Modeling Interaction between Tone and Phonation Type in the Northern Wu Dialect of Jinshan

Phil Rose#1, Tianle Yang*2*

*Independent researcher, Australia
#ANU Emeritus Faculty, Australia
[1] https://philjohnrose.net, [2] u6512077@alumni.anu.edu.au

Abstract

Impressionistic and acoustic data are presented for the seven tones of the Wu Chinese dialect of Jinshan 金山, where tone is much more than just pitch. The independence of extrinsic phonation type from syllable Onsets is exemplified, and it is argued using quantified tonatory parameters that phonation type determines tonal pitch, not *vice versa*. Command-response modeling is then used to factor tone into depression and tonal target components, which enable a more precise understanding of Jinshan tonological structure.

Index Terms: tonal acoustics, Wu dialects, phonation type, depression, command-response model.

1. Introduction

"Tone is seldom, if ever, a matter of pitch alone." Thus Eugenie Henderson, one of the pioneer descriptive phoneticians and linguists of South East Asian languages [1]. Of course, pitch is criterial for tone: the definition of a tone language is, after all, one in which pitch is part of the phonological representation of words [2 p.4, 3 p.229]. However there are many tone languages, especially in S.E. Asia, where tonal pitch is closely intertwined with other segmental and suprasegmental aspects of the syllable and word, and this paper describes the tones of one of them: the northern Wu dialect of Jinshan in southern Jiangsu province.

Jinshan belongs to the Tàihú-Sūhùjiā 太湖苏户嘉 sub-subgroup of Wu. Although close to Shanghai, the two varieties differ a lot, with Jinshan having more complex tones and tone sandhi. But it is a typical Wu dialect in its complex interaction between tonal pitch, phonation type, duration, vowel quality, syllable-structure and syllable Onsets. This paper aims to show how, with speech science and, some might say, speech technology, two of these components – tone and phonation type – can be quantitatively disentangled.

Jinshan has, we think, not been previously described. A comparison of Wu dialect descriptions of 33 sites in 1928 and again in 1992 [4, 5] shows that they changed considerably in this sixty year period, and more recent socio-phonetic findings on closely related Wu varieties [6] suggest that tonal change is accelerating in metropolitan areas due to urbanisation. In a sense, therefore, this description may also constitute a salvage operation.

2. Procedure

2.1. Informants, elicitation

Because of recent changes in the speech of younger Shanghai speakers [7], it was considered advisable to collect data from older Jinshan speakers, and so nine speakers over 60 years old were selected and recorded by the second author, who is a native Jinshan speaker (albeit a youngish one). Three of them are described here: two males and a female.

Informants were given the list of 453 basic words for exemplifying Chinese dialect lexicon in [8 pp.18-26] and asked to read out the equivalent Jinshan word. Some of the recordings may be listened to at [9].

The recordings were first phonetically transcribed, and then manually labeled in *Praat*. Transcription is an essential part of the process: it enables one to become familiar with a voice and note features of possible phonetic and/or phonological importance (to take an actual example from the recordings, between-speaker variation in the use of implosives as opposed to voiceless unaspirated stops).

Tone acoustics were quantified with the same method used in previous studies of Wu varieties, e.g. [10 11]. A wideband spectrogram was generated in *Praat*, together with its wave-form and superimposed F0. The token's tonally relevant F0 was then identified, extracted with a *Praat* script, and modeled in *R* by an 8th order polynomial. This enabled F0 values to be sampled from the polynomial F0 curve with a sufficiently high sampling frequency (at 10% points of the curve as well as 5% and 95%) to capture the details of its time-course. Phonation type was quantified with *VoiceSauce* [12]. Interaction between tonal pitch and phonation type was modeled with an extended version of Fujisaki's *command-response* model [13 - 15].

