. However, the differences between the high and low register tones are in general similar for the older and the middle-aged voices, despite some variations. We did not choose the base syllables produced by a younger speaker because the tone changes in T2 and T8 have rendered the pitch contour dramatically different from the typical rising contour (Fig. 1) and there is heavy creaky voice in the middle of T2 syllables produced by younger speakers. The vowel was /ɑ/ for the unchecked tone pairs and /aʔ/ for the checked tone pairs. More details on the phonetic realization of the checked tones can be found in Ge et al. (2023) and Gao and Kuang (2022). The onsets were all phonetically voiceless, with no prevoicing in the voiced stop /b/ and no voicing during the frication in the fricative /z/. The transcription using /b z/ is to indicate their historical origin and also their realization to be truly voiced in intervocalic positions (for more information, see Ge et al., 2023 and Ge & Mok, 2023). The durations of the onsets were not manipulated, but we intentionally selected syllables with similar onset duration. Table A2 in Appendix A displays the duration of the onsets of the base syllables. The words used and their transcriptions are shown in Table 3, together with the Chinese characters and the English gloss.

Using the base syllables, a series of pitch continua were superimposed on the syllables of both high and low register tones using the Pitch Synchronous Overlap Add (PSOLA) method implemented in Praat (Boersma, 2020; Valbret et al., 1992). Before the superimposition, the duration of the rimes was neutralized such that the duration of T1 and T7 was lengthened to that of T2 and T8, respectively. The reason is that as the duration of T2 and T8 is longer than T1 and T7, compressing the duration of T2 and T8 would result in alteration of the pitch slope. For T1 and T7, the adverse effect is smaller, as T1 is a level tone and the shape of T7 is also largely level (though slightly falling). After the neutralization of duration, the pitch contour was manipulated. The pitch height was manipulated because the pitch contours of low register tones (T2 and T8) do not rise to the level of high register tones (T1 and T7), and manipulating the starting point alone is not enough. The pitch contours of the pitch continua for T1–T2 and T7–T8 pairs are shown in Fig. 2. The ends of the pitch continua were determined by the average pitch contours of female speakers. For T1 and T2, eight steps of F0 were created for the

Table 2
The number of valid participants in each gender and age group.

Age group	Female	Male	Age range
Younger	10	10	21–32
Middle-aged	10	10	37–51
Older	8	8	58–72
Total	28	28	–

Table 3
Transcriptions and glosses of the base syllables.

Phonation	Onset	Tone	IPA	Gloss	Tone	IPA	Gloss
Modal	Stop	T1	/pɑ/	巴 “bar”	T7	/paʔ/	百 “hundred”
	Fricative	T1	/sɑ/	篩 “sieve”	T7	/saʔ/	殺 “kill”
Breathy	Stop	T2	/bɑ/	牌 “card”	T8	/baʔ/	拔 “pull”
	Fricative	T2	/zɑ/	柴 “firewood”	T8	/zaʔ/	閘 “floodgate”

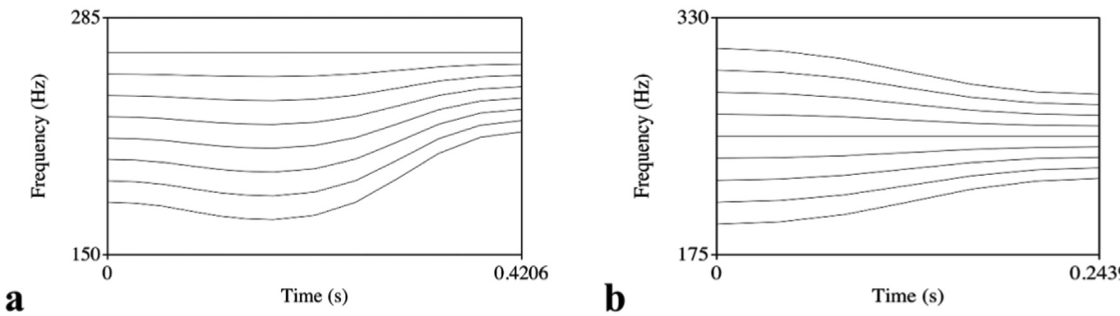

Fig. 2. Pitch continua of the (a) unchecked and (b) checked tone pairs.

onset, turning point (0.4 of the whole vowel) and offset of the pitch contour. The whole pitch contour was then interpolated using the three points using the quadratic interpolation function in Praat. Nine steps of F0 were created for T7 and T8. The reason for this was that the difference between T7 and T8 is slightly greater than that between T1 and T2, and the value of each step was controlled to be around the just-noticeable difference of F0, i.e., ~10 Hz (Qin et al., 2019). After the manipulation of F0, the intensity of all the stimuli was scaled to 70 dB. The perception experiment consisted of 136 randomized stimuli, i.e., 64 tokens of unchecked tones (8 pitch contours × 2 phonation types × 2 onsets × 2 voices) and 72 tokens of checked tones (9 pitch contours × 2 phonation types × 2 onsets × 2 voices). All stimuli were presented twice, thus there were 272 stimuli in total for each participant. The base syllables, the stimuli, and the scripts used to resynthesize the stimuli can be found online (https://osf.io/xdajf/).

There is one point worth noting about the base syllables. As pointed out by a reviewer, the syllable /paʔ/ 百 "hundred" used to have the vowel /ɑ/. However, Ling (2009, pp 94–97) has shown that the two vowels /a/ and /ɑ/ have been merged in checked tone syllables in the new accent of Suzhou Wu and is in the process of merging even for the old accent. The formants at the midpoint of the vowel of the two base syllables we used were very similar, especially in F2 (see Table A1 in Appendix A). The base syllables for both the middle-aged and the older speakers were pronounced as a low front vowel /a/, not a back /ɑ/. As the stimuli we used do not differ in the front-back dimension (F2), the listener would rely on pitch and phonation to identify the syllables. Therefore, the perception results are valid. The word /paʔ/ 百 "hundred" was selected because it is the only suitable word in our production materials and during the pandemic, we were unable to revisit the local community to collect more production data. Of course, this limitation should be avoided in future studies and a thorough study is needed to determine whether the ongoing merger has any effect on the perception of pitch and phonation.

2.3. Procedure

The COVID-19 pandemic prevented us from conducting the perception experiment on-site. The experiment was thus run online using PsyToolkit (Stoet, 2017, 2010), with the assistance of a local speaker, particularly to help the middle-aged and older participants to use the computer. The participants were instructed to take the test wearing earphones or headphones. The experiment was a word identification task. Two Chinese characters appeared on the screen. The order of the characters was randomized. There was a practice session prior to the test session. Because we did not manipulate the word frequency of the test words, the practice session familiarized the participants with the task and the stimuli. During the practice, all the base syllables were presented so as to activate the words and mitigate the influence of word frequency in the test session. The participants were asked to place their index and middle fingers of their dominant hand on the keys "N" and "M". If what they heard was the character on the left, the participant would press "N", and if it was the character on the right, they would press "M". There was a fixation cross in the middle of the screen between trials, which disappeared and changed into the characters once the stimuli started to play. The next stimulus would be played if there was no response after 5000 ms (5 s) or as soon as a response was given.

After the perception experiment, they also completed a questionnaire on their language background. The whole experiment was loaded into the browser beforehand and thus the measurement of response time did not rely upon internet connection. It has been reported that the performance of PsyToolKit is comparable to E-Prime in measuring response time (Kim et al., 2019). The entire experiment together with the questionnaire took approximately 30 minutes. Each participant was given a nominal payment for their participation.

