Tonal Aspects of Languages 2016
24-27 May 2016, Buffalo, New York

An Interplay Between F0 and Phonation in Du'an Zhuang Tone

Jeremy Perkins [1], *Seunghun Lee* [2], *Julián Villegas* [1]

[1] University of Aizu, Aizu, Japan
[2] International Christian University, Tokyo, Japan
and University of Venda, Thohoyandou, South Africa

jperkins@u-aizu.ac.jp, seunghunjlee@gmail.com, julian@u-aizu.ac.jp

Abstract

This paper undertook an acoustic study of the tone system of one speaker of Du'an Zhuang, finding that his tone system involved phonation contrasts in addition to F0 contrasts. It was found that two of the six tones in unchecked syllables involved significant creakiness near the midpoint of the vowel. These results suggest a phonological tone contrast that involves both F0 and creakiness. Among pairs of tones that differed in their phonation, significant differences in the timing of F0 fall were discovered. Additionally, the two creaky tones differed in the timing of the maximum creakiness. Future research could establish whether and to what extent Du'an Zhuang speakers utilize creakiness and F0, and their relative timing, in discerning between tonal categories. Production studies are needed with larger samples of speakers and perception studies are needed to confirm the contrastive roles of phonation and F0 in the Du'an Zhuang tone system.

Index Terms: Creakiness, phonation, tone, Zhuang

1. Introduction

This paper illustrates the importance of considering phonation in tonal studies of languages via a specific example with a single speaker of the Du'an dialect of the Zhuang language, a member of the Tai-Kadai family spoken in southern China. The paper constitutes a production study where both F0 and a composite acoustic measure of creakiness are together taken to distinguish tonal categories.

Previous work has been done on Wuming Zhuang, the standard variety, which is spoken in the Yongbei (邕北) region. Wuming Zhuang belongs to the Hongshui He (红水河) group as a variety of Northern Zhuang. Wei and Qin [1] have established that Wuming Zhuang employs a system of tonal contrast in which F0 contours alone can fully describe the tone system. Hansen and Castro [2] and Kullavanijaya [3] have reported that dialectal variations of Zhuang are so vast that various dialects are mutually unintelligible.

We explore the possibility that the tonal system of Du'an Zhuang involves contrasts in both F0 and creaky phonation. While pitch has been at the center of describing tone, studies have noted that phonetic correlates such as phonation and/or duration are also important in characterizing tone in general (see [4] for discussion). Many authors have offered proposals of Chinese languages where phonation contrasts are involved [5, 6, 7].

In the Cambodian language, a report on 'register' appears, in which one register is described as being associated with a breathy or sepulchral voice and lower pitch, and another register with a normal voice with higher pitch [8]. DiCanio [9] provides a detailed phonetic study of the register system in Chong, a Mon-Khmer language. The interplay between phonation and tone has been reported in various other languages, including Burmese [10], Green Mong [11], Southern Yi, Hani and Bo [12], to name a few.

A traditional description of Chinese is based on Chao Tone Letters [13], which represent tones using numbers 1 to 5 in ascending pitch order. In the case of Zhuang, each tone category is split into two, allegedly to an upper register and a lower register as shown in Table 1. Note the traditional use of "register" in Chinese is distinct from "register languages" involving contrasts in F0 and/or phonation.

Table 1. *Tone categories and corresponding Chao Tone Letters in Wuming Zhuang (based on* [1]).

Tone	1	3	5	7 Short	7 Long
Chao	24	55	35	55	35
Example	[na]	[na]	[na]	[nap]	[na:p]
Gloss	'thick'	'face'	'arrow'	'to put into'	'to be stuck'

Tone	2	4	6	8 Short	8 Long
Chao	31	42	33	33	33
Example	[na]	[na]	[na]	[nap]	[na:p]
Gloss	'field'	'aunt'	'meat'	'to bind'	'to turn in tax'

Note that checked syllables, which from the point of view of tone behave as a distinct sub-system contrasting two tones (tones 7 and 8), are not included in the scope of this study.

The main focus of discussion in the rest of the paper is to investigate the roles of pitch and phonation and how these are used to distinguish tonal categories in unchecked syllables for our Du'an Zhuang speaker. Note that results from this single speaker cannot be generalized to the language.

2. Materials

The data for this study was acquired from one male native speaker of Du'an Zhuang in his mid 20's. He received his education in Zhuang until middle school. Since then, his education was in Mandarin Chinese. At the time of the elicitation, the consultant communicated with his immediate family members and relatives in Du'an Zhuang over the phone. In the US, however, he mainly spoke Mandarin Chinese.

The words used in the recordings were taken from [1], a short grammar that describes Wuming Zhuang. The word list used for recordings was taken from the sections on sounds and tone. The stimuli list includes all onset types, vowel qualities, vowel lengths and coda types in order to control for contextual effects on F0. 160 items were used in total, with each item repeated multiple times.

Six tokens with duration less than 40 ms were discarded since they were too short.

http://dx.doi.org/10.21437/TAL.2016-12

Influence of onset and coda segments on tone, duration and F0 contour is important. Unfortunately, with the available data it is impossible to do such an analysis since not all combinations of tones and onsets (or codas) were present. Therefore, only differences in tones are reported here.

