Interspeech 2018
2-6 September 2018, Hyderabad

Pitch or phonation: On the glottalization in tone productions in the Ruokeng Hui Chinese dialect

Minghui Zhang[1], *Fang Hu*[2]

[1]Si-Mian Institute for advanced Studies in Humanities, East China Normal University
[2]Institute of Linguistics, Chinese Academy of Social Sciences

goldenzhangminghui@gmail.com, hufang@cass.org.cn

Abstract

This paper examines the interplay of glottalization and tones in tonal phonology of the Ruokeng Hui Chinese. Acoustic data from 10 native speakers were analyzed in terms of pitch (F_0), duration, H_1-H_2, H_1-$A_{1/2/3}$, CPP, HNR, SHR, etc. Fine-grained phonetic details reveal the interactions between phonation and tones and shed light on the ongoing tonal change from a glottalized tone to a plain high falling tone in the Ruokeng dialect.

Index Terms: pitch, phonation, glottalization, Ruokeng Hui Chinese, tonal change

1. Introduction

How tones distinguish with each other is an essential issue in the phonetics and phonology of tones ([1], [2], [3], [4]). It is well-known that tones are correlated with pitch, but languages may differ in linguistic utilization of pitch. For instance, tones in African languages usually distinguish in pitch height and are therefore termed as register tones in the literature ([5]), whereas tones in Chinese and southeastern Asian languages are dynamically complex and are thus termed as contour tones ([4]). And additionally, phonation may interact with pitch and consequently contribute to tonal phonology in languages. On the one hand, phonation could enhance pitch targets, especially those extreme ones. For instance, creaky voice could be a concomitant phonatory setting to facilitate an extreme low pitch target, as is commonly observed especially in female speakers in the production of the dipping tone in Mandarin Chinese; in contrast, falsetto could enhance the realization of an extreme high pitch, as is observed in Hui Chinese ([6]). On the other hand, phonation could be pitch-independent and directly serve as tonal targets in languages ([7], [8]). Furthermore, interactions between pitch and phonation could induce tonal variations ([9]). In summary, it has been widely acknowledged that tonal contrasts in languages could be multidimensional ([7], [10], [11]). And thus a phonetic theory of tones should provide enough dimensions for descriptive adequacy of potentially unbounded number of tones within a restricted range and to satisfy the requirements of general phonetic and phonological comparability of tones ([12]).

This paper examines the interaction of pitch and phonation in tone productions in the Ruokeng Hui Chinese dialect. There are 6 tones in the Ruokeng: T1 (*Yin-ping*), T2 (*Yang-ping*), T3 (*Shang*), T4 (*Yin-qu*), T5 (*Yang-qu*) and T6 (*Ru*), among which T4 is characterized by a special voice quality of glottalization and has been transcribed as a dipping tone in previous impressionistic dialectological work ([13]).

This paper adopts a gradient view of glottalization ([14]), which includes different types of creaky voice ([15]) and full or incomplete glottal stops in the continuum of phonation ([16], [17]). Variations of glottalization were described on the basis of fine-grained acoustic phonetic details. The aim is to explore the interactions between phonation and pitch in tone productions. How pitch and phonation characterize the production of T4 in Ruokeng? Which is essential in tone productions? Pitch or phonation? Are these variations and interactions subphonemic that manifests phonetic granularity in tone production? Or should they be understood as part of tonal phonology?

2. Methodology

10 native speakers, 5 male and 5 female, provided speech data. Five meaningful monosyllabic words in CV syllables, where C is a stop, are selected as the test words for each tone. The test word X is embedded in a carrier sentence [X_1. ŋa35 tʰu33 X_2 fɛ̃31 n̩35 tʰɛ̃31] (X_1, I read X_2 for you to listen). That is, there are two contexts for each test word X: the citation form X_1 and sentence-mid X_2. This paper focuses on citation forms due to the space limit. The sampling rate is 11,025 Hz. 5 repetitions were recorded into a laptop PC through an E-MU 0404 USB sound card with a SHURE SM86 microphone.

