TONATORY PATTERNS IN TAIZHOU WU TONES

Phil Rose

Emeritus Faculty, Australian National University
philjohn.rose@gmail.com

ABSTRACT

Recordings of speakers of the Táizhou subgroup of Wu Chinese are used to acoustically document an interaction between tone and phonation first attested in 1928. One or two of their typically seven or eight tones are shown to have what sounds like a mid-Rhyme glottal-stop, thus demonstrating a new pattern in Wu tonatory typology. Possibly reflecting gradual loss, larygealisation appears restricted to the north and north-west, and is absent in Huángyán dialect where it was first described. A perturbatory model of the larygealisation is tested in an experiment determining how much of the complete tonal F0 contour can be restored from a few centiseconds of modal F0 at Rhyme onset and offset. The results are used both to acoustically quantify laryngealised tonal F0, with its problematic jitter and shimmer, and to help investigate the conditioning factors of its realisation.

Keywords: Wu dialects, Taizhou, tonal acoustics, non-modal phonation, glottal-stop, creaky voice

1. INTRODUCTION

Although the primary auditory and phonological feature of the linguistic category of tone is pitch [1, 2, 3], some tone languages also show interaction with extrinsic phonation type. Tone languages where extrinsic phonation type is an integral part of tonal realisation have been dubbed *tonatory* [4]. Northern Vietnamese, with its creaky and 'constricted' tones is a well-known example, e.g. [5]. Another example is the contrastive tonation in Thai Phake, with its two high falling tones, one of which ends in creak [6]. Such interaction is to be expected, as both phonation types and tone share a laryngeal articulator, resulting from valvular action at different levels at and above the vocal folds [7, 8, 9].

In his 1928 survey of Wu dialects – the first western-linguistic description of Chinese dialect phonetics – Chao Yuenren [10] noted of two tones in Huángyán 黃岩 that they had a glottal-stop in the middle of the vowel giving the impression of two syllables.

The aim of this paper is to investigate this tonation by quantifying the tonal acoustics for four speakers with laryngealised tones from the Táizhou 台州 subgroup of Wu to which Huángyán belongs. The issue has significance within descriptive tonetics, tonatory typology and historical linguistics. Wu dialects – at least the conservative varieties – show a wide range of tonatory behaviour [11]. One finds breathy or ventricular phonation in groups of tones characterising natural tonal classes of importance for phonotactics and Wu's complex tone sandhi. One also finds a single tone characterised by a different non-modal phonation type [12]; or even two different non-modal phonation types in two tones. However, the Huangyan-type tonation seems to involve a new variation, with the same phonation type in two different tones from the same historical tonal category, thus prompting speculation that it developed before the tonal split.

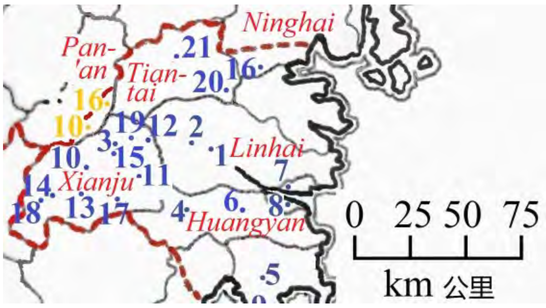
2. PROCEDURE

My main source of data was the large number of tape-recordings made by Prof. Zhu Xiaonong and his assistant in the mid to late 90's as part of his post-doctoral survey of Wu tones and tone sandhi funded by the *Australian Research Council*. These were supplemented by tape-recordings made about a decade earlier by Prof. William Ballard as part of his survey of tones and tone sandhi in southern Wu varieties, funded by the *US National Endowment for Humanities* [14].

Recordings from 23 Taizhou sites were identified, digitised at 22.05k, and their citation tones saved for further processing. The sites are shown in figure 1, which is based on maps in [15] and [16], amended according to [13].