2.4. Data analysis

There were 272 responses for each participant. Among all the responses, 29 tokens were without response. They were removed from analysis, resulting in 0.23% of data loss.

There were four factors in total. Pitch step was 0–7 for the unchecked tone pairs and 0–8 for the checked tone pairs, in which 0 was the high pitch end and 7/8 was the low pitch end. There were two kinds of Phonation, modal vs. breathy; two types of Onset, stop vs. fricative onset; and two kinds of Talker Voice, older vs. middle-aged voice. The last variable was the Age group of the listeners, either younger, middle-aged or older. The two-level factors Phonation and Talker voice were coded using sum/deviation coding, such that the coefficients indicate main effects. The three-level factor Age group was coded using Helmert coding, to allow comparisons across three levels. The levels of Age group were ordered as older vs. middle-aged and younger listeners. Thus, the first comparison was between the older and the mean of the middle-aged and the younger speakers, while the second comparison was between the middle-aged and the younger speakers. In addition to identification accuracy, response time (RT) was also analyzed. RT was log-transformed to approximate a normal distribution. As different steps were used for the unchecked and checked tone pairs, they were analyzed separately.

3. Results

In presenting the results, we first describe the patterns of identification results by tone pairs (unchecked and checked) and then the results for RT.

3.1. Identification data

3.1.1. The unchecked tone pairs

The identification curves of the unchecked tone pairs are shown in Fig. 3. The vertical axis is the proportion of low register tone responses, and the horizontal axis is the step of the pitch continuum, which ranges from 0 to 7 (the high end to the low end). Responses to the stimuli with stop onsets are shown in Fig. 3(a) and those to the stimuli with fricative onsets in Fig. 3(b). In each figure, the column indicates the age group (older, middle-aged, or younger) of the participants, while the row indicates the talker voice (older or middle-aged) of the stimuli. The phonation of the base syllable is indicated by line type. Responses to the stimuli with modal voice are in a solid line, and those to the stimuli with breathy voice are in a dashed line. The identification proportions are averaged over the participants within each age group, and the error bar shows the standard error.

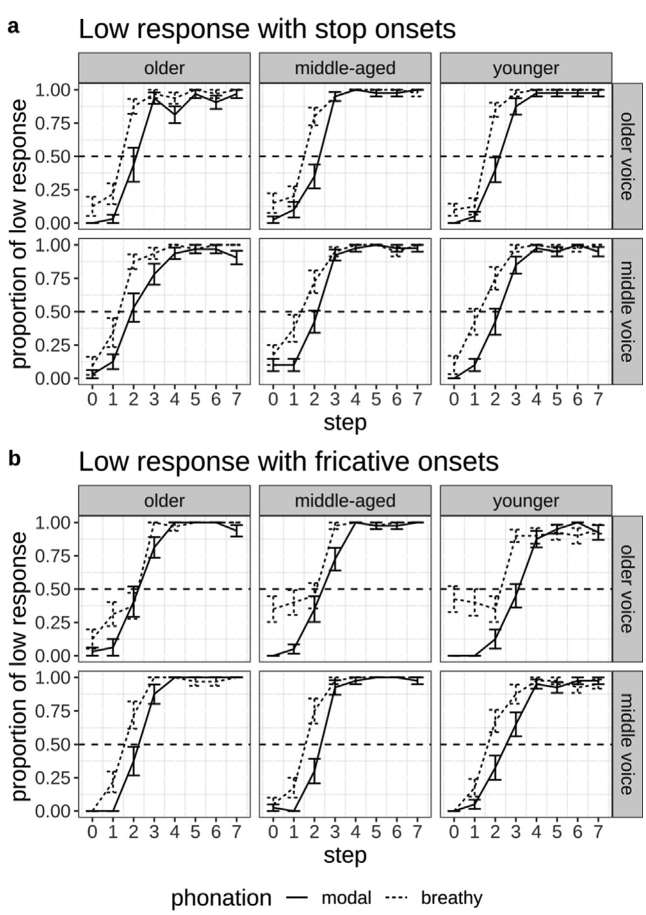

Fig. 3. The identification curves of the unchecked tone pairs with stop (a) and fricative (b) onsets.

It is clear that the identification of T1 vs. T2 is categorical. The dashed line is generally to the left of the solid line, i.e., breathy voice can shift the identification curve to the higher end of the pitch continuum. In other words, when the stimuli were with breathy voice, the participants would accept a higher pitch as the low register tones, compared to the stimuli with modal voice. This effect was present for participants of all age groups, which indicates that breathy voice can still influence the perception of the younger listeners in a way similar to the older and middle-aged listeners.

To compare the effect of breathy voice across the three age groups, mixed-effect logistic regression models were conducted on the binary response data, using the glmer function in the lmerTest package (Kuznetsova et al., 2017). The binary response (T1 as 0 vs. T2 as 1) was the dependent variable, and Phonation (modal vs. breathy), Step (0–7), Age group (older vs. middle-aged vs. younger), talker Voice (older vs. middle-aged), and their interactions were fixed effects, with participant as random intercepts. As noted by Barr et al. (2013:275), the analysis of categorical data can adopt a more data-driven approach and in the present analysis, random slopes were not included because models including these terms encountered issues of convergence. The results of the maximum model that converged are reported here. If the effect of breathy voice is different across age groups, we would find a significant interaction between Phonation and Age group. Similarly, a significant interaction between Phonation and Voice would indicate differences of the effect of breathy voice between older and middle-aged voices. If the effect of breathy voice is larger for the younger/middle-aged listeners than for the older listeners (i.e., the stimuli with breathy voice induce more low register tone responses), the estimate would be positive, and negative otherwise.

For the stimuli with stop onsets, Phonation (Estimate = −0. 72, SE = 0.07, z = −10.87, p < 0.001) and Step (Estimate = 1.62, SE = 0.06, z = 0.20, p < 0.001) were significant predictors, but not any other factors. Specifically, for Age group, there was no significant difference between the older and the mean of the middle-aged and younger listeners (Estimate = 0.16, SE = 0.30, z = 0.52, p = 0.60), and also no significant difference between the younger and middle-aged listeners (Estimate = 0.16, SE = 0.30, z = 0.52, p = 0.60). For the factor Talker voice, there was also no significant difference between the older and middle-aged voices (Estimate = −0.12, SE = 0.28, z = −0.41, p = 0.68). Therefore, for the unchecked tone pairs with stop onsets, the breathy voice had a similar effect across age groups, and the talker voice of the stimuli did not influence the listeners' response.

The results of the stimuli with fricative onsets were slightly different. In addition to Phonation and Step, the younger listeners had significantly more low register tone responses than the middle-aged listeners did (Estimate = 0.76, SE = 0.33, z = 2.32, p < 0.05). Moreover, the interactions between Phonation and Age group and Voice were also significant (Estimate = 0.33, SE = 0.13, z = 2.43, p < 0.015 for the interaction between Phonation and the middle-aged listeners, Estimate = −0.12, SE = 0.06, z = −2.00, p < 0.05 for the interaction between Phonation and the middle-aged Voice). To better illustrate the effects of both age group and talker voice, interaction plots of these factors for both stop and fricative onsets were made using the R package emmeans, as shown in Fig. 4. The y-axis is the predicted log odds ratio of a low tone response against a high tone response. It is clear that when hearing the stimuli with breathy voice the listeners were more likely to give a low tone response than hearing the stimuli with modal voice. For the stimuli with stop onset, the effect of the older voice was similar to the middle-aged voice, while middle-aged and younger listeners were sensitive to the difference between older and middle-aged voices for the stimuli with fricative onset. Older voice with breathy voice was more likely to induce a low tone response than middle-age voice was.