Annotations were completed in praat 6.0.36 ([18]). The rime was labelled as the tone-bearing unit (TBU). 5 equidistant sampling chunks were identified for cases of Jitter and Shimmer; otherwise, 11 equidistant sampling chunks were identified. And then, F_0, H_1*, H_1*-H_2*, H_1*-$A_{1/2/3}$* (asterisks indicate that the harmonic/spectral amplitudes are calibrated by formant corrections), CPP, $HNR_{05/15/25/35}$, SHR, Jitter and Shimmer are extracted by praat or VoiceSauce ([19]) on each chunk.

Normalization procedure were applied to F_0 and other acoustic parameters, in order to minimize anatomical and physiological variations in gender and speaker while still preserve phonological as well as intrinsic variations of tones. A Revised D-value Procedure ([20]) was employed to normalize the F_0 data into Chao's five-digit tone scale ([21]). As postulated in Formula (1), ST*i* is the semitone value of each sampling F_0 point, while STmin and STmax are the minimal and maximal semitone values within the modal-voiced pitch range of each speaker, respectively. And for the other acoustic parameters, Z-score Standardization was applied.

$$Di = 0.5+(8\times STi-9\times STmin+STmax)/(2\times(STmax -STmin)) \quad (1)$$

3. Results

3.1. Pitch and phonation: phonetic details

10.21437/Interspeech.2018-1638

Figure 1 shows the pitch curves for Ruokeng tones in terms of mean D-values ±1 standard deviation (SD), wherein normalized pitch heights, which are comparable to Chao's tone letters, are labeled on the ordinates, and the 11 time points in percentage are calibrated along the abscissas. The pitch curve for T4 is drawn alone in the right, while the pitch cures for the other tones, T1, T2, T3, T5 and T6, are grouped together in the left. Table 1 summarizes mean SDs of each tone of 10 speakers. Figure 1 reflects inter-speaker difference in tone production, and Table 1 indicates inter-token variance of each speaker.

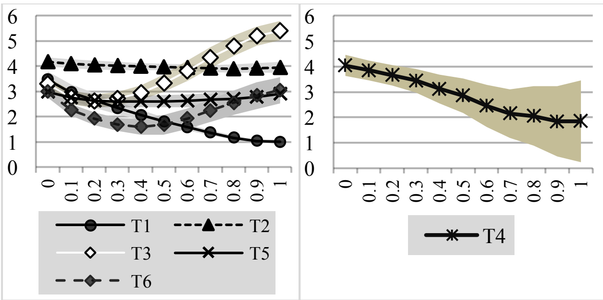

Figure 1: *Pitch curves (means ±1 SD) along the normalized time series for the 6 tones in Ruokeng: T4 (right) and the other 5 tones (left).*

Table 1: *Mean SDs of D-values for each tone from 5 male (M1~M5) and 5 female speakers (W1~W5).*

	M1	M2	M3	M4	M5	W1	W2	W3	W4	W5
T1	0.27	0.50	0.21	0.37	0.44	0.31	0.35	0.39	0.44	0.28
T2	0.36	0.49	0.27	0.24	0.41	0.32	0.44	0.27	0.49	0.30
T3	0.29	0.51	0.28	0.36	0.46	0.38	0.35	0.35	0.44	0.33
T5	0.29	0.33	0.32	0.32	0.39	0.37	0.37	0.34	0.32	0.36
T6	0.33	0.76	0.42	0.49	0.40	0.35	0.53	0.30	0.44	0.27
T4	**0.91**	**0.88**	**0.79**	**0.62**	**0.76**	**1.06**	**0.61**	**0.84**	**0.96**	**0.77**

It can be seen from the left panel in Figure 1 that these 5 tones are characterized by distinctive pitch contours, namely 2 level tones (T2 [44] and T5 [33]), 1 falling tone (T1 [31]), 1 rising tone (T3 [35]), and 1 concave tone (T6 [323]). Moreover, all these 5 tones have a limited SD values within 0.5, suggesting a stable control in tone production. In contrast, the production of T4 is quite different. T4 could be described as a high falling tone [42] according to the mean D-value curve, rather than a dipping tone as transcribed in dialectological work ([13]). But T4 has greater inter-speaker variations, as indicated by the shade in the right panel of Figure 1; also, T4 has inter-token variance with mean SDs greater than 0.5 in all speakers, indicating a worse control of pitch production vis-à-vis the other 5 tones. In other words, there is certain overlap between the two falling tones T4 [42] and T1 [31] in terms of pitch contour. It is then of great interest to examine the phonetic details of T4 and its interaction with concomitant glottalization, and to further determine what role the glottalization plays in the characterization of tones.