It was very clear from even casual listening that some recorded speakers had extrinsic mid-tone larygealisation. (Chao's *glottal-stop* characterisation has been questioned and the Taizhou phonation said to be better described as creaky [13]. Since glottal-stops are commonly realised in continuous speech with creak or creaky voice, the more general term *larygealisation* is preferable.) Space permits the description of only four of these speakers, from Tiāntái county sites blue 20 (Sānhézhèn 三合镇) and blue 21 (Yìzháixiāng 义宅乡); Nínghǎi county site blue 16 (Sāngzhōuzhèn 桑州镇); and Pán'ān county site yellow 16 (Fāngqiánzhèn 方前镇) (see figure 1). These four speakers, referred to below as SH, YZ, SZ and FQ, all had larygealisation in two tones, which can be heard by clicking on the site names above. The speakers' complete tonal data may be listened to, and their acoustics inspected, on the author's website [17].

Figure 1: Map of Taizhou region, showing putative Taizhou sites documented. Names of counties referenced in paper are shown in red. Thin black lines = county boundaries. Red dashed lines = putative Taizhou subgroup boundary.

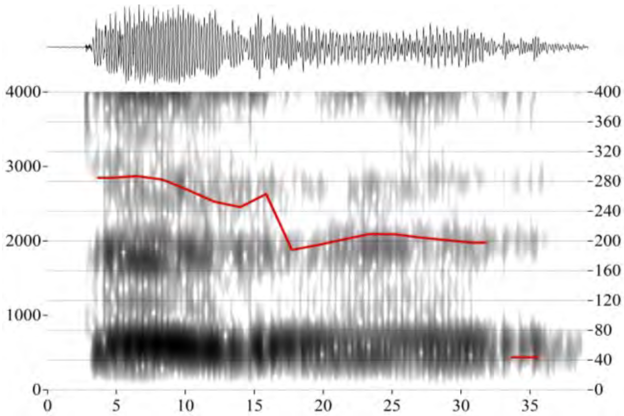


Data were first transcribed auditorily with respect to pitch, length and phonation type – a necessary step when confronting unfamiliar data and also to avoid mechanically extracting mean F0 values that conflate different tonal targets. Both Ballard and Zhu used an elicitation common in Chinese dialectology by requiring their informants to read out Chinese characters they expected, from received knowledge of historical linguistics, would exemplify specific phonological categories, like tones, onsets, and rhymes. Thus their tonal examples were chosen according to the eight traditional Middle Chinese categories. This is a sensible heuristic. However, for several reasons to do with historical tonal changes in the area, just because a character is known to belong to a historical category is no guarantee that its modern reading will directly reflect that. In the absence of standard face-to-face fieldwork interaction with informants, all this can most easily be resolved by auditory transcription followed by confirmation by acoustic analysis.

Tonal acoustics were measured by first generating a wideband spectrogram in *Praat*, together with an aligned waveform and superimposed F0 contour. The good time-domain resolution of the spectrogram was then used to hand-segment the onset and offset points of the Rhyme. Onset was deemed to be at the first high-amplitude glottal pulse of the Rhyme; offset was adjudged to be the point where irregularities in either amplitude or period, or both, became evident. The F0 over each Rhyme was then extracted with *Praat*.