Thus, for the unchecked tone pairs, the effect of breathy voice was similar (for stop onsets) and even larger (for fricative onsets) for the younger listeners than for the older listeners. The voice of the stimuli also had an effect for the stimuli with fricative onset, in that the effect of breathy voice was larger for the older voice than for the middle-aged voice.

3.1.2. The checked tone pairs

The identification curves of the checked tone pairs are shown in a similar manner to the unchecked tone pairs in Fig. 5. The results were in general similar to those for the unchecked tone pairs, but the stimuli with fricative onsets showed notably different results. In what follows, the results for the stimuli with stop and fricative onsets will be presented in turn, using similar statistical analyses to those for the unchecked tone pairs.

First of all, it should be noted that although a merging pair was used for the checked tone pair with stop onsets, the results were similar with those of the unchecked pair with stop onsets. This indicates that the results of this pair are still valid and listeners did rely mainly on phonation and pitch cues to identify T7 and T8. For the stimuli with stop onsets,

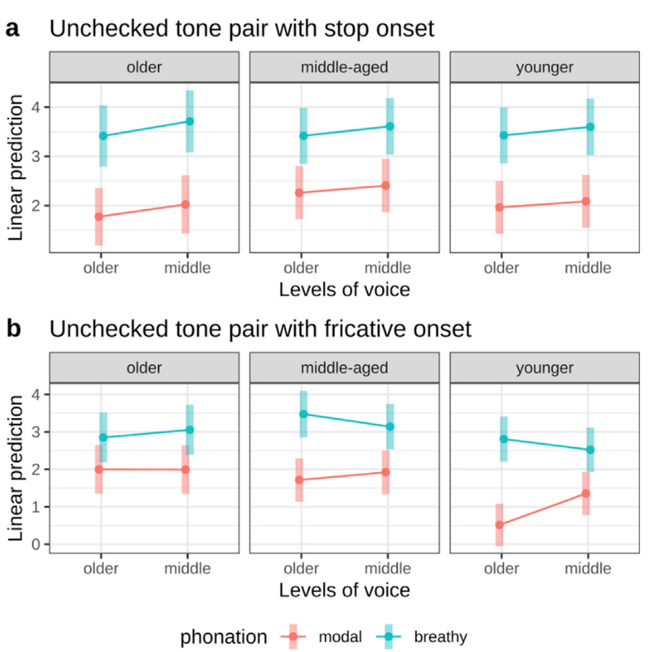

Fig. 4. Interaction plots of the unchecked tone pairs with stop (a) and fricative (b) onsets, with the vertical bars showing 95% confidence interval.

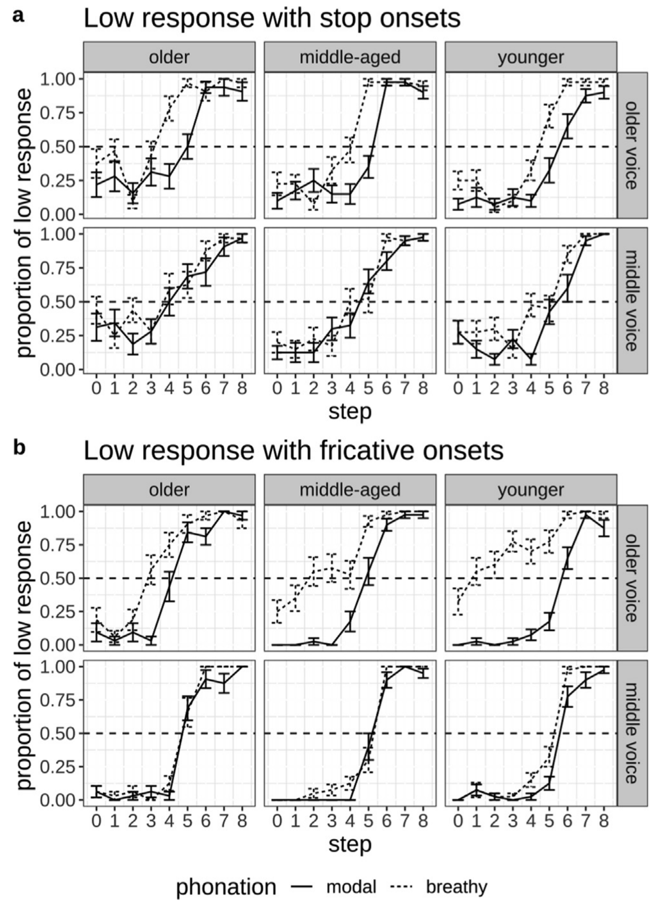

Fig. 5. The identification curves of the checked tone pairs with stop (a) and fricative (b) onsets.

there was also a significant difference between the mean of the middle-aged and younger listeners and that of the older listeners (Estimate = 0.74, SE = 0.34, z = 2.18, p < 0.05), while the difference between the middle-aged and younger listeners was not significant (Estimate = 0.41, SE = 0.36, z = 1.12, p = 0.26). The interaction between Phonation and Age group was not significant for both the younger (Estimate = 0.18, SE = 0.10, z = 1.74, p = 0.08) and the middle-aged listeners (Estimate = −0.02, SE = 0.10, z = −0.26, p = 0.80). Talker voice did not have a significant effect (Estimate = −0.03, SE = 0.04, z = −0.67, p = 0.51), but its interaction with Phonation was significant (Estimate = −0.16, SE = 0.04, z = −3.59, p < 0.001). Therefore, breathy voice also had a similar effect for the middle-aged and the younger age groups for the checked tone pairs with stop onsets. Furthermore, while the middle-aged voice generally had a greater proportion of low register tone (T8) responses than the older voice did, when the stimuli were with breathy voice, the low register tone responses were considerably fewer for the middle-aged voice than for the older voice, as the dashed line in the top panel (the older voice) is higher than the dashed line in the lower panel (the middle-aged voice). Thus, whilst the effect of breathy voice was similar for different age groups, the effect of breathy voice was smaller for the middle-aged voice than for the older voice. This can also be seen from the interaction plots in Fig. 6(a). For breathy voice (the blue line), the older

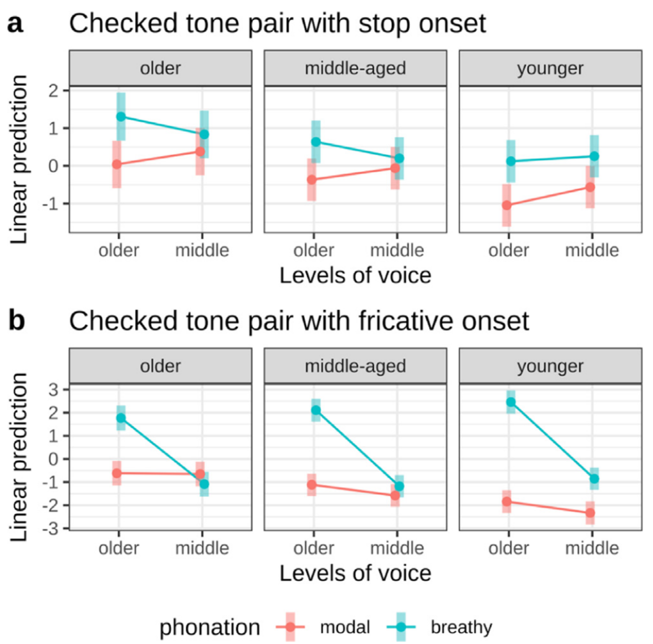

Fig. 6. Interaction plots of the checked tone pairs with stop (a) and fricative (b) onsets, with the vertical bars showing 95% confidence interval.

voice had a larger effect than the middle-aged voice for both the older and the middle-aged listeners, and the two voices had similar effects for the younger listeners.