Figure 2 shows the spectrograms of the test words [pa], [pʰa] or [pɛ] with different tones. Note that all the 9 cases were extracted from the X2 position in carrier sentences. Glottalization on target syllables could be examined clearly, as the target TBU syllables are followed by a fricative [f]. It can be seen clearly from T4a/b/c/d in Figure 2 that irregular, highly-damped pulses occur in syllable-final position, signifying complete/incomplete glottal closures. The four typical cases of T4a/b/c/d demonstrate a gradient scale of glottalization in the production of T4: T4a and T4b are both creaky voice /̰ /, and T4c and T4d are incomplete and complete glottal stops /ʔ̰/ and /ʔ/ respectively. And the tone values for these T4 tokens were labeled as [40] in the figure, where [0] indicates an extreme low pitch target. In contrast, the other 5 tones are realized on syllables with regular/periodic voicing in general.

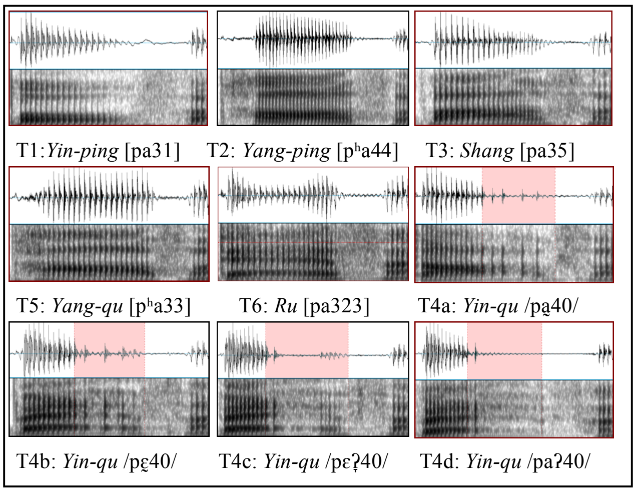

Figure 2: *Spectrograms of the Ruokeng tones in carrier sentence: typical cases from a male speaker.*

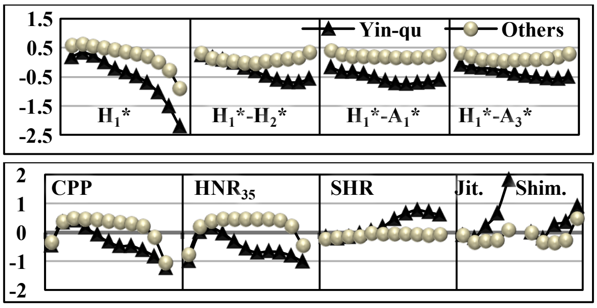

Figure 3: *The 9 acoustic parametric curves for T4 (Yin-qu: black) and the mean of the other 5 tones (grey) in Ruokeng.*

The measured 9 acoustic parameters, H_1*, H_1*-H_2*, H_1*-A_1*, H_1*-A_3*, CPP, HNR_{35}, SHR, Jitter and Shimmer, confirm an apparent difference in phonation between T4 and the other 5 tones in Ruokeng. As shown in Figure 3, the 9 acoustic parametric curves exhibit different patterns for T4 (*Yin-qu*: black) and the mean of the other 5 tones (grey). The ordinates represent the normalized Z-score of each parameter; and the abscissas represent the normalized time series that mark the sampled time points in percentage. As compared to the mean of the other 5 tones, T4 demonstrates comparatively smaller contact quotient and milder abruptness of closure in the glottis (lower H_1-related measures), stronger noise energy (lower CPP and HNR_{35}), and more irregular vibrations (higher SHR, Jitter and Shimmer but lower CPP). These are all typical acoustic properties of glottalization ([7]). That is, there is a clear difference in phonation between T4 and the other 5 tones: namely glottalized vs. modal voice.

