



# A Potential New Sound Change after Tonogenesis: A Preliminary Perceptual Study on the Tonal Contrast of Wenzhou Wu Chinese

Weijun Zhang, Peggy Pik Ki Mok<sup>1</sup>

<sup>1</sup>The Chinese University of Hong Kong, Hong Kong  
zhangweijun@link.cuhk.edu.hk, peggymok@cuhk.edu.hk

## Abstract

Wu Chinese, spoken in Southeast China, developed high and low tonal register contrast from onset voicing contrast in history. A new sound change observed in production is that the phonation type contrast of the vowel has emerged and meanwhile the pitch difference across tonal registers arising from tonogenesis is narrowing in young speakers' speech. The current study aims to investigate the perceptual cue(s) that distinguishes the tonal register contrast for speakers of different age groups in Wenzhou dialect of Wu. A generational change is found on the perceptual cues in perceiving the tonal contrast. Old speakers use both low pitch and vowel breathiness to perceive the low register tones, and any type of cue conflict may lead to a bias to the high register tones. The pitch still has an effect for young speakers but the cue weighting has been decreased. The vowel phonation type difference has become the primary perceptual cue of tonal register contrast for young speakers. A potential new sound change after tonogenesis could be observed, that the phonetic realisation of the contrast has changed from a segmental feature (consonantal voicing) to a suprasegmental feature (tone), and now is further changing to another segmental feature (vowel phonation).

**Index Terms:** Wu Chinese, tonogenesis, sound change, tone, phonation type

## 1. Introduction

Wu is a group of Chinese varieties spoken in Southeast China and many diaspora Chinese communities with about 70 million native speakers. The four tones of Middle Chinese were developed into eight tones of two tonal registers in Wu by the consonantal voicing contrast, i.e., high register tones in syllables with voiceless onsets and low register tones in those with voiced onsets, which reflects a common tonal split mechanism in the history of tonogenesis of many East and Southeast Asian languages [1, 2, 3]. However, a specific feature of Wu is that the phonological voicing contrast, which is lost in most Chinese languages, is still reported to be preserved in Wu, with a hand-in-hand relationship between the tonal (pitch) and phonological voicing contrast [4]. The coexistence of tonal register and voicing contrasts suggests that Wu Chinese is at an intermediate stage of tonogenesis [5]. In other words, the synchronic data of Wu Chinese may to some extent reflect a common areal historical sound change.

Previous studies on some Wu dialects (e.g. Shanghai Wu) indicate that the phonological voicing contrast in the initial position is no longer realised as consonantal voicing, but as phonation type (breathiness) of the following vowel [6, 7]. In our previous production research on Wenzhou Wu, regarding the phonetic realisation of voicing contrast, significantly

different age variants could be observed in plosive-onset syllables as shown in Figure 1: the young speakers produced breathy vowels and modal vowels in the syllables with low and high register tones respectively, while the old speakers did not make the contrast by vowel breathiness [8].

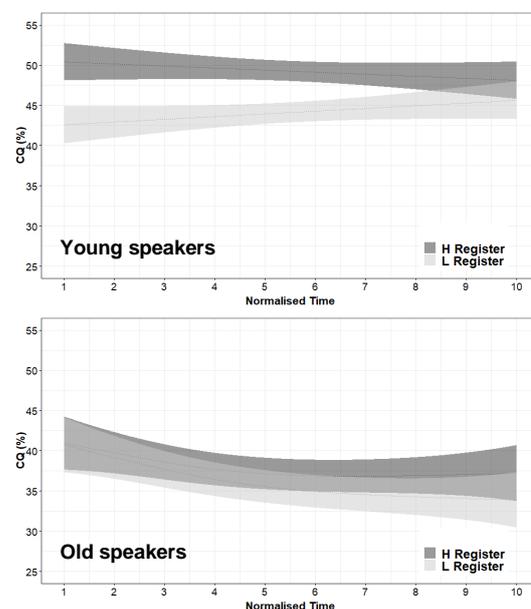


Figure 1: *The closed quotient (an articulatory correlate of breathiness) of the plosive-onset syllables.*