The effects of breathy voice were similar for the checked tone pairs with fricative onsets to those with stop onsets. The stimuli with breathy voice induced more low register tone responses (T8) for both the mean of the middle-aged and the younger listeners than the older listeners (Estimate = 0.69, SE = 0.12, z = 5.74, p < 0.001) as well as the younger listeners than the middle-aged listeners (Estimate = 0.54, SE = 0.13, z = 4.21, p < 0.001) and it induced fewer low register tone responses for the middle-aged voice than for the older voice (Estimate = −0.71, SE = 0.08, z = −12.29, p < 0.001). Both effects are also clear from Fig. 5(b). However, the stimuli with fricative onsets were different from those with stop onsets in that there is an interaction effect between the phonation effect and the talker voice effect. On the one hand, while the stimuli with modal voice have similar identification curves (the solid line) across age groups for both the older and the middle-aged voices, the identification curves of the stimuli with breathy voice (the dashed line) are higher for the older than for the middle-aged voice and also higher for the younger and the middle-aged listeners than for the older listeners. The effects of age group and talker voice can also be seen from the interaction plots in Fig. 6(b), which plots the predicted log odds ratio of low tone response. Breathy voice induced more low tone response than modal voice did and the effect was larger for the older voice than for the middle-aged voice, which holds for all three listener groups.

Although there were some notable differences between the checked tone pairs and the unchecked tone pairs, the results were generally consistent. The effect of breathy voice was similar for the younger and the older listeners, and it was larger for the older voice than for the middle-aged voice.

3.2. Response time (RT)

Because the effect of breathy voice is quite subtle, it may also influence RT. Therefore, RT was also analyzed to see whether the effect of breathy voice differed across age groups. The RTs were log-transformed with the natural base. Because older listeners naturally have longer RTs than middle-aged and younger listeners, the RT of the middle-aged listeners was chosen as the reference level. Dummy coding was used here instead of Helmert coding, because we wanted to compare the RTs of the older and the younger listeners to those of the middle-aged listeners.

As the identification is categorical, and the response near the category boundary is ambiguous, we followed Gao et al. (2020) and selected response data from the steps in which there were the most responses of either high or low register tones. For the unchecked tone pairs, the category boundary was near step 2, and the RTs of the high register tone responses in steps 0–1 (high pitch end) and the low register tone responses in 4–7 (low pitch end) were selected. Similarly, the RT of the high register tone responses in steps 0–3 (high pitch end) and the low register tone responses in 6–8 (low pitch end) were selected for the checked tone pairs. At the high pitch end, a high register tone response (T1/T7) with modal voiced stimuli was called congruent, and a T1/T7 response with breathy voiced stimuli was called incongruent. Likewise, at the low pitch end, a low register tone response (T2/T8) with modal voiced stimuli was incongruent, and a T2/T8 response with breathy voiced stimuli was congruent. The relationship is summarized in Table 4.

Fig. 7 shows the mean RT of the congruent and incongruent conditions at the high and low pitch ends by talker voice, with the error bars showing standard errors. The RT (with log transformation) was analyzed using linear mixed-effects models with the relation between response and phonation of the stimuli (Condition hereafter, congruent vs. incongruent), Age group (middle-aged vs. younger vs. older), End of the pitch continuum (high vs. low) and their interactions as fixed effects. Model selection followed the two-stage procedure proposed by Lin et al. (2013). In the first stage, models of different random effect structures were fitted using restricted maximum likelihood and were compared using anova based on likelihood ratio. Random effects which did not significantly improve model fit were removed and those which did were included. The second stage is to fit models which include different fixed effects using maximum likelihood. Fixed effects and interactions were step-wisely added to the base model and model comparison was also done using the anova function. The final models of both the unchecked and checked tone pairs included participant as random intercepts, and also by-participant random slopes of Condition and End. The formulae of the final models were included in Tables B5 and B6. The

Table 4
Summary of the relation between response and phonation of the stimuli (condition) at the two ends of the pitch continuum.

Pitch end	Response	Phonation	Condition
High	T1/T7	Modal voice	Congruent
	T1/T7	Breathy voice	Incongruent
Low	T2/T8	Modal voice	Incongruent
	T2/T8	Breathy voice	Congruent

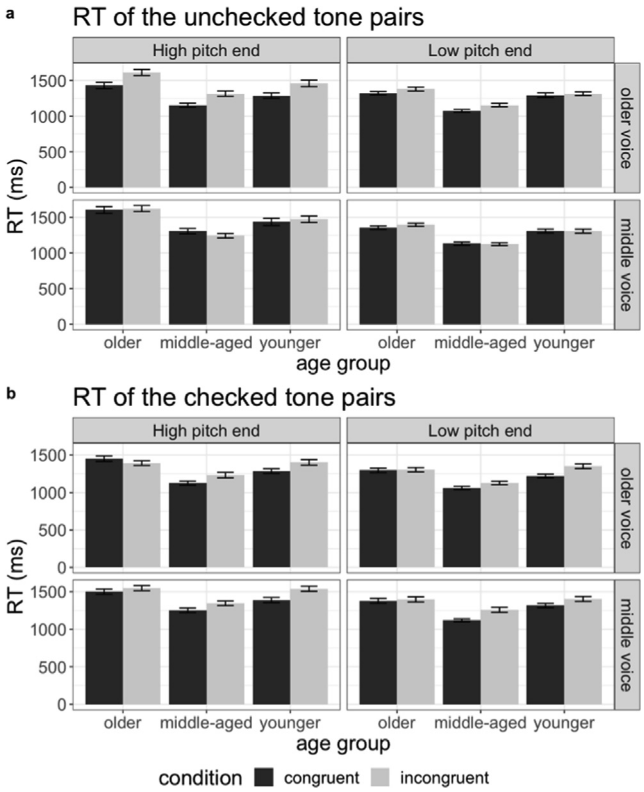

Fig. 7. RT in congruent and incongruent conditions of the unchecked (a) and checked (b) tone pairs pooled by pitch continuum end and talker voice.

coefficients and *p*-values of the final models were calculated using the lmerTest package in R. Unchecked and checked tone pairs were analyzed separately. For the unchecked tone pairs, the older listeners (Estimate = 0.22, t = 3.61, p < 0.01) and the younger listeners (Estimate = 0.10, t = 1.76, p = 0.08) both had longer RTs than the middle-aged listeners. That is to say, both the older and younger listeners were in general slower than the middle-aged listeners. For the middle-aged listeners, the incongruent condition had a significantly longer RT than the congruent condition (Estimate = 0.04, t = 4.48, p < 0.01). However, the inclusion of the interaction between Age group and the relation between response and the stimuli's phonation did not significantly improve the model fit ($\chi^2(2)$ = 0.09, p = 0.95). This means that when a listener listened to the stimuli in which the phonation did not match his/her expectations, he/she would respond slower than when the phonation matched expectations, but this delay was similar for the middle-aged, older and younger listeners. In other words, the effect of breathy voice was similar for listeners of the three age groups. Talker voice also had a significant effect on RT, in that the stimuli with the middle-aged voice (Estimate = −0.03, t = −4.37, p < 0.01) induced a slightly shorter RT than did the older voice. The interaction between Condition and pitch continuum End was also significant (Estimate = −0.04, t = −2.81, p < 0.01). At the low pitch end, the incongruent condition had a shorter RT than at the high pitch end. Therefore, breathy voice caused some confusion and delayed the response at the high pitch end.