3. Results

3.1. Auditory analysis

Table 1. Examples of Jinshan tones on monosyllabic words

Tone name	Example
high fall	pɔ 包*wrap,* pɪã 冰*ice*, sɔ 烧 *burn*, lɔ 捞 *carry*, tsʰu 搓 *roll*
low rise-fall	bi̯ã̯ 平 *flat*, ʑɛ̯ 裁 *cut*, li̯ɛ̯ 晾 *to dry*, lɑ̯ 狼 *wolf*
high level	tsʰɛ *to rub*, sɔ 扫 *to sweep* tsɔ 早*early*
(delayed) mid rise	tsɔ 罩 *cover*, tsʰɛ 踩 *to trample,* tɔ 到 *arrive*, tʰɔ 套 *sheath,* su 漱 *to rinse*
(delayed) low rise	ɦʊɛ̯ *to throw*, ʑɛ̯ 站 *to stand*, mã̯ 问 *to ask*
short stopped high	səʔ 塞 *block*, pɑʔ 剥 *peel*, tʰəʔ 脱 *take off,* vəʔ 勿 *not*
short stopped low rise	ŋi̯ɛ̯ʔ 热 *hot*, ʑə̯ʔ 直 *straight,* və̯ʔ 活 *live*

Conventional auditory phonetic and phonological analysis showed our speakers have seven tones, which can be named

after their pitch features as follows: *high fall*, *low rise-fall*, *high level, mid rise*, *low rise, short stopped high* and *short stopped low rise*. Table 1 gives some examples.

3.2. Acoustic description and tonological structure

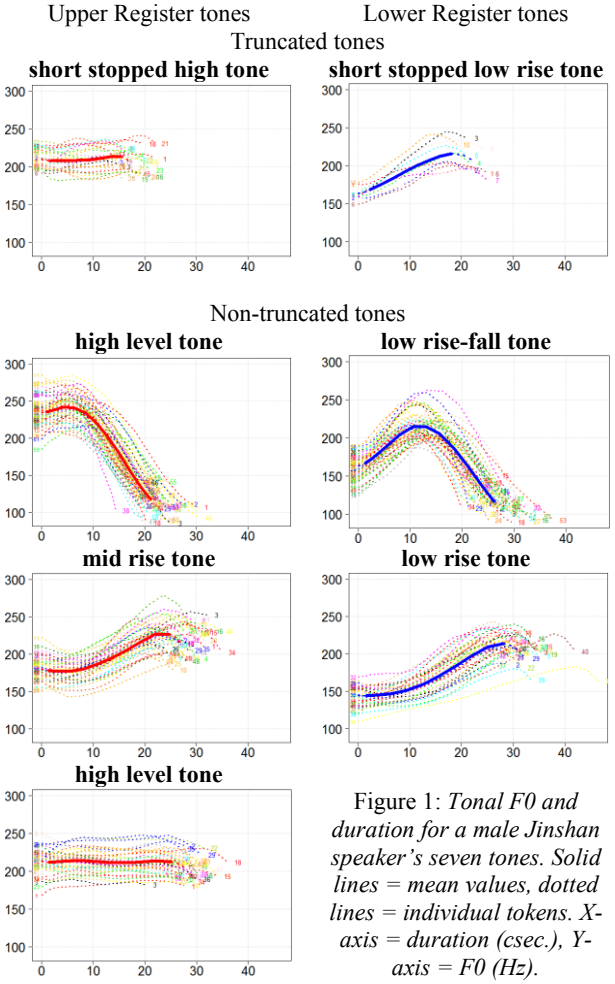

Figure 1: *Tonal F0 and duration for a male Jinshan speaker's seven tones. Solid lines = mean values, dotted lines = individual tokens. X-axis = duration (csec.), Y-axis = F0 (Hz).*

Figure 1 shows mean values for tonal F0, plotted as function of duration, for one of the male speakers. Individual tokens are also shown to give an idea of the amount of variation due to intrinsic factors. It can be seen that the tones' F0 shapes resemble their pitch descriptors fairly closely. It can also be seen that the short tones have about half to two-thirds of the duration of the long tones. The arrangement of the panels in figure 1 is important, as it shows how the seven tones are cross-classified in the typical Northern Wu manner by features of *truncation* and *register* [16]. Thus the truncated *short stopped high* and *short stopped low rise* tones have glottal-stop codas and shorter Rhymes, compared to the longer Rhymes, with gradual phonation offset, of the five other non-truncated tones.