Meanwhile, we noticed a trend of tone change in Wenzhou Wu as well. As shown in Table 1, the tone inventory of Wenzhou Wu is conventionally described as a canonical eight-tone system, whereas T1 and T6, T3 and T4, T5 and T2, as well as T7 and T8 actually share identical contours respectively, that is, the four high register tones (T1, T3, T5 and T7) and the four low register tones (T2, T4, T6 and T8) are highly consistent with pitch value distinction only. In our production data, the normalised pitch differences within each pair of tones with the same contour were larger in old speakers' speech than that of young speakers, which implied a narrowing of pitch contrast across tonal registers in young speakers' speech. Taking the two level tones as an instance, for the old male speakers, the difference between the average normalised pitch values (logarithmic z-scores) of the high and low level tones was 1.56, while the corresponding data of young male speakers was 1.03 [9]. It is also reported that the difference between the two rising tones and the two dipping tones is hardly perceptible in their informant's speech [10].

Therefore, in Wenzhou Wu, it seems that an ongoing sound change in production is in progress that the pitch contrast arising from tonogenesis is gradually being replaced by the phonation type contrast of the vowel. Since the pitch difference between the high and low tonal registers is only weakened but not completely vanished in production, the emergence of vowel phonation contrast between the two types of syllables with different tonal registers would result in potential acoustic cue redundancy in young speakers' speech. Investigating the perceptual cue(s) that distinguishes the tonal register contrast and whether age variation or diachronic change also exists in perception can contribute to our understanding of the interaction between segmental and suprasegmental features in sound change.

Table 1: *Wenzhou Wu tone inventory.*

T1	T2	T3	T4	T5	T6	T7	T8
44	31	45	34	42	22	323	212

## 2. Method

### 2.1. Participants

Eight old speakers (60-72 years old) and twelve young speakers (23-30 years old) with normal hearing balanced between the two sexes participated in the experiment.

All of the participants were native Wenzhou Wu speakers born and raised in Wenzhou downtown with no experience living in other places before adulthood. The core family members of the participants were all local in Wenzhou and Wenzhou Wu was the dominant language or even the only language used in their family.

Participants were tested individually in a quiet room with the experimenter.

### 2.2. Experiment design and stimuli

A two-alternative forced-choice task was designed, in which participants had to choose an answer between a syllable pair potentially contrasting in both tonal register and phonation type of vowels. Referring to the acoustic and articulatory data of our previous production experiment, within the eight speakers, the data of a 26-year-old male speaker with significant differences on both pitch and phonation type of vowels when producing the two types of syllables in contrast with high and low register tones was used to generate the stimuli for the current study.

Six monosyllabic words with level tones covering all unaspirated plosives in Wenzhou Wu were chosen for resynthesis as shown in Table 2:

Table 2: *Words used in the current study (the contrast conventionally transcribed as tones as well as voicing).*

IPA	Gloss	Character	Tonal register
/pa <sup>44</sup> /	'class'	班	High
/ba <sup>22</sup> /	'handle (affairs)'	办	Low
/ta <sup>44</sup> /	'list'	单	High
/da <sup>22</sup> /	'bullet'	弹	Low
/ka <sup>44</sup> /	'close'	关	High
/ga <sup>22</sup> /	'throw'	攒	Low

The duration of the tokens was normalised as identical to the mean of 393.6 ms. The six tokens were then resynthesised into a 12-step pitch continuum with equal semitone steps ranging from 113 Hz to 136 Hz in Praat [11]. To avoid the unnaturalness of the completely level pitch curves, the original contours of the pitch curves were kept, with the midpoints of the curves aligned with the target pitch values for resynthesis. The phonation type of vowels was not changed in resynthesis.

The experiment hence consisted of 72 trials of both modal and breathy vowels with 12 continuous pitch levels. All participants heard three repetitions for each token, with breaks between repetitions. The order of stimuli presentation was randomised for each participant and each repetition.

### 2.3. Procedure

Participants sat at a comfortable distance from the computer screen and wore earphones. The session began with six practice trials for participants to get familiar with the experimental procedure. In the practice phase, participants heard six typical tokens, i.e., three with modal vowels and highest pitch and three with breathy vowels and lowest pitch. The Chinese characters of the token and its competitor (the other one in the contrastive pair) were shown on the screen with a fixed order of the high tone one on the left and the low tone one on the right. The tokens presented during the practice phase were also included in the experiment (as tokens with the most extreme pitch at both ends). The purpose of giving an advance preview of the stimuli was to familiarise the participants with the experimental words and its Chinese characters to exclude any lexical influence on participants' feedback as much as possible, since the word /ga<sup>22</sup>/ meaning 'throw' was not used in formal written text and thus participants may not know the character, and the character of /da<sup>22</sup>/ had a homograph.