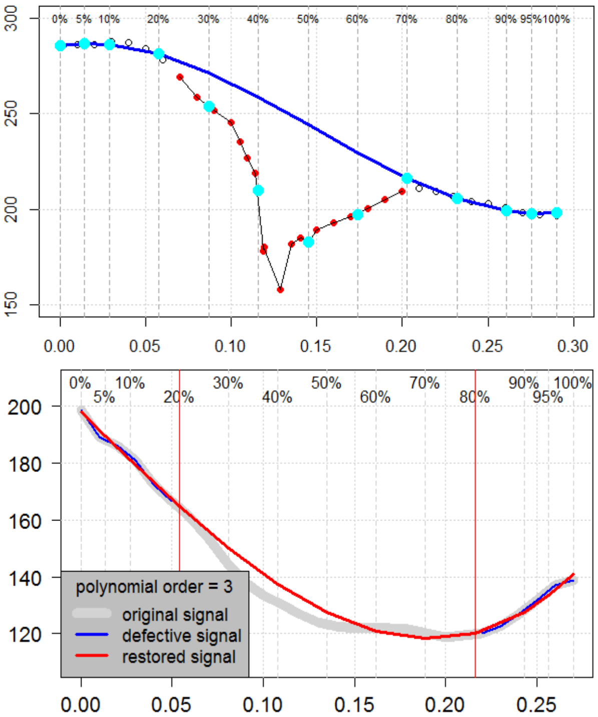
It is well-known that larygealisation is usually accompanied by relatively heavy jitter and shimmer. Figure 2 shows the acoustics of a laryngealised tone in the word /tø 5^{ʔ}3/ 短 *short* from SH. High jitter and shimmer are obvious over a relatively short stretch in mid-Rhyme of about 6 csec., between ca. csec. 12 and ca. csec. 18. Less obviously, but still auditorily clear, is a less jittery stretch from ca. csec.20 to ca. csec.25. (The Rhyme offset can also be seen, indicated by onset of irregular pulsing, at ca. csec. 31). It can also be appreciated that, despite a low pitch floor setting in *Praat* of 50 Hz, the F0 has not been well extracted over the initial high jitter stretch, in the middle of which actually occur periods corresponding to between 150 and 200 Hz.

Figure 2: Acoustics of laryngealised tone, showing typical jitter and shimmer. Axes – *left*: spectral frequency (Hz), right: F0 (Hz), *bottom*: duration (csec.).



From a perceptual point of view, of course, the precise values of F0 during the laryngealised portion of a tone are irrelevant: it is the jitter and shimmer which generate the percept cuing the category. For the purposes of this paper, however, it was necessary, quite apart from correcting failures in automatic F0 extraction, to devise a method of accurately sampling the F0 time course throughout the Rhyme, in order to facilitate inferences on production. Thus, depending on whether the tone was laryngealised, different methods were used for sampling the F0. For non-laryngealised tokens, F0 was modelled in *R* with an eighth-order polynomial, the high order being necessary to avoid over-smoothing at this point of the procedure. For laryngealised tones, separate models were used for a central portion of the F0 containing the obviously laryngealised portion and the remaining peripheral modal values. F0 from the peripheral portions was modelled with a low order polynomial (cubic or quartic); F0 from the central portion was modelled with R's *lowess* function. The F0 was then sampled using these models, with a sufficiently high frequency to capture the details of its time-course: at 10% points of the Rhyme as well as at 5% and 95%. The top panel of figure 3 illustrates this procedure with the F0 from the token shown in figure 2. The

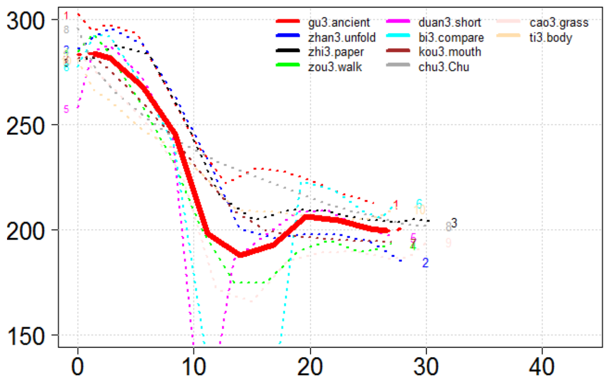
Figure 3: Top: Illustration of procedure for extracting F0 in laryngealised tones. Bottom: Cubic polynomial restoration of excised F0 in a *non-laryngealised* high falling-rising tone. X-axis = duration (csec.), y-axis = F0 (Hz). See text for explanation.



central F0 values containing the putative laryngealised F0 are shown with red dots. Because the first half of these were incorrectly extracted by *Praat* (see figure 2) they were measured manually, and pitch-synchronously, from the waveform. The cubic polynomial fitted to the peripheral portion is shown in blue, the *lowess* fit to the central portion (smoothing = 0.2) is shown with a thin black line). The sampling points for tonal F0 are shown along the top, and with vertical dashed lines; and the sampled tonal F0 values at these points are shown with larger cyan dots. It can be seen that this method results in a close fit to the observed F0 values.