The same analyses were also conducted for the checked tone pairs. Similar to the unchecked tone pairs, both the older (Estimate = 0.21, t = 3.86, p < 0.01) and the younger listeners (Estimate = 0.14, t = 2.66, p = 0.01) had a longer RT than the middle-aged listeners. Also, the incongruent condition also had a longer RT than the congruent condition (Estimate = 0.07, t = 4.24, p < 0.01) for the middle-aged listeners and the effect was almost identical for the younger listeners (Estimate = −0.01, t = −0.27, p = 0.79) and the older listeners (Estimate = 0.00, t = 0.12, p = 0.91). Simply put, the effect of breathy voice on RT was smaller for the older and younger listeners than the middle-aged listeners. It can be seen from Fig. 5(b) that there is barely any difference between the congruent and incongruent conditions for the older and younger listeners. The middle-aged voice also induced a slightly shorter RT than the older voice (Estimate = −0.01, t = −3.8, p < 0.01). The End of the pitch continuum does not interact with Condition, in that the inclusion of their interaction did not significantly improve the model fit ($\chi^2(1)$ = 0.70, p = 0.40).

To summarize, the RT shows consistent results with the identification data. The younger listeners in general responded more slowly than the middle-aged listeners did, indicating their lower proficiency in Suzhou Wu. But the effect of breathy voice had a similar effect on the speed of the younger and middle-aged listeners' responses. Moreover, the older voice sped up the listeners' responses compared to the middle-aged voice.

4. Discussion

4.1. Summary of results

This study aimed to answer the question of whether younger listeners of Suzhou Wu used breathy voice in their perception despite its decrease in their production, and whether the effect of breathy voice was the same for younger and older/middle-aged listeners. The identification experiment based on resynthesized stimuli showed that the younger listeners still used breathy voice in their tone identification, and the stimuli with breathy voice could induce more low tone response as compared to the stimuli with modal voice. Also, breathy voice could influence how fast they responded to the stimuli. They responded faster to the stimuli with breathy voice at the low pitch end, but slower to the breathy voiced stimuli at the high pitch end. For the stop onset, the shift of identification by breathy voice for the younger listeners was similar to that for the middle-aged and older listeners, and the effect of breathy voice on RT was also similar for the younger listeners and the middle-aged listeners. But the shift of identification by breathy voice was larger for younger listeners than it was for older listeners on the checked tone pairs, which is especially true for the fricative onset. Therefore, we can conclude that the younger listeners still use breathy voice in their perception and the effect of breathy voice is similar for the younger and at least the middle-aged listeners.

Another question this study asked is whether or not the older and middle-aged voices would induce the same effect of breathy voice, and it was hypothesized that the effect would be larger for the older voice. To answer this question, natural tokens produced by an older speaker and a middle-aged speaker were used as the base syllables. The results confirmed the prediction. The shift of identification by breathy voice was larger for the older voice than for the middle-aged voice, and the older voice sped up listeners' responses compared to the middle-aged voice.

Stop and fricative onsets were also used and there were some differences between the two kinds of onset. The stimuli produced with fricative onsets and breathy voice were perceived less categorically than those with modal voice. This effect was limited to the older voice and to the middle-aged and younger listeners. It is unclear why there was such an effect. It was perhaps due to the fact that the perception of stops is intrinsically more categorical than that of fricatives (Healy & Repp, 1982; Liberman et al., 1967; Pinget et al., 2020, p. 678). Inspection of our recordings showed that there was barely any voicing during the frication of fricative onsets for all the older and middle-aged speakers in the production experiment, and the base syllables for /z/ that we used were also totally voiceless. Although the durations of onsets were not manipulated, the durations of the high and low tone syllables were similar, as shown in Table A2 in Appendix A. However, duration can be a potential cue and voiced fricatives are approximately 60 ms shorter than voiceless fricatives in Shanghai Wu, as reported by Zhang and Yan (2018). But as the difference in most of our stimuli is much smaller than that (less than 5 ms, only one pair with ~50 ms difference), the effect of duration should be negligible. Nevertheless, it is possible that the older speakers produced the fricative onsets in breathy voiced syllables with some subtle characteristics which were perceptually clear to the native listeners, and further study is needed to investigate these subtle characteristics.

4.2. The effect of talker voice

The results also indicated a larger effect of breathy voice for the older voice than for the middle-aged voice under some conditions. It is likely that talker voice as an indexical cue is integrated with other linguistic cues and listeners use the indexical cues to guide their perception of linguistic distinctions. Another possibility is that the larger effect of breathy voice on perception is due to the generally breathier voice of older speakers compared to middle-aged speakers. As pointed out in Section 2.2, the older voice is indeed breathier than the middle-aged voice, and even for high register tones, the older voice also has larger spectral tilts than the middle-aged voice (see Tables A3–A6). Hirano et al. (1988) found that older women usually have larger posterior glottal opening, indicating breathier voice. Stathopoulos et al. (2011) supported this finding with acoustic measurements using data from a large age range. Therefore, the generally breathier voice in our older voice can be attributed to the aging effect. However, the differences between the high and low register tones in phonation are in general similar between the two voices. Even in the case that older speakers are in general breathier than middle-aged speakers, it would not contradict the effect of older voice in general because talker voice can include a vast variety of information, and how speakers produce speech is also included in our representation of talker voice (Pierrehumbert, 2016; Nygaard & Tzeng, 2021). Compelling evidence is provided by talker identification in sinewave speech (Sheffert et al., 2002) for the idea that the same acoustic attributes are used for both speech perception and talker identification. Sheffert et al. (2002) trained listeners to identify different talkers using natural, sinewave, and reversed speech and they found that listeners could identify the talkers' voices even from sinewave speech in which all natural vocal quality cues had been removed and only some phonetic and suprasegmental properties were retained. Therefore, it is highly likely that our listeners were sensitive to the talker voice difference and used this information to guide their perception of breathy voice.

The effect of talker voice on how breathy voice influences tone perception may have some implications for the process of the loss of breathy voice as a secondary cue. Because younger listeners adapt their perception to a talker's voice, and in turn modulate the effect of breathy voice in perception, when the voices of older talkers become less and less encountered in the speech community, the effect of breathy voice in perception will also decrease. In the course of time, when the younger listeners encounter only the voices of their peers (the present younger voices) and no longer those of even the present middle-aged talkers, the effect of breathy voice on their perception would decrease even further. But this is only a speculation, as the voices of younger talkers were not tested in the current study. Yet a possible account for the loss of a secondary cue in speech perception is that the secondary cue of a linguistic distinction could take on indexical meanings (e.g., age), and when the indexical cue is encountered less and less, the importance of the secondary cue also gradually decreases in speech perception (see also Eckert, 2019 for a model of sound change in the semiotic landscape).

4.3. Does the loss of breathy voice originate from production or perception?

Besides the effect of age group and talker voice, the results could shed light on the origin of the loss of breathy voice, i.e., whether the loss originates from production or perception. The results showed that younger listeners in Suzhou Wu still use breathy voice in their perception and their use of breathy voice is in general similar to that of middle-aged and older listeners. If younger speakers produce less breathy voice because they use it less in their perception, they would be less sensitive to breathy voice than older and middle-aged listeners are in perception, which is contrary to the findings in this study. The identification results showed that the younger listeners were sensitive to breathy voice, and their sensitivity was similar to that of the older and middle-aged listeners, and under some conditions they had even higher sensitivity. The results are more in line with the hypothesis that the loss of breathy voice originates in production. Younger speakers produce less breathy voice but keep their sensitivity to breathy voice in perception. Therefore, the decrease in breathy voice is more likely to be driven by production than by perception. Breathy voice is still used by younger listeners as a secondary cue in their perception, but they produce less breathy voice than older and middle-aged speakers (see Gao et al. (2020) for similar findings in Shanghainese).