In summary, the production of T4 is heavily influenced by glottalization, which results in types of creaky voices (prototypical creak, multiply pulsed voice, aperiodic voice, etc.) or incomplete/complete glottal stops, and induces unstable control in F_0 production, problematic pitch trace in acoustic measurement, and ambiguity in pitch perception ([9], [15]).

3.2. Pitch or phonation: phonology

It has been shown so far that the falling tone T4 in Ruokeng is characterized by a non-modal phonation, i.e. glottalization. It is quite natural to suppose that glottalization would also play a role in tonal phonology such as distinguishing T4 from the other falling tone T1. But which one is phonologically essential? Pitch or phonation? Or both? As an inquiry into this issue, Linear Discriminant Analyses (LDA) based on Fisher Method were conducted on pitch and the 9 acoustic parameters of phonation: the normalized Z-score of semitone, H_1*, H_1*-H_2*, H_1*-A_1*, H_1*-A_3*, CPP, HNR_{35}, SHR, Jitter and Shimmer. And each LDA exclusively select one type of the parameters, within which all the 11 or 5 sampling points are defined as the independents. As a result, the 3 most accurate parameters in classifying the 6 tonal categories are listed in Table 2.

Table 1: *The 3 most effective LDA parameters in classifying the 6 tones in Ruokeng: ST stands for semitone and Jit. for Jitter.*

T1		**T2**		**T3**		**T5**		**T6**		**T4**		**Mean**	
ST	96	ST	96	ST	97	ST	94	ST	86	**H_1**	**76**	ST	89
H_1A_1*	78	Jit.	74	HNR_{35}	73	H_1*	57	H_1*	65	**Jit.**	**71**	H_1*	64
H_1A_2*	77	CPP	68	HNR_{25}	71	HNR_{15}	56	CPP	62	**CPP**	**68**	H_1H*_2	62

First, pitch is the most important cue for the classification of the 5 modal-voiced tones and nearly 90% of all tonal samples are correctly sorted by using semitone values alone. Second, the 3 most effective parameters (H_1*, Jitter and CPP) in discriminating T4 are all phonation-related, which manifest the 2 essential properties (constricted glottis and irregular voicing) in specifying glottalization. And pitch is not a valid cue for the classification of T4, which explains only 62% of the data.

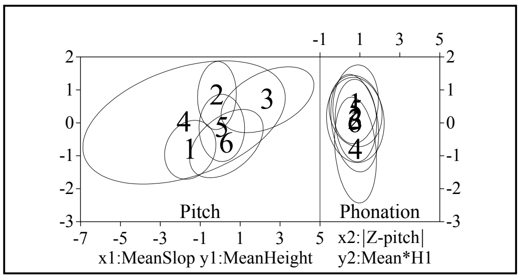

Figure 4: *Pitch (left) and phonation contrasts (right) of the six tones in Ruokeng both in terms of confidence ellipses specified by two-dimensional features. The number of scatter plots for each tonal ellipse is 250.*

The effectiveness of pitch and phonation for the classification of tonal categories was graphically represented in Figure 4. Both the pitch (left) and phonation (right) contrasts of Ruokeng tones can be represented in forms of 95% confidence ellipses determined by two parameters. Pitch contrast is specified by mean slope (x1) and mean height (y1) of pitch curves, while phonation contrast is quantified by the absolute value of Z-score (x2) and the mean value of H_1* (y2).

The contrast pattern in Figure 4 resembles territorial maps of LDA for the classification of Ruokeng tones. It confirms the fact that pitch and phonation make different contributions to the classification of different tones. As shown in the left panel, 5 of the 6 tones can be generally discriminated from each other by pitch-related features; the only exception is T4, whose ellipse is scattered. In the right panel, only the ellipse of T4 can be discriminated from the other 5 tones that are heavily overlapped with each other.