Although the configuration in the top panel of figure 3 invites rather strongly the interpretation of an impulse response in an underlying gradually decaying parameter (vocal fold tension perhaps?), a prior ontological question arises: are the interpolated values real? (The danger of polynomial inter- and extrapolation is, after all, a commonplace.) So an experiment was conducted to ascertain how much of a *non-laryngealised* tonal F0 contour excised from mid-rhyme can be restored from polynomial modelling of the type demonstrated above. Non-laryngealised contour tones from different Wu varieties were used, and increasing portions of their F0 removed from the middle of their time course. The resulting defective F0 was then modelled with a low-order (cubic or quartic) polynomial and the results compared with the original intact F0 shape. The bottom panel of figure 3 shows one such trial on the F0, in thick grey, of a *non-laryngealised* high falling-rising tone of a speaker from Xiānjū 仙居 blue 3, where slightly more than 60% of the F0 has been removed (from 20% to 80% of duration). The remaining data, in blue, obviously contain sufficient information on the original F0 contour to enable interpolation with a cubic, in red, with a low mean-squared error. It was found that this was generally the case, with better approximation, of course, the less F0 removed.

Figure 4: Sampled F0 in 10 tokens of a laryngealised tone. Red line = mean. X-axis = duration (csec.), y-axis = F0 (Hz).



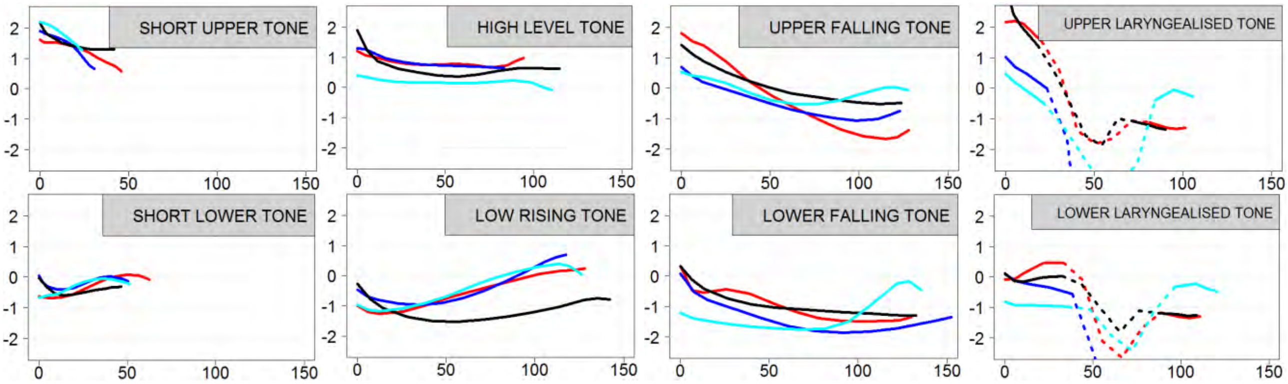
Mean F0 trajectory was then calculated from the sampling points over all tokens of a tone. Figure 4 shows the resulting plot for ten tokens of a laryngealised tone. It can be seen that the different tokens show perturbatory effects of different magnitude. Inspection of the plots for the laryngealised tones showed that the perturbations in mid time-course have a weak positive correlation with vowel height.

The mean F0 data were then z-score normalised [18] using normalisation parameters from all except the laryngealised tones. This approach has been shown to give by far the best results in tonal F0 normalisation [19]. Duration was normalised by converting raw duration to a percentage of a speaker's mean tonal duration [20]. Z-score normalised tonal F0 as a function of normalised duration for the four speakers is shown in figure 5.

3. RESULTS & DISCUSSION

Results of the auditory transcription indicated rather similar eight-tone systems for the four speakers, with some between-speaker differences. (In the absence of data from more speakers, it is not of course possible to know whether these between-speaker differences represent differences between- or within-varieties.)

Figure 5: Normalised tonal acoustics for eight tones of four Taizhou speakers (SH YZ FQ SZ) with two laryngealised tones. x-axis = normalised duration, y-axis = z-score normalised F0, dotted lines = larygealisation.