The reason why younger speakers produce less breathy voice in the first place is harder to answer, however. Although the present study does not provide direct answers, some speculation can be made (see also our discussion in Ge et al. (2023). There are two main possibilities. The first is that breathy voice is a secondary or concomitant cue in the transphonologization from the onset voicing contrast to the tone contrast and when f0 becomes more dominant, breathy voice becomes less and less important. Another possibility is that frequent contact with another tone language (Standard Chinese) which does not employ breathy voice influences younger speakers' production, as they are all bilingual in Suzhou Wu and Standard Chinese. The answer to this question requires further study on Wu Chinese as well as on other languages with tone split or tonogenesis induced by consonant voicing.

4.4. Limitations

Although this study had some interesting findings, there were also some limitations. In constructing the stimuli, only natural speech tokens were used, and no synthesized tokens were included. This was because we wanted to manipulate talker voices, and as mentioned above, talker voice can include a vast array of cues, which are hard to finely control. Nevertheless, synthesized stimuli can provide some advantages over resynthesized stimuli based on naturally produced speech. Because phonation can be more finely controlled in synthesized stimuli, it can be used to test the effect of different degrees of breathy voice on tone identification without using different voices from different speakers. This approach can be employed to test listeners' sensitivity to degrees of breathy voice, especially for younger listeners. Moreover, only two talker voices were used which have limited the understanding of the effects of talker age and talk voice. Also, as acknowledged in Section 2.2, one vowel-merging pair (the checked tone pair with stop onsets) was used in this study due to logistic constraints. However, there was little acoustic difference in the vowels, and that the listeners responded similarly to the checked tone pair with stop onsets and the unchecked tone pair with stop onsets, suggesting that the ongoing vowel merger had not influenced the perception of the participants in this study. Future studies should have more sophisticated control on the stimuli and thorough studies are needed to investigate the influence of such a vowel merger on tone perception.

Another limitation is that this study did not investigate individual differences of perception and the relationship between each individual's production and perception. The reason for this is that due to travel restrictions, we were unable to revisit all the speakers in the production study to do the perception experiment (and unfortunately, one of our older speakers even passed away during the hiatus between the production and perception experiments). Recent studies have shown that individuals' perception is linked to their own production (e.g., Beddor et al. (2018) on the perception of coarticulatory nasalization and Pinget et al. (2020) on the perception of devoicing of Dutch obstruent) and individual differences in cognitive processing style (Yu, 2016), and phonetic imitation (Pinget, 2022) can have a critical influence on sound change (see also Stevens and Harrington (2014) and Yu and Zellou (2019) for comprehensive reviews). Our production study has shown that younger female speakers are more advanced than younger male speakers in the decrease in breathy voice, and it is possible that other factors of individual variation may also be at play. It will be a fruitful area for future research.

5. Conclusion

In the tone split of Northern Wu Chinese, breathy voice is disappearing and our previous studies found that younger speakers of Suzhou Wu also produce less breathy voice than older and middle-aged speakers. This study furthers our work by investigating the effect of breathy voice on younger listeners' tone identification, compared to that of older and middle-aged listeners. The results showed that younger listeners are still sensitive to breathy voice and their sensitivity is similar to older and middle-aged listeners. There was also some influence of talker voice on the effect of breathy voice. The effect was larger for the older voice than for the middle-aged voice. This study also discusses the implications of this influence for the mechanism of the loss of breathy voice as a secondary cue. Some limitations of this study have also been noted. We hope that more sophisticated experiments can be conducted on more diverse sound change patterns to deepen our understanding of the mechanisms of sound change with regard to speech production and perception.

Conflict of Interest

We have no conflict of interest to declare.

CRediT authorship contribution statement

Chunyu Ge: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Peggy Mok:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization.

Acknowledgements

This study was supported by the RGC General Research Fund 2019/20 Project Reference 14607619 awarded to the second author and the CUHK Research Fellowship Scheme 2019-20, 2020-21 awarded to the first author. We would like to thank Jingyu You for his assistance in recruiting participants and running the experiment in Suzhou. We are also grateful to the three anonymous reviewers for their valuable comments and suggestions.

Appendix A. . Acoustic measurements of the base syllables

Table A1
The value of the first two formants of the base syllables /paʔ/ 百 "hundred" and /baʔ/ 拔 "pull" (in Hz).

Voice	Base syllable	F1	F2
Older	paʔ7 百"hundred"	1029	1576
	baʔ8 拔 "pull"	1034	1527
Middle-aged	paʔ7 百"hundred"	1064	1611
	baʔ8 拔 "pull"	1182	1613

Table A2
The duration of the onsets of the base syllables (in millisecond).

Voice	Onset	Unchecked pair	Checked pair
Older	p	11	10
	b	10	9
	s	199	205
	z	194	202
Middle-aged	p	11	10
	b	14	10
	s	267	255
	z	209	253

Table A3
H1*–H2* of the base syllables (in dB).

Tone pairs	Voice	Onset	High register	Low register
Unchecked	Middle-aged	Fricative	1.20	4.07
		Stop	1.87	4.94
	Older	Fricative	8.50	9.38
		Stop	5.28	6.95
Checked	Middle-aged	Fricative	−0.70	4.38
		Stop	0.20	1.40
	Older	Fricative	7.24	9.75
		Stop	6.14	11.17

Table A4
H1*–A1* of the base syllables (in dB).

Tone pairs	Voice	Onset	High register	Low register
Unchecked	Middle-aged	Fricative	13.43	23.53
		Stop	14.21	23.59
	Older	Fricative	30.87	37.78
		Stop	20.35	31.56
Checked	Middle-aged	Fricative	14.03	20.33
		Stop	13.53	14.94
	Older	Fricative	23.53	30.45
		Stop	27.04	35.92

Table A5
H1*–A2* of the base syllables (in dB).

Tone pairs	Voice	Onset	High register	Low register
Unchecked	Middle-aged	Fricative	12.69	25.35
		Stop	11.32	25.71
	Older	Fricative	34.97	38.78
		Stop	9.10	30.37
Checked	Middle-aged	Fricative	7.22	22.58
		Stop	6.47	13.92
	Older	Fricative	17.27	27.29
		Stop	27.48	33.35

Table A6
H1*–A3* of the base syllables (in dB).

Tone pairs	Voice	Onset	High register	Low register
Unchecked	Middle-aged	Fricative	0.57	16.65
		Stop	−1.63	12.74
	Older	Fricative	16.32	20.59
		Stop	−5.57	14.12
Checked	Middle-aged	Fricative	7.20	23.50
		Stop	−5.25	12.47
	Older	Fricative	9.98	16.68
		Stop	12.95	23.94

Appendix B. Summary tables of the mixed-effects models of identification data and RT

Table B1
Summary table of the unchecked tone pairs with stop onsets (formula: lowResponse ~ phonation*(voice + ageGroup) + voice*ageGroup + step + (1|participant)).