In summary, glottalization is responsible for specifying T4, while pitch is essential for the other tones. The LDA results from Ruokeng tones reflect a general linguistic preference in tonal contrasts: pitch is utilized to distinguish the majority of tones, whereas phonation is limited to certain specific tonal categories ([7]).

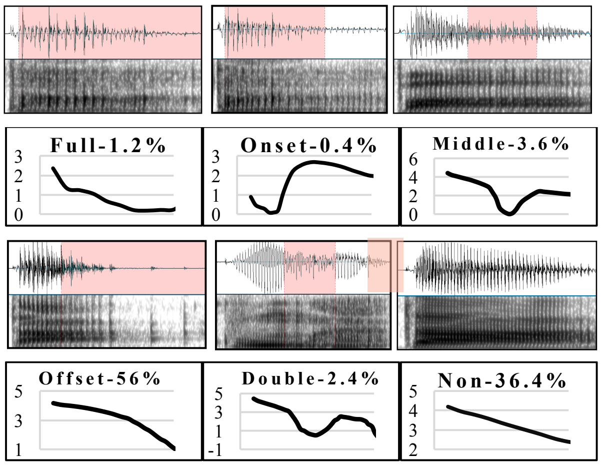

Figure 5: *Varieties of glottalized timings and the corresponding pitch contours for T4: [ta4] ('bring') or [kuːɐ4] ('past') from W1, W5, W1, M1, M2 and W2.*

It is mentioned above in Figure 2 in 3.1 that glottalization is gradient in nature. A related issue is that glottalization also has great variations in timing and has a further consequence on deviations of pitch contours as well. As shown in Figure 5, spectrograms (in 1st and 3rd panels) and the corresponding pitch contours (in 2nd and 4th panels) illustrate 6 types of glottalized timings that are representative for the total 250 samples: full-glottalization, onset-glottalization, mid-glottalization, offset-glottalization, double-glottalization, and non-glottalization (modal-voice), respectively. The glottalized part is designated by the rectangular shadows in each spectrogram; the frequency of each type is expressed as percentage of the total samples and is labelled in the title for each pitch diagram. Variations in timing suggest that speakers can utilize different articulatory strategies for glottalization. The most common case is a glottalized offset. And the second most common case is non-glottalization: there are nevertheless 36.4% of the samples that are not glottalized. That is, they are realized as a plain falling tone within modal voice. This means that pitch contours could be an alternative in the production of T4. It further implies that pitch and glottalization both could be options for the realization of a phonological low tonal target, as the glottalized part of T4 in Ruokeng always co-occurs with low or extremely low F_0, which is a typical but not necessary property for glottalization ([15], [22]).

3.3. The interaction between pitch and phonation: an ongoing tonal change

A closer examination of data reveals that observed variations on T4 production is not purely irregular, free synchronic variations. Rather, they are age-related. A regular pattern emerges when variations are correlated with speakers' ages. That is, it seems T4 is chronically changing from a glottalized tone to a plain high falling tone. Figure 6 compares generational differences in terms of average pitch curves (means ± 1 SD) of T4 among the 3 age groups: 3 old (＞55 yrs), 4 middle-aged (40~48 yrs), and 3 young (20~36 yrs). It shows that a glottalized pitch production in old generation is changing gradually into a smoother curve in younger generations.

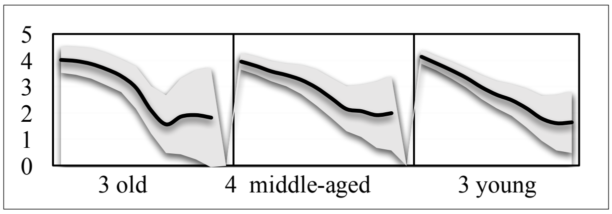

Figure 6: *Generational differences in average pitch curves (means ± 1 SD) of T4 for the 3 old (left), 4 middle-aged (mid) and 3 young speakers (right).*

The process of tonal change can be quantified in terms of the following 5 parameters. Parameter 1: accuracy of pitch (semitone) in LDA; Parameter 2: accuracy of all phonation-related measures in LDA; Parameter 3: proportion of modal-voiced samples of T4 in the samples of each speaker; Parameter 4: mean SD of the entire pitch curves of T4; Parameter 5: mean pitch slope for the modal-voiced part.