The description of the individual tones is best done in conjunction with the normalised tonal acoustics in figure 5. Typical of many Wu dialects, all speakers had the typical Wu tone template of two upper and lower register sets of tones, including an upper and lower pair of short tones [11, 21]. The **short upper tone** had either a level or a slightly falling pitch in the top of the speakers' pitch range: 5 ~ 54 , e.g. [pǐʔ] ~ [pǐ] *pen* 笔, [çʏǒ̈ʔ] *blood* 血. The **short lower tone** had a level or slightly rising pitch in the speakers' lower pitch range: 23 ~ 2, often with whispery voice, e.g. [b̥æ̤̆], *white* 白, [d̥ɔ̤̆] *read* 读. Notable was that many of the short tone tokens lacked the expected final glottal-stop which occurs in other Wu varieties which have preserved short reflexes of the Middle Chinese (MC) *ru* category. All speakers also shared a **high level tone** (55 ~ 44) and a (often whispery) **low rising tone** (23 ~ 113) as reflexes of the MC categories *yinqu* and *yangqu* respectively. High level examples are [tɕɪŋ] *correct* 正, [tɕy] *expensive* 贵. Low rising examples are [lṳ] *road* 路, [d̥i̤] *ground* 地. The reflexes of MC *yinping*, which can be summarised as **upper-falling,** have high or mid falling pitch followed by varying degrees of final pitch rise: SH 51(1), YZ 322, FQ 433, SZ 434). Examples are [ʔe] *safe* 安, [tɪŋ] a surname 丁. Variation in the **low-falling tone** (< MC *yangping*) is similar to the upper-falling, with whispery voice: SH 211, YZ 212, FQ 21~11, SZ 213. Examples are: [g̊ɪɔ̤ŋ] *poor* 穷, [b̥ɪ̤ŋ] *level* 平. The **upper laryngealised tone** (< MC *yinshang*) shows between-speaker differences in relative pitch height of its two portions. In SH YZ and FQ, the first is high and the second lower; in SZ the second is higher. Speakers also differ in recovery of modal F0 in the second portion: SH FQ and SZ mostly recover modal phonation but YZ has creaky voice phonation to the end of the syllable. This variation can be represented as 5ˀ3 ~ 5ˀ2 (SH FQ); 5ˀ2̰ ~ 5ˀ1̰ (YZ); 3ˀ4 (SZ). Examples are [kuˀu] *ancient* 古, [tɕɪˀŋ] *tight*. The **lower laryngealised tone** shows parallel variation to this, but with a lower first portion, often showing a rising pitch from word-initial depression [22]: 23ˀ2 (SH FQ); 3ˀ1̰ (YZ); 2ˀ3 (SZ). Examples are: [ŋ̍ˀŋ] *five* 五, [lɐˀʊ] *old* 老.

Given the involvement of what sounds like glottal-stop and creaky voice in the realisation of the two laryngealised tones, the most likely cause is ventricular incursion. It has been demonstrated in many laryngoscopic studies, e.g. [7, 8], that the ventricular folds are centrally involved in the production of both glottal-stop and creaky voice [9], whereby the false vocal folds are medially adducted to load the vocal folds, the damping resulting in irregular lower frequency oscillation (creak, creaky voice), or its cessation (glottal-stop). A sensible inference would be that the laryngealised tonal acoustics demonstrated in this paper result from a constricting epilaryngeal tube gesture of the ventricular folds initiated just after onset of phonation, then withdrawn in SH FQ and SZ, and partially withdrawn in YZ. Following [9, table 2] the superscript ˀ symbol was chosen to reflect this phonemically.

4. SUMMARY

Prompted by Chao's 1928 description of Huángyán, this paper has described the tonal acoustics of four speakers that appear to show the same tonation, and hypothesised that the observed larygealisation is due to ventricular incursion. This tonation is probably in decline: of the 19 other Táizhou speakers investigated, a few – from sites in Xiānjū county – showed larygealisation only in the upper tone. The remainder – including speakers from Huángyán county where the tonation was originally documented – have none [13]. Consonant with the *Endangered Languages* theme of this conference, perhaps one should speak in this case of *endangered tone*.