Participants were instructed to place their two index fingers on 'z' and 'm' on the keyboard throughout the whole experiment and chose the answer based on their intuition by pressing 'z' or 'm'. The same laptop with keyboard was used by all of the participants. The response and the reaction time of each trial were recorded.

## 3. Results & Discussion

### 3.1. Accuracy of response

#### 3.1.1. Low register tone (breathy vowels)

Since the perceptual cue(s) of the tonal contrast per se is a crucial research objective and thus no presumable pattern was expected, the accuracy of response is simply calculated as percent correct. According to Figure 2, the accuracy pattern of the low register tone shows an apparent age variation. The old speakers, who do not distinguish tones with voice quality difference in their production, strongly relied on the cue of pitch to distinguish the tonal contrast when the cue of vowel breathiness is present. The negative linear correlation between pitch difference and accuracy of response strongly suggested a rather typical continuous perception pattern. The old speakers tended to regard the high pitch tokens as words with the high register tone despite the existence of a conflicting cue (breathiness). As for the young speakers, the decrease on accuracy was much slower and the accuracy had been consistently higher than chance level (0.5) till the end of the continuum. Furthermore, different from the linear pattern

found in old speakers' data, young speakers seem to be insensitive within the pitch steps 6 to 8, as suggested by the plateau on the curve.

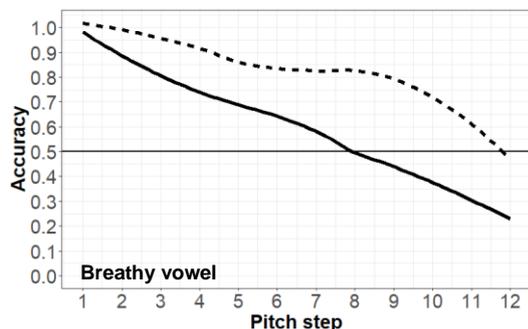


Figure 2: Accuracy of response to words with the low register tone (Solid line: old; Dashed line: young).

In general, it can be concluded that all participants use pitch as one perceptual cue to distinguish the contrast, while age variation appeared in cue weighting: pitch played a more prominent role in old speakers' perception of the contrast.

### 3.1.2. High register tone (modal vowels)

As for words with the high register tone and modal vowels (i.e., no conflicting cue), the two groups of participants showed similar patterns. Regardless of the pitch values, all of the participants could have accuracy higher than chance level. Besides, differences in pitch within the range of steps 12 to 8 seem to have no effect on the 100% perceptual accuracy of both groups of participants.

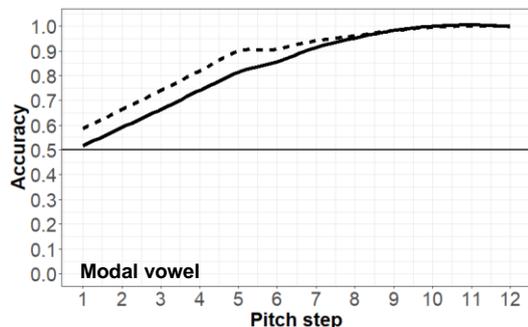


Figure 3: Accuracy of response to words with the high register tone (Solid line: old; Dashed line: young).

Based on Figures 2 and 3, an asymmetrical effect on old speakers' accuracy of response can be observed. The old speakers would regard a sound as a low tone syllable only when both segmental (vowel breathiness) and suprasegmental (low pitch) cues were present. In other words, for the old speakers, the syllables with low register tones were more marked than those with high register tones. An interesting fact is that the old speakers did not use phonation type difference to realise the contrast in their own production as indicated in Figure 1, but vowel breathiness has become an important cue for perceiving the low register tones (at least when listening to the young speakers' speech), which is presumably owing to the quality of stimuli which is apparently produced by a young

speaker. Therefore, apart from pitch and vowel phonation type, recordings of old or young speakers seems also necessary to be included as a variable for stimuli synthesis and be assessed whether it has an effect on the participants' performance.