For this paper, though, Register is the important dimension. Register partitions the tones into two sets. High fall, high level, mid rise and short high level belong to the upper register; low rise-fall, low rise and short low rise are lower register. Upper register tones are plotted on the left in figure 1, with their means in red; lower register tones on the right, with means in blue. Register correlates with many phonetic and phonological features. The most important of these, for this paper, are phonation type and depression (*depression* refers to the lowering of pitch onset as a function of phonetic and/or phonological factors producing, e.g., a rising-falling from a falling pitch [17 18]). The upper register tones have modal voice; the lower register tones have breathy voice and depressed pitch onset.

Register also correlates with the nature of Onset obstruents. Jinshan has the typical Wu three-way contrast within stop and affricate phonemes, e.g. /p^h t^h ts^h k^h, p t ts k, b d dz g/; and two-way contrast within fricatives, e.g. /f v, s z/. As in most Wu dialects, phonemic obstruent voicing is in complementary distribution with register: the voiceless phonemes co-occur with the upper register tones, and have the expected allophones e.g. /p/ → [p], /p^h/ → [p^h]. The voiced obstruent phonemes occur with the lower register tones and have different realisations conditioned by position in word: voiceless lenis word-initially, e.g. /b/ → [b̥] / #__ , and voiced word-internally, e.g. /b/ → [b] / V_V.

Register also correlates to a certain extent with overall pitch height: upper register tones have pitch contours mostly in the upper half of the pitch range and lower register tones have pitch contours mostly in the lower half of the pitch range. There is, however, considerable overlap between the high and low register tones' F0 values. In order to show this better, and to move from individual values to values representative of the variety, figure 2 is a plot of three speakers' normalised tones (z-score normalised F0 plotted against normalised duration [19 20]. The individuals' normalised F0 trajectories cluster fairly tightly except for the mid and low rise tones, for which there seems to be greater between-speaker variation: the female's trajectories rise immediately after onset, whereas the males have a delayed rise (so perhaps they should be kept separate). It can be seen that two thirds of the contour of the upper register mid-rise tone, and half the values of the high fall tone, lie below the mid-range value of 0; and the peak values of the lower register low rise-fall are above the mid-range value. This means one cannot define Register – at least as far as these varieties are concerned – in terms of location of pitch/F0 in the upper or lower half of the pitch/F0 range, as is commonly assumed [2].

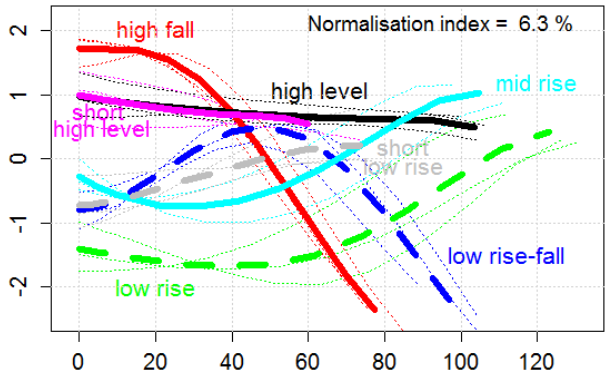

Figure 2: *Normalised values for the seven tones of three Jinshan speakers. Thick lines = mean normalised values. Thin lines = normalised values of individual speakers. Dashed lines = lower register tones.*

Figure 2 also shows how the lower register tones (low rise-fall, low rise, and short low rise) have the same F0 contours as the upper register high fall, mid rise and short high tones respectively, but with a depressed onset. The high level tone lacks a depressed counterpart. It will be shown below how this depression effect can be modeled.

The seven Jinshan tones are thus actually a constellation of pitch, phonation type, duration and segmental quality, and this is again typical of most conservative Wu varieties.

Some of these features are exemplified acoustically in Figure 3, with synchronous wide-band spectrograms and F0, using data from the female speaker. The top panel of figure 3 shows typical segmental and phonatory differences between the upper register high fall tone and lower register low rise-fall tone in words with bilabial stop Onset: [pɪã 51] 冰 *ice* and [b̥ɪ̰ã̰ 231] 平 *flat*. The upper register word has a voiceless unaspirated [p] as allophone of /p/ and has an expected very short VOT lag of less than 1 centisecond. The bilabial Onset in the lower register word is voiceless unaspirated lenis [b̥], which is the word-initial allophone of /b/. A longer duration of about 2 centiseconds between stop release and onset of phonation can be seen. This small difference in VOT, documented for several Wu dialects, presumably reflects a slightly greater distance between the arytenoidal vocal processes at stop release for [b̥] than [p], which in turn reflects the phonatory difference associated with upper and lower register.