	Estimate	Std.Error	z	Pr
Intercept	−2.87	0.17	−16.82	<0.001***
phonationBreathy	−0.72	0.07	−10.87	<0.001***
voiceMiddle	−0.10	0.06	−1.59	0.11
ageGroupMiddle	−0.12	0.28	−0.41	0.68
ageGroupYounger	0.16	0.30	0.52	0.60
step	1.62	0.06	26.47	<0.001***
phonationBreathy:voiceMiddle	0.01	0.06	0.20	0.84
phonationBreathy:ageGroupMiddle	−0.17	0.14	−1.21	0.22
phonationBreathy:ageGroupYounger	0.15	0.15	1.06	0.29
voiceMiddle:ageGroupMiddle	−0.06	0.14	−0.42	0.68
voiceMiddle:ageGroupYounger	−0.01	0.15	−0.07	0.94

Table B2
Summary table of the unchecked tone pairs with fricative onsets (formula: lowResponse ~ phonation*voice*ageGroup + step + (1|particpant)).

	Estimate	Std.Error	z	Pr
Intercept	−3.25	0.19	−17.46	<0.001***
phonationBreathy	−0.70	0.07	−10.71	<0.001***
voiceMiddle	−0.05	0.06	−0.85	0.40
ageGroupMiddle	0.29	0.31	0.95	0.34
ageGroupYounger	0.76	0.33	2.32	<0.05*
step	1.58	0.06	27.27	<0.001***
phonationBreathy:voiceMiddle	−0.12	0.06	−2.00	0.05
phonationBreathy:ageGruopMiddle	0.33	0.13	2.43	<0.05*
phonationbreathy:ageGroupYounger	0.12	0.14	0.83	0.40
voiceMiddle:ageGroupMiddle	0.00	0.13	0.02	0.99
voiceMiddle:ageGroupYounger	0.17	0.14	1.19	0.23
phonationBreathy:voiceMiddle:ageGroupMiddle	0.26	0.13	1.95	0.05
phonationBreathy:voiceMiddle:ageGroupYounger	0.15	0.14	1.01	0.31

Table B3
Summary table of the checked tone pairs with stop onsets (formula: lowResponse ~ phonation*voice*ageGroup + step + (1|particpant)).

	Estimate	STd.Error	z	Pr
Intercept	−2.87	0.18	−15.9	<0.001***
phonationBreathy	−0.41	0.05	−9.21	<0.001***
voiceMiddle	−0.03	0.04	−0.66	0.51
ageGroupMiddle	0.74	0.34	2.17	0.03
ageGroupYounger	0.41	0.36	1.12	0.26
step	0.75	0.02	31.83	<0.001***
phonationBreathy:voiceMiddle	−0.16	0.04	−3.59	<0.001***
phonationBreathy:ageGroupMiddle	−0.02	0.10	−0.26	0.80
phonationBreathy:ageGroupYounger	0.18	0.10	1.74	0.08
voiceMiddle:ageGroupMiddle	0.09	0.10	0.96	0.34
voiceMiddle:ageGroupYounger	0.19	0.10	1.81	0.07
phonationBreathy:voiceMiddle:ageGroupMiddle	−0.07	0.10	−0.68	0.49
phonationBreathy:voiceMiddle:ageGroupYounger	−0.10	0.10	−0.94	0.35

Table B4
Summary table of the checked tone pairs with fricative onsets (formula: lowResponse ~ phonation*(voice + ageGroup) + voice*ageGroup + step + (1|participant)).

	Estimate	Std.Error	z	Pr
Intercept	−5.38	0.21	−25.42	<0.001***
phonationBreathy	−0.94	0.06	−15.56	<0.001***
voiceMiddle	0.87	0.06	14.57	<0.001***
ageGroupMiddle	0.39	0.27	1.48	0.14
ageGroupYounger	0.20	0.28	0.70	0.48
step	1.24	0.04	30.86	<0.001***
phonationBreathy:voiceMiddle	−0.71	0.06	−12.29	<0.001***
phonationBreathy:ageGroupMiddle	0.69	0.12	5.74	<0.001***
phonationBreathy:ageGroupYounger	0.54	0.13	4.21	<0.001***
voiceMiddle:ageGroupMiddle	−0.22	0.12	−1.87	0.06
voiceMiddle:ageGroupYounger	−0.01	0.13	−0.11	0.91

Table B5
Summary table of RT for the unchecked tone pairs (formula: logRT ~ end*ageGroup + ageGroup*condition + voice*end + (condition + end | participant)).

	Estimate	Std.Error	t	df	Pr
Intercept	7.08	0.04	177.83	53	<0.001***
ageGroupOlder	0.22	0.06	3.61	53	<0.001***
ageGroupyounger	0.10	0.06	1.76	53	0.08
conditionIncongruent	0.04	0.01	2.99	51	<0.05
voiceMiddle	−0.03	0.01	−4.37	2520	<0.001***
endLEnd	−0.10	0.01	−6.62	4965	<0.001***
conditionIncongruent:endLEnd	−0.04	0.02	−2.81	4987	<0.05*
ageGroupOlder:conditionIncongruent	0.00	0.02	0.15	55	0.88
ageGroupYounger:conditionIncongruent	0.00	0.02	−0.17	55	0.87
ageGroupOlder:endLEnd	−0.01	0.02	−0.52	4972	0.60
ageGroupYounger:endLEnd	0.04	0.02	2.24	4979	<0.05*
voiceMiddle:endLEnd	0.02	0.01	2.75	4956	<0.05*

Table B6
Summary table of RT for the checked tone pairs (formula: logRT ∼ ageGroup*condition + voice + end*condition + (condition + end| participant)).

	Estimate	Std.Error	t	df	Pr
(Intercept)	7.05	0.04	186.99	60	<0.001***
ageGroupOlder	0.21	0.05	3.86	53	<0.001***
ageGroupYounger	0.14	0.05	2.66	53	0.01
conditionIncongruent	0.08	0.02	4.24	118	<0.001***
voiceMiddle	−0.01	0.00	−3.8	4896	<0.001***
endLEnd	−0.09	0.01	−6.42	103	<0.001***
conditionIncongruent:endLEnd	−0.05	0.02	−3.01	4944	<0.05**
ageGroupOlder:conditionIncongruent	0.00	0.02	0.12	50	0.91
ageGroupYounger:conditionIncongruent	−0.01	0.02	−0.27	50	0.79

References

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.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language, 85*(4), 785–821. https://doi.org/10.1353/lan.0.0165.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language, 94*(4), 931–968. https://doi.org/10.1353/lan.2018.0051.

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

Brunelle, M., Brown, J., & Hà, P. T. T. (2022). Northern Raglai voicing and its relation to Southern Raglai register: Evidence for early stages of registrogenesis. *Phonetica, 79*(2), 151–188. 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.

Cao, J., & Maddieson, I. (1992). An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics, 20*, 77–92.

Chao, Y. R. (1928). *Studies in the modern Wu dialects*. Tsinghua University.

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

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

Chen, Z. (2015). Breathy voice and low tone. *Journal of Chinese Linguistics, 43*(1A), 90–118.

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 Decker, P. (2010). Sounds shifty: Gender and age differences in perceptual categorization during a phonetic change in progress. *University of Pennsylvania Working Papers Linguistics, 15*(2), 7.

Eckert, P. (2019). The individual in the semiotic landscape. *Glossa: A Journal of General Linguistics, 4*(1). https://doi.org/10.5334/gjgl.640 1.

Esposito, C. M., & Khan, S. ud. D. (2020). The cross-linguistic patterns of phonation types. *Language and Linguistics Compass, 14*(12), e12392.

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., & Draxler, C. (2020). Breathy voice and low-register: A case of trading relation in Shanghai Chinese tone perception? *Language and Speech, 63*(3), 582–607. https://doi.org/10.1177/0023830919873080.

Gao, X., & Kuang, J. (2022). Phonation variation as a function of checked syllables and prosodic boundaries. *Languages, 7*(3). https://doi.org/10.3390/languages7030171 3.