For the young speakers, however, the difference between breathy vowels and modal vowels had become the most significant perceptual cue to distinguish the tonal contrast, while pitch still served as an influential factor as well. The age variation shown in the experimental results may reflect a potential new stage after the completion of basic tonogenesis. The phonetic realisation of the contrast experienced a sound change from onset voicing contrast to pitch contrast, and may further change into vowel phonation type contrast.

## 3.2. Reaction time

### 3.2.1. Low register tone (breathy vowels)

For the pitch range from steps 1 to 7, though the accuracy of response to words with the low register tone by old speakers was obviously lower than young speakers as demonstrated in Figure 2, there was no difference observed regarding the reaction time. As mentioned above, pitch was still at least one of the perceptual cues for all speakers, it could be noticed that the perceptual uncertainty implied by reaction time increased as expected when pitch increased, whereas the reaction time curves of old and young speakers became increasingly further apart when the pitch rose to step 8 or above, where old speakers had even faster reaction during this phase.

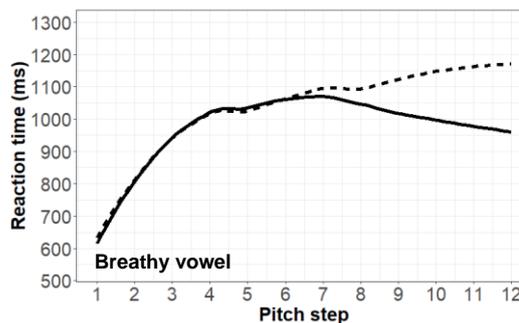


Figure 4: Reaction time of words with the low register tone (Solid line: old; Dashed line: young).

As Figures 2 and 4 show, in light of the synchronisation of decrease in reaction time and accuracy falling below chance level, the acceleration of old speakers' reaction time in these high pitch trials higher than step 8 is likely the result of the old speakers' bias of perceiving these syllables as the high register tone, and hence the perceptual uncertainty (reaction time) would be negatively correlated with pitch in these trials. Since young speakers still perceived the presence of vowel breathiness (an important cue for the low register tone in their own production), the further rising of pitch kept intensifying the cue conflict. Thus, the increased reaction time of the young speakers at the high end of the pitch step continuum reflects perceptual uncertainty due to cue conflict.

### 3.2.2. High register tone (modal vowels)

As the old and young speakers showed similar patterns in accuracy without vowel breathiness, the reaction time patterns shown in Figure 5 by the two age groups were also not very

different. The reaction time of the young speakers was generally shorter during mid pitch steps, while basically the same as the old speakers in peripheral pitch ranges. A plateau effect could be observed during mid pitch steps (among the pitch steps around 6 to 8) in young speakers' reaction time pattern, similar to the pattern shown in accuracy of response.

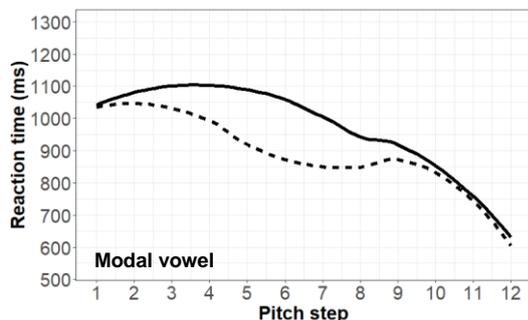


Figure 5: Reaction time of words with the high register tone (Solid line: old; Dashed line: young).

Taking the accuracy and reaction time patterns of the low register tone together, it seems that a significant boundary was centred around the pitch step 8, rather than the central value of the pitch range. Previous research on other tonal languages reported similar asymmetrical effect in tone perception as well. In the perception of Cantonese high and mid level tones, it is found that the maximal perceptual uncertainty, i.e., the participants' category boundaries did not position at the central pitch value but a higher value [12]. The asymmetrical effect found may reflect language-specific phonological properties that more tones concentrate in the lower tone space rather than the upper tone space in both Cantonese and Wenzhou Wu, and thus the pressure from the low tone space has an effect on the position of perceptual boundaries.