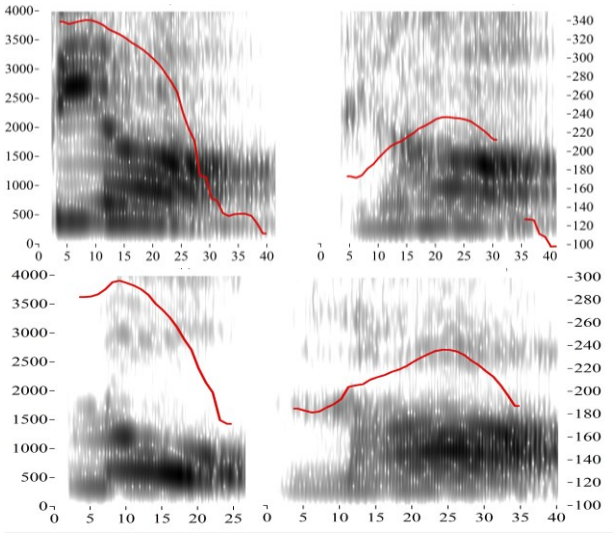

Figure 3: *Acoustics of phonation type and segmental structure differences associated with register differences. Top* = [pɪã 51] 冰 ice *(left) and* [b̥ɪ̰ã̰ 231] 平 flat. *Bottom* = [lɔ 51] 捞 carry *(left) and* [lɥ̩ 231] 狼 wolf. *X-axis = duration (csec.), y-axis left = spectral frequency (Hz), y-axis right = F0 (Hz).*

A far more salient difference between the upper and lower register words is the phonation itself: modal in upper register and breathy in lower. In the lower register word [b̥ɪ̰ã̰ 231] about the first third of the Rhyme – lasting to about the peak F0 point at csec. 20 – shows noisy periodicity, especially clear in the noise-excited F2. This is absent from the modally phonated upper register word (the abrupt spectral change at ca. csec.12 reflects the opening of the velum for the nasalisation in /ɪã/.)

It has often been assumed, e.g. [21 p.91, 22], that breathy voice in Wu is a function of syllable onset *stops*; but in Jinshan, and many other Wu varieties, it characterises all low register words irrespective of whether the Onset is stop, fricative, sonorant, glide or zero. As illustration, the bottom panel of figure 3 shows typical segmental and phonatory differences between the upper register high fall tone and lower register low rise-fall tone in words with *sonorant* Onset: [lɔ 51] 捞 *to carry* and [lɥ̩ 231] 狼 *wolf*. Once again, in the low register word, noisy periodicity is evident in the higher frequency regions (F3, F4) over the first 10 centiseconds after Rhyme onset, as well as during the /l/. Attenuated broadband energy extending from F2 downwards is also seen after Rhyme onset.

3.3. Quantification of phonation type differences

In order to quantify the phonation type differences associated with register, *VoiceSauce* was used to extract spectral slope measures expected to correlate with breathy *vs* modal phonation type. Common practice is to sample parameters at a single point in the tone's time course. It is more informative, however, to quantify the whole of the time course of the parameter. The methodology is described in [23].

Figure 4 shows the time course of two common phonation type parameters: the difference in amplitude between the fundamental and the second harmonic; and the difference in amplitude between the fundamental and the harmonic closest to the first formant centre frequency. *VoiceSauce* calls these "H1H2c" and "H1A1c", where *c* stands for corrected for vowel quality (i.e. using an all-pole LPC transfer function). To provide even tighter control, tokens with non-high vowel nuclei and non-nasal Onsets were used. This means that the estimation of the energy of the fundamental will not be compromised by its being in the vicinity of a low F1 associated with a high vowel or, with nasal Onsets, the lowest nasal formant.