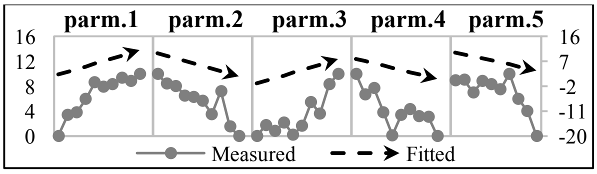

Figure 7: *Generational tonal changes of T4 represented the measured scatter curves and the linear-fitted arrow lines of 5 parameters (pram.1~5).*

Figure 7 shows generational changes of the 5 parameters in terms of scatter curves (grey) and their linear-fitted arrow lines (black). All the measured curves have been rescaled into a unified range [0, 10] marked along the ordinate on the left; and the fitted arrow lines which hint the tendency of generational changes are labelled according to the ordinate on the right. The 10 speakers are arranged along each abscissa in a descending order of age: $W1_{66}$, $M2_{62}$, $W4_{54}$, $M5_{48}$, $M4_{46}$, $W5_{45}$, $M1_{42}$, $M3_{40}$, $W3_{36}$ and $W2_{20}$. And it should be noted that the 5 female speakers are from the same family or neighbour who live together (W1-grandma, W4-mother, W3-aunt, W2-daughter, and W4-neighbour). The rising curve in pram.1 and the falling curve in pram.2 mean increasing and decreasing contributions of pitch and glottalization in the classification of T4, respectively. The rising curve in pram.3 indicates more and more modal-voiced samples detected in younger speakers. And the falling curves in pram.4 and pram.5 suggest more T4 samples have falling contours with a better controlled articulation. In conclusion, the production of T4 in Ruokeng is transforming from a glottalized tone into a plain high falling tone.

The histograms in Figure 8 compare the plain falling tone T1 (grey), level tone T2 (white) and the modal-voiced portion of T4 (black) in terms of pitch slope (increasing from left to right along the x-axes) and the slope frequency (the baselines of each tone are calculated from the trisection or quartering points along the y-axes). The values skewness (SK), kurtosis (KT) and S-W tests listed on the right-bottom in the figure were adopted to quantify the distribution modes of T4. A wave-like shift can be clearly observed from a comparison of the distributions of T4 from W1 to W2: normal > left-skewed > bimodal > right-skewed > normal. And, the values of SK, KT and S-W tests also confirmed this observation, indicating that the pitch curves in T4 is becoming more and more similar to a falling contour from elder speakers to younger speakers.

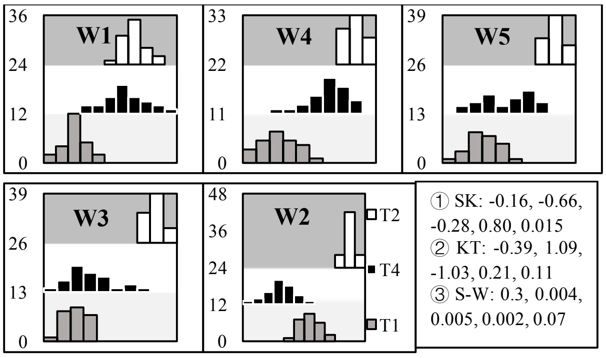

Figure 8: *Generational changes of the pitch slope in the modal-voiced portion within T4 (black) by reference to the distribution modes of the falling tone T1 (grey) and the level tone T2 (white).*

4. Conclusions

Fine grained phonetic details explained interactions between pitch and phonation in the production of Ruokeng tones.

Phonetically, tones with modal voice are characterized by pitch contours, whereas the production of the glottalized tone T4 in Ruokeng is characterized by a high falling contour with concomitant glottalization, which is gradient in nature and is variable in timing. Phonologically, pitch is the essential feature for tonal contrasts, whereas glottalization contributes to the specific tone that it influences.

The interaction between glottalization and pitch influences the production of T4 in Ruokeng in several aspects. On the one hand, glottalization co-occur with low or extremely low pitches, and thus could function as a phonological low tone target. On the other hand, glottalization induces instability in pitch production and ambiguity in pitch perception, and thus could trigger sound change in long term.