5. REFERENCES

[1] Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: CUP.

[2] Yip, M. 2002. *Tone.* Cambridge: CUP.

[3] Pike, K. 1972. *Tone Languages.* Ann Arbor: U. Michigan Press.

[4] Bradley, D. (ed), 1982. *Tonation*. Papers in South East Asian Linguistics No. 8. *Pacific Linguistics*.

[5] Earle, M.A. 1975. *An acoustic phonetic study of North Vietnamese tones*. Monograph 11, Santa Barbara: Speech Communication Research Laboratories Inc.

[6] Rose, P. 1990. Thai-Phake tones: acoustic, aerodynamic and perceptual data on a Tai dialect with contrastive creak. In: Seidl, R (ed), *Proc. 3rd Australian Intl. Conf. on Speech Science and Technology,* Melbourne, 394–399.

[7] Esling, J., Harris, J. 2005. States of the Glottis: An Articulatory Phonetic Model Based on Laryngoscopic Observations. In: Hardcastle, W., Beck, J. (eds), *A figure of speech: A Festschrift for John Laver*. Mahwah, New Jersey: Erlbaum, 347–383.

[8] Edmondson, J., Esling, J. 2006. The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology* 23, 157–191.

[9] Moisik, S., Esling, J. 2011. The 'Whole Larynx' approach to Laryngeal Features. *Proc. 15th ICPhS*, Hong Kong, 1406–1409.

[10] Chao, Y. 趙元任. 1928. 現代吳語的研究 *Studies in the Modern Wu Dialects*. Peking: Tsing Hua College Research Institute Monograph 4.

[11] Rose, P. 2015. Tonation in Three Chinese Wu Dialects. *Proc. 18th ICPhS* Glasgow, no page numbers.

[12] Steed, W. 2006. Phonation Type and Tone Sandhi as Evidence of Stress in Longquan Wu. In: Warren, P., Watson, C. (eds), *Proc. 11th Australasian Intl. Conf. on Speech Science and Technology*, Sydney, 82–87.

[13] Zhu, X. 2006. Creaky voice and the dialectal boundary between Taizhou and Wuzhou Wu. *Journal of Chinese Linguistics* 34/1, 121–133.

[14] Ballard, W.L. 1992. Oujiang Wu Tone Sandhi: Visi-Pitch Results. Chinese Language and Linguistics 1: Chinese Dialects, Symposium Series of the Institute of History and Philology Academia Sinica, vol 2, 41–66.

[15] Wurm, S. et al. 1987. *Language Atlas of China*. Longmans.

[16] Fu, G. 傅国通, Fang, S. 方松熹, Cai, Y. 蔡勇飞, Bao, S. 鲍士杰, Fu, Z. 傅佐之 1985. 浙江吴语分区 [Wu dialect subgroups of Zhejiang]. Zhejiang Linguistics Society, Hangzhou.

[17] http://philjohnrose.net

[18] Rose, P. 1987. Some considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication,* 6/4, 343–352.

[19] Rose, P. 2016. Comparing Normalisation Strategies for Citation Tone F0 in Four Chinese Dialects. In: Carignan, C., Tyler, M. (eds), *Proc.16th Australasian Int'l Conf. on Speech Science & Technology*, *Sydney,* 221–224.

[20] Rose, P. 2000. Hong Kong Cantonese Citation Tone Acoustics: A Linguistic-Tonetic Study. In: Barlow M. (ed), *Proc. 8th Australian Int'l Conf. on Speech Science and Technology,* Canberra, 198–203.

[21] Chao, Y. 1976. Contrastive Aspects of the Wu Dialects. In: Dil, A. (ed), *Aspects of Chinese Socio-Linguistics*, 34–47, Stanford: Stanford University Press.

[22] Rose, P. 2002**.** Independent depressor and register effects in Wu dialect tonology: Evidence from Wenzhou tone sandhi. *Journal of Chinese Linguistics* 30/1, 39–81.