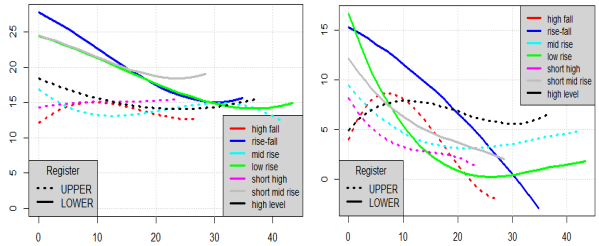

Figure 4: *Time course for VoiceSauce parameters H1A1c (left) and H1H2c for the seven tones of a female Jinshan speaker showing differences between high and low register tones. X-axis = duration (csec.); Y-axis = VoiceSauce parameter (dB).*

Figure 4 shows a clear difference in both parameters associated with register: low register tones have higher values at Rhyme onset than lower. Thus immediately after Rhyme onset the lower register tones have considerably more low frequency energy. This difference disappears towards the end of the Rhyme – more quickly for H1H2c than for H1A1c.

Now, H1A1c and H1H2c differences can be found intrinsically varying with F0 in tones without extrinsic phonation type, e.g. Cantonese [23]. The crucial finding for Jinshan (and other Wu dialects) is that the phonatory parameter is independent of F0. This can be seen by comparing the phonatory parameters of tones *of different register but similar F0*. Fortunately, Jinshan allows us to do this with the lower register rise-fall tone (plotted in blue) and the upper register mid rise tone (plotted in cyan). Figures 1 and 2 show that both these tones have similar F0 over the first 10 centiseconds of their Rhyme. Figure 4 shows their phonatory parameters are very different over this stretch, however: the low rise-fall tone has much greater lower frequency energy. This indicates that, rather than constituting an intrinsic accompaniment to a deliberate low pitch onset in

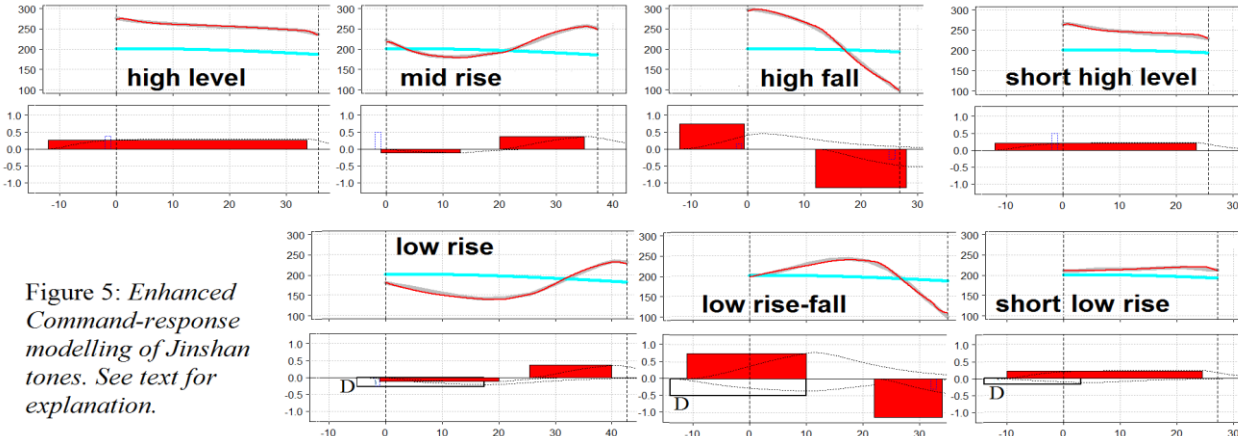

Figure 5: *Enhanced Command-response modelling of Jinshan tones. See text for explanation.*

the low register tones, it is actually the other way round. A low pitch onset is not the cause of the breathy phonation; it is the deliberate breathy voice phonatory setting which causes the low pitch onset. This in turn allows a tonological interpretation of the low register tones as having underlyingly *the same tonal target* as the high register tones but with an additional component which results in depression and breathy voice. The best guess as to the articulatory nature of this component is a constricted epilarynx, an insight of the *larynx-as-articulator* concept described in [24]. The following section shows how this can be modeled phonetically.