#### 4. Conclusion

Based on the discussion above, it can be concluded that there is age variation on the perceptual cue(s) in perceiving the high and low register tonal contrast in Wenzhou Wu. For old speakers, both low pitch and vowel breathiness are necessary for the perception of the low register tone (of young speakers' voice), while the sole existence of either low pitch or vowel breathiness would tend to be regarded as high register tones. For young speakers, the cue weighting of pitch has been decreased to a large extent, which may suggest a potential diachronic change in progress that the vowel phonation type contrast is playing an increasingly important role in distinguishing this contrast, consistent with the change of acoustic correlates of the contrast happening in production. Therefore, a potential new sound change after tonogenesis could be observed in both production and perception.

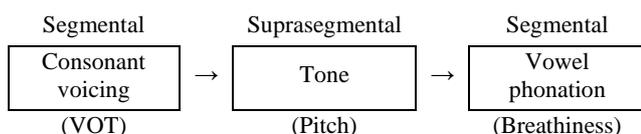


Figure 6: Three stages of the potential sound change.

The different features listed in Figure 6 are all potentially used as phonetic realizations of the same phonological contrast in Wu, and hence these features or corresponding parameters seem to be closely related. What is the internal acoustic interaction mechanism leading to the phonetic compensation we observed is still largely unexplored.

The current study only generalise results to the level tone pair with different plosive onsets as a preliminary experiment, the involvement of the other three tonal pairs, the rising, falling and dipping tones and more participants seem to be necessary in further investigation. And the vowel phonation was not resynthesised as a continuum of spectral quality from modal to breathy voice in the current research, the phonation type of vowels would be resynthesised regarding degree of breathiness in further studies to examine the mixed effects of various parameters and to construct statistical analysis. Besides, since the onset voicing contrast is still preserved in all fricative pairs in Wenzhou Wu [8], the coexistence of all of the three factors in Figure 6 allows us to do further perceptual research on fricatives and assess the perceptual cue(s) and their differential cue weighting as well as more sophisticated interaction mechanism among various segmental and suprasegmental features during sound change.

#### 5. References

- [1] D. Jiang, *Studies in the Tibetan Phonological History*. Beijing: The Ethnic Publishing House, 2002. (in Chinese)
- [2] J. A. Matisoff, *The Loloish Tonal Split Revisited*. Taipei: Institute of Linguistics, Academia Sinica, 2002.
- [3] J. A. Matisoff, "Tonogenesis in Southeast Asia," in *Consonant Types & Tone*, Los Angeles: University of Southern California, 1973, pp.71-96.
- [4] Y. R. Chao, *Studies in the Modern Wu-dialects*. Peking: Tsing Hua College Research Institute, 1928. (in Chinese)
- [5] L. M. Hyman, "Phonologization," in *Linguistic Studies Offered to Joseph Greenberg on the Occasion of his Sixtieth Birthday*, vol. 2, pp. 407-418, 1976.
- [6] Z. Chen, "An acoustic study of voiceless onset followed by breathiness of Wu dialects: based on the Shanghai dialect," *Studies in Language and Linguistics*, vol. 30, no. 3, pp. 20-34, 2010. (in Chinese)
- [7] J. Gao, *Interdependence between Tones, Segments and Phonation Types in Shanghai Chinese*. Ph.D. dissertation, University Sorbonne Nouvelle – Paris 3, 2015.
- [8] W. Zhang, W. Xu, and P. P. K. Mok, "A preliminary study on the age variation of the voicing contrast in Wenzhou Wu Chinese," *Proceedings of the 19th International Congress of Phonetic Sciences*, pp. 2411-2415, 2019.
- [9] W. Zhang, *An Experimental Study on Contemporary Variation of Wenzhou Wu Tones*. MA thesis, The Chinese University of Hong Kong, 2017.
- [10] M. Y. Chen, *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press, 2000.
- [11] P. Boersma and D. Weenink, Praat, version 6.1.08, 2019.
- [12] J. S. Nixon, J. van Rij, P. Mok, R. H. Baayen, and Y. Chen, "The temporal dynamics of perceptual uncertainty: eye movement evidence from Cantonese segment and tone perception," *Journal of Memory and Language*, vol. 90, pp. 103-125, 2016.