5. Acknowledgements

This work is supported by CASS Innovation Program and Chinese National Social Science Foundation (Project number: 15BYY073; P.I.: Dr. HU Fang).

6. References

[1] W. S-Y. Wang, "Phonological features of tone," *International Journal of American Linguistics*, 33, 93-105, 1967.

[2] V. A. Fromkin (ed.), *Tone: A Linguistic Survey*, New York: Academic Press, 1978.

[3] I. Maddieson, "Universals of tone," in J. H. Greenberg (ed.) *Universals of Human Language: Phonology, 2*, Stanford: Stanford University Press, 1978.

[4] M. Yip, *Tone*, Cambridge: Cambridge University Press, 2002.

[5] D. Odden, "Tone: African languages," in Handbook of Phonological Theory, Oxford: Basil Blackwell, 1995.

[6] M. Zhang, F. Hu, "Tone features in Qimen Hui Chinese Dialect," in *18th International Congress of Phonetic Sciences Proc.*, *August 10-14, Glasgow, Scotland*, 2015.

[7] J. Kuang, *Phonation in Tonal Contrasts*. Ph. D. dissertation, University of California, Los Angeles, 2013.

[8] J. Kuang, "The tonal space of contrastive five level tones," *Phonetica*, vol. 70, no.1-2: pp. 1-23, 2013.

[9] K. M. Yu, "The tonal space of contrastive five level tones," in *INTERSPEECH 2010-11th Annual Conference of the International Speech Communication Association, September 26-30, Makuhari, Chiba, Japan*, 2010, pp. 1529-1532.

[10] J. T. Gandour and R. A. Harshman, "Crosslanguage differences in tone perception: A multidimensional scaling investigation," *Language and Speech*, vol. 21, no. 1, pp. 1-33, 1978.

[11] P. Keating, J. Kuang, C. Esposito, M. Garellek and S. Khan, "Multi-dimensional phonetic space for phonation contrasts," Poster presented in *13th Conference on Laboratory Phonology*, *July 27-29, Stuttgart, Germany,* 2012.

[12] S. R. Anderson, "Tone features," in V. A. Fromkin (ed.) pp. 133-175, 1978.

[13] L. Wang, "The Phonetic System of Ruokeng Dialect," *Journal of Huangshan University*, Vol. 12, No.2, pp. 17-20, 2010.

[14] M. Garellek, *Production and perception of glottal stops*. Ph. D. dissertation, University of California, Los Angeles, 2013.

[15] P. Keating, M. Garellek and J. Kreiman, "Acoustic properties of different kinds of creaky voice," in *18th International Congress of Phonetic Sciences Proc.*, *August 10-14, Glasgow, Scotland*, 2015.

[16] P. Ladefoged, *Preliminaries to linguistic phonetics*, Chicago: University of Chicago, 1971.

[17] M. Gordon, & P. Ladefoged, "Phonation types: a cross-linguistic review," *Journal of Phonetics*, 29, 383-406, 2001.

[18] P. Boersma, D. Weenink, *Praat: doing phonetics by computer* [Computer program], Version 6.0.36, retrieved 11 November 2017 from http://www.praat.org/.

[19] Y.-L. Shue, *The voice source in speech production: Data, analysis and models*. Ph. D. dissertation, University of California, Los Angeles, 2010.

[20] F. Ling, "Comparisons on Normalization Algorithms for Cross-dialectal Tone Researches," *Studies on Wu Chinese Dialects*, edited by Z. Chen. Shanghai: Shanghai Educational Publishing House, pp. 47-52, 2016.

[21] Y. R. Chao, "A System of Tone-Letters," *La Maitre Phonetique*, vol.45, pp. 24-47, 1930. Reprinted in *Fangyan*, vol.2, pp. 81-82, 1980.

[22] M. Blomgren, Y. Chen, L. Ng. Manwa, H. R. Gilbert, "Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers," *The Journal of the Acoustical Society of America*, Vol. 103, No.1, Pt. 1, pp. 2649-2658, 1998.