3.4. Modeling interaction of phonation type and tone

The interaction between phonation type and tone was quantified using Fujisaki's *command-response* (CR) model for tonal F0 [13 - 15 ch.3]. This model, which has been applied to several tone languages including Mandarin, Cantonese and Shanghai [25 - 27], factors the time-varying F0 into two types of component, both modelled as impulses of given amplitude and duration. A *tonal* component represents the response of the speech production mechanism – in this case the *pars recta* of the crico-thyroid – to impulse commands for implementing tone. The second, or *phrasal*, component represents a much slower time-varying response and accounts for more gradual, declinational, change in F0 throughout an utterance or intonational phrase.

As it stands, the CR model does not allow for anything other than the phrasal and tonal commands, and the tonal commands must not overlap. In order to model the effect of the phonation type in lower register tones, and its interaction with tone, we significantly modified the model to incorporate an additional component which, because it is produced by a different mechanism, can overlap with tonal commands. We have called this a depression component because it models the effect of lowering the F0 at the onset of the Rhyme.

Figure 5 shows the CR modelling of the seven Jinshan tones of the female speaker. Each tone is represented vertically in two panels. The bottom panel shows the impulses and impulse responses: tonal impulses are shaded in red; depression impulses are unshaded and marked with a "D"; impulse responses are plotted in a thin line. The top panel shows, in red, the F0 predicted from the summation of the impulse responses. The speaker's actual tonal F0 is plotted in grey, but is difficult to see as the fit between predicted and observed F0 is very good: mean squared errors range from 3.5 Hz for low rise-fall, down to 1.5 Hz for high level. The cyan line represents the response of the system with just the phrasal command.

It can be seen that the speaker's high level tone is modelled with a single tone impulse, its slight F0 decay accounted for by the decay in the phrasal command. The short high tone is similar, but without such a long tonal command. The high fall tone is modelled with an early positive and a later negative tonal command. The mid rise tone has the opposite arrangement, with an early slightly negative and a late positive tonal command. These results are typical.

It can be seen in figure 5 that each of the lower register tones has been generated *with tonal impulses of the same amplitude as their high register counterparts* plus an additional depression impulse. (The relative *timing* of the impulses has to be such to accommodate the durational differences between the upper and lower register tones, and is not exactly the same.)

This CR analysis-by synthesis buttresses a tonological analysis of the seven tones which allows upper and lower register pairs of tones to be represented with the *same* tonal target, whilst differing in a depression component. Thus the high fall and low rise-fall tones can be analysed as sharing a falling [HL] tonal target, but with the low rise-fall tone having in addition a depression component: [HL, D]. Moreover, the CR analysis supplies an obvious productional interpretation of the tonological constructs H and L as crico-thyroid and strap muscle activity, with D as epilaryngeal constriction. The same analysis applies *mutatis mutandis* for the other tones. Mid rise [MH] and low rise [MH, D] tones share a rising [MH] tonal target, and short high [H] and short low rise [H, D] tones both have a [H] tonal target. The fact that the lower register tones can be shown to have the same tonal target as their upper register counterparts also helps explain the fact that they pattern together in Jinshan's typically complex tone sandhi.

4. Summary

This paper has described the seven isolation tones of Jinshan, a dialect where tone is not just pitch but a constellation of pitch, phonation type, duration and segmental effects. Quantification of the phonation type differences associated with register using conventional spectral slope parameters showed that they were extrinsic and not a function of F0. An extended version of the CR model was then used to demonstrate how the lower register tones can be modeled with the same tonal commands as the upper register tones, but with the addition of a depression component.

5. Acknowledgements

Many thanks to all our Jinshan informants, but especially to teacher 卢迅, for taking their time to read out such a long list so carefully for us. Thanks also to our three anonymous reviewers for some extremely useful comments: we have restructured the paper to take many of these into account.