Garellek, M. (2022). Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics, 94*. https://doi.org/10.1016/j.wocn.2022.101155 101155.

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.

Ge, C., Xu, W., Gu, W., & Mok, P. (2022). An electroglottographic study of phonation types in tones of Suzhou Wu Chinese. *Speech Prosody, 2022*, 570–574. https://doi.org/10.21437/SpeechProsody.2022-116.

Ge, C., & Mok, P. (2023). The phonetic realization of intervocalic obstruents in Suzhou Wu Chinese. *Proceedings of International Congress of Phonetic Sciences*, 773–777.

Ge, C., Xu, W., Gu, W., & Mok, P. P. K. (2023). The change in breathy voice after tone split: A production study of Suzhou Wu Chinese. *Journal of Phonetics, 98*. https://doi.org/10.1016/j.wocn.2023.101239 101239.

Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics, 29*(4), 383–406. https://doi.org/10.1006/jpho.2001.0147.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34*(4), 458–484. https://doi.org/10.1016/j.wocn.2005.10.001.

Healy, A. F., & Repp, B. H. (1982). Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance, 8*(1), 68–80. https://doi.org/10.1037/0096-1523.8.1.68.

Hirano, M., Kiyokawa, K., & Kurita, S. (1988). Laryngeal muscles and glottal shaping. In O. Fujimura (Ed.), *Vocal physiology: Voice production, mechanisms, and functions* (pp. 249–264). Raven Press.

Jiang, B., Clayards, M., & Sonderegger, M. (2020). Individual and dialect differences in perceiving multiple cues: A tonal register contrast in two Chinese Wu dialects. *Laboratory Phonology: Journal of the Association for Laboratory Phonology, 11*(1), 1–30. https://doi.org/10.5334/labphon.266.

Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics, 45*, 76–90. https://doi.org/10.1016/j.wocn.2014.03.005.

Kettig, T., & Winter, B. (2017). Producing and perceiving the Canadian Vowel Shift: Evidence from a Montreal community. *Language Variation and Change, 29*(1), 79–100. https://doi.org/10.1017/S0954394517000023. Scopus.

Kim, J., Gabriel, U., & Gygax, P. (2019). Testing the effectiveness of the Internet-based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. *PLOS ONE, 14*(9), e0221802. https://doi.org/10.1371/journal.pone.0221802.

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.

Kleber, F., Harrington, J., & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech, 55*(3), 383–405. https://doi.org/10.1177/0023830911422194.

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., & 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.

Kuang, J., & Liberman, M. (2018). Integrating voice quality cues in the pitch perception of speech and non-speech utterances. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.02147.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13.

Kingston, J. (2011). Tonogenesis. In *The Blackwell companion to phonology* (pp. 1–30). Ltd: John Wiley & Sons. https://doi.org/10.1002/9781444335262.wbctp0097.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431–461. https://doi.org/10.1037/h0020279.

Lin, B., Pang, Z., & Jiang, J. (2013). Fixed and random effects selection by REML and pathwise coordinate optimization. *Journal of Computational and Graphical Statistics: A Joint Publication of American Statistical Association, Institute of Mathematical*

Statistics, Interface Foundation of North America, 22(2), 341–355. https://doi.org/10.1080/10618600.2012.681219.

Ling, F. (2009). *A phonetic study of the vowel system in Suzhou Chinese* [Doctoral thesis, City University of Hong Kong]. https://scholars.cityu.edu.hk/en/theses/theses(52b089ed-b109-4b7f-968d-5e9a96e6f9b1).html.

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.

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.

Nygaard, L. C., & Tzeng, C. Y. (2021). Perceptual integration of linguistic and non-linguistic properties of speech. In *The handbook of speech perception* (pp. 398–427). Ltd: John Wiley & Sons. https://doi.org/10.1002/9781119184096.ch15.

Ohala, J. J. (1981). The listener as a source of sound change. In C. S. Masek, R. A. Hendrick, & M. F. Miller (Eds.), *Papers from the Parasession on language and behavior* (pp. 178–203). Chicago Linguistics Society.

Ohala, J. J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (Eds.), *Language change: Contributions to the study of its causes* (pp. 173–198). Mouton de Gruyter.

Pharao, N., Appel, L., Wolter, K., & Thøgersen, J. (2015). Raising of /a/ in Copenhagen Danish – Perceptual consequences across two generations. *Proceedings of the international congress of phonetic sciences*.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics, 2*(1), 33–52. https://doi.org/10.1146/annurev-linguistics-030514-125050.

Pinget, A.-F. (2022). Individual differences in phonetic imitation and their role in sound change. *Phonetica*. https://doi.org/10.1515/phon-2022-2026.

Pinget, A.-F., Kager, R., & Van de Velde, H. (2020). Linking variation in perception and production in sound change: Evidence from Dutch obstruent devoicing. *Language and Speech, 63*(3), 660–685. https://doi.org/10.1177/0023830919880206.

Qin, Z., Tremblay, A., & Zhang, J. (2019). Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *Journal of Phonetics, 73*, 144–157. https://doi.org/10.1016/j.wocn.2019.01.002.

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

Schertz, J., Kang, Y., & Han, S. (2019). Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast. *Laboratory Phonology, 10*(1). https://doi.org/10.5334/labphon.67 1.

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., & Remez, R. E. (2002). Learning to recognize talkers from natural, sinewave, and reversed speech samples. *Journal of Experimental Psychology. Human Perception and Performance, 28*(6), 1447–1469.

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.

Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. (2011). Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *Journal of Speech, Language, and Hearing Research, 54*(4), 1011–1021. https://doi.org/10.1044/1092-4388(2010/10-0036).

Stevens, M., & Harrington, J. (2014). The individual and the actuation of sound change. *Loquens, 1*(1), e003.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24–31. https://doi.org/10.1177/0098628316677643.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096.

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. (2007). Tonogenesis revisted: Revising the model and the analysis. In J. G. Harris, Somsonge Burusphat, & J. E. Harris (Eds.), *Studies in Tai and Southeast Asian Linguistics* (pp. 261–291). Ekphim Thai.

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.

Valbret, H., Moulines, E., & Tubach, J. P. (1992). Voice transformation using PSOLA technique. *Speech Communication, 11*(2), 175–187. https://doi.org/10.1016/0167-6393(92)90012-V.

Wang, P. (1987). A reanalysis of Suzhou phonology (苏州音系再分析). *Studies in Language and Linguistics (语言研究), 01*, 41–48.

Warren, P., Hay, J., & Thomas, B. (2007). The loci of sound change effects in recognition and perception. In J. Cole & J. I. Hualde (Eds.). *Laboratory phonology* (9, pp. 87–112). Mouton de Gruyter.

Yang, R. (2015). The role of phonation cues in Mandarin tonal perception. *Journal of Chinese Linguistics, 43*(1), 453–472. https://doi.org/10.1353/jcl.2015.0035.

Ye, X. (1979). Tone sandhi in Suzhou dialect (苏州方言的连读变调). *Dialect (方言), 01*, 30–46.

Yu, A. C. L. (2016). Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *The Journal of the Acoustical Society of America, 139*(4), 1672–1690. https://doi.org/10.1121/1.4944992.

Yu, A. C. L., & Zellou, G. (2019). Individual Differences in Language Processing: Phonology. *Annual Review of Linguistics, 5*(1), 131–150. https://doi.org/10.1146/annurev-linguistics-011516-033815.

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., & 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.

Zhang, Y., & Kirby, J. (2020). The role of F0 and phonation cues in Cantonese low tone perception. *The Journal of the Acoustical Society of America, 148*(1), EL40–EL45. https://doi.org/10.1121/10.0001523.