6. References

[1] Henderson, E., "The topography of certain phonetic and morphological characteristics of South East Asian languages", *Lingua* 15: 400-434, 1965.
[2] Yip, M., Tone, CUP, 2002.
[3] Hyman, L., "Word Prosodic Typology", Phonology 23: 225-257, 2006.
[4] Chao Y. 趙元任，現代吳語的研究 Studies in the Modern Wu Dialects, Tsing Hua College Research Institute Monograph 4, 1928.
[5] Qian N. 钱乃荣，当代吴语研究 [Studies in the Contemporary Wu Dialects], Shanghai Educational Press, 1992.
[6] Zhang J., A Sociophonetic Study on Tonal Variation of the Wúxī and Shànghǎi Dialects, LOT Netherlands Graduate School of Linguistics, 2014.
[7] Gao, J-Y. and Hallé, P., "Are Young Male Speakers Losing Tone 3 Breathyness in Shanghai Chinese? An Acoustic and Electroglottographic Study." Proc. 2nd International Congress on the Phonetics of the Languages in China, 163-166, 2013.
[8] Editorial office for linguistics teaching and research, Hanyu Fangyan Cihui 漢語方言詞彙 [Chinese Dialect Vocabulary], 文字改革出版, 1964.
[9] http://philjohnrose.net/Wu_tones/index.html
[10] Rose, P., "Complexities of Tonal Realisation in a Right-Dominant Chinese Wu dialect – Disyllabic Tone Sandhi in a Speaker from Wencheng", Journal of the South East Asian Linguistics Society 9: 48-80, 2016.
[11] Shen R. and Rose, P., "Preservation of Tone in Right-Dominant Tone Sandhi: A Fragment of Disyllabic Tone Sandhi in Maodian Wu Chinese", in C. Carignan & M. Tyler [Eds] *Proc. 16th Australasian Int'l Conf. on Speech Science & Technology*: 345-348, Sydney, 2016.
[12] Shue, Y.L., Keating, P., Vicenik C. and Yu, K., "VoiceSauce: A Program for voice analysis", in *Proc. 17th Int'l Congress of Phonetic Sciences*, Hong Kong: 1846-1849, 2009.
[13] Fujisaki, H., "Dynamic Aspects of Voice Fundamental frequency in Speech and Singing", in F. MacNeilage [Ed], The Production of Speech, Springer: 39-55, 1983.
[14] Fujisaki, H., "In Search of Models in Speech Communication Research." Proc. *INTERSPEECH,* Brisbane, Australia: 1-10, 2008.
[15] Mixdorff, H., *Intonation Patterns of German – Model-based Quantitative Analysis and Synthesis of F0 Contours*, Ph.D. TU Dresden, 1998.
[16] Rose, P**.,** "Tonation in Three Chinese Wu Dialects"**,** *Proc. Int'l Congress of Phonetic Sciences*, (no page numbers), Glasgow, 2015.
[17] Rycroft, D., "Tone in Zulu Nouns", African Language Studies 4: 43-68, 1963.
[18] Rose, P., "Independent depressor and register effects in Wu dialect tonology: Evidence from Wenzhou tone sandhi", *Journal of Chinese Linguistics* 30(1): 39-81, 2002
[19] Rose, P., "Considerations in the normalisation of the fundamental frequency of linguistic tone", Speech Communication 6(4): 343-352, 1987.
[20] Rose, P., "Comparing Normalisation Strategies for Citation Tone F0 in Four Chinese Dialects", in C. Carignan & M. D. Tyler [Eds], *Proceedings 16th Australasian Int'l Conf. on Speech Science & Technology*, Sydney: 221-224, 2016.
[21] Ramsey, S., The Languages of China, Princeton University Press, 1987.
[22] Cao J. and Maddieson, I., "An exploration of phonation types in Wu dialects of Chinese", Journal of Phonetics 20: 77-92, 1992.
[23] Rose, P., "Variation in Spectral Slope and Interharmonic Noise in Cantonese Tones", Proc. *INTERSPEECH,* Shanghai, 2020.
[24] Esling, J., Moisik, S., Benner, A. and Crevier-Buchman, L., Voice Quality – The Laryngeal Articulator Model, CUP, 2019.
[25] Fujisaki, H., Hirose, K., Hallé, P. and Lei H., "Analysis and modeling of tonal features in polysyllabic words and sentences of the Standard Chinese", Proc. *ICSLP 1990, Kobe, Japan:* 841-844, 1990.
[26] Fujisaki, H., Ohno, S. and Gu W., "Physiological and Physical Mechanisms for Fundamental Frequency Control in Some Tone Languages and a Command Response Model for generation of Their F0 Contours", *International Symposium on Tonal Aspects of Languages*, Beijing, 2004.
[27] Gu W., Hirose, K. and Fujisaki, H., "Analysis of Shanghainese F0 Contours based on the Command-Response Model", Proc. ICSLP: 81-84, 2004.