Four two-hour recording sessions were conducted in a soundproof booth at a U.S.-based university. The participant took breaks every 30–45 minutes. He wore a head-mounted unidirectional cardioid microphone Shure WH-30, which was connected to a solid-state recorder Marantz PMD671 with an XLR cable. Most collected samples were stored in mono .WAV files at a sampling rate of 48 kHz; some of them (111 of 549) were sampled at 44.1 kHz due to a setup error. Words were read in list form to exclude possible tone sandhi effects, and were presented using Chinese characters. Their transcriptions, which are for Wuming Zhuang, were used as a basis for comparing each tone with the corresponding elicited Du'an Zhuang versions. Most tokens were repeated at least three times (there were ten words that were only recorded once). A total of 559 tokens were recorded. Of these, 436 tokens were unchecked syllables. 162 of these had nasal codas; the remaining 274 tokens had no coda.

The annotation process was eased using Praat [14]. The annotation included a segment tier, on which the syllable rhyme was marked. The decision to focus on syllable rhymes follows since tone is associated to rhymes. Rhyme boundaries were determined manually. The initial boundary was set by looking for the point at which the nucleus vowel formants became apparent. The end of the rhyme was determined by the absence of voicing. Boundaries were then adjusted to the nearest zero-crossing sample.

Each token was inspected by the first author and categorized as creaky or non-creaky based on aural and visual inspection of the waveform. The latter process involved looking for irregularly spaced voicing pulses. This classification was used to determine parameters in F0 analysis, as later explained.

3. Analysis

3.1. F0 contour analysis

F0 values were estimated using an autocorrelation method [15] implemented in [14] every 10 ms. This algorithm has several parameters from which ‘Pitch Floor’ (minimum F0 threshold) and ‘Voicing Threshold’ are important for this analysis. For non-creaky tones, they were set at their defaults (75 Hz and 0.45, respectively), but for creaky tones (i.e., those tones that were manually identified as creaky), these parameters were lowered to 40 Hz and 0.2 as a way to minimize errors on the voiced/unvoiced classification performed by the algorithm. A slightly higher duration threshold for exclusion of tokens (50 ms, instead of 40 ms) was used for creaky tokens to accommodate the lowered frequency of those tokens. F0 contours generated in this fashion were visually and aurally inspected and corrections were manually made when octave jump errors were found.

Each token was time normalized (*normT*) before fitting a smoothing cubic spline ANOVA model as implemented in [16] and used in [17] and [18].

In our model, F0 is explained by the factors *Tone* (after reclassification) and *normT*, and their interaction. Smoothing parameters were selected by a generalized cross-validation method using the default parameter values (e.g α =1.4).

A coefficient of determination R^2 = .722 was found for the unchecked tone model. To evaluate the fitted model, F0 values were predicted every 2 percentage points of the normalized time and were plotted along with 95% Bayesian confidence intervals as shown in Fig. 1. Overlapping between confidence intervals corresponds to regions where no evidence of significant difference between tones was found. Results in Fig. 1 can be used to describe the tone system of Du'an Zhuang, but it is not sufficient to classify the tones. Details of these results will be discussed in Section 4.1

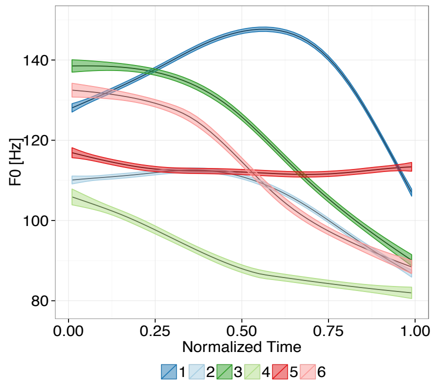

Figure 1: F0 contour for unchecked tones along with 95% confidence intervals. Non-significant differences between contours correspond to overlapping regions.

3.2. Creakiness analysis

This section will discuss a composite method used to measure creakiness in detail. Creakiness was estimated every 10 ms using the method described in [19] and [20]. This algorithm determines the odds of a frame being creaky based on a combination of acoustic features, although the authors note that the output should not be taken as a strict probability. Information theoretic methodology was applied to assess how well a set of acoustic features correlated with actual creakiness judgments of recordings from various databases. Their model outperformed other recent models in its ability to correctly identify instances of creakiness [19].

The acoustic features considered included the difference between the first two harmonics H2–H1, F0 contour, residual peak prominence (RPP), as well as a group of features used by [21], including power peak parameters, inter-pulse similarity (IPS), and intra-frame periodicity. RPP refers to prominence of local minima of the residual signal after applying an LPC filter to the signal; IPS is the maximum of the cross-correlation between two adjacent audio frames in the vicinity of a peak candidate in the signal. Three further measures (normalized signal energy, number of zero-crossings, and variance in the very short-term power contour) were also included to minimize false positives. Annotated recordings from a number of databases were used to train an artificial neural feedforward network consisting of a single hidden layer of neurons as implemented in [22], yielding an output, which is treated as a probability of creakiness. The annotated recordings included examples from four languages that do not use creakiness contrastively: American English, Swedish, Finnish, and Japanese. Drugman et al. [20] determined the creakiness of the samples manually (via aural and visual inspection of the waveform).

There are a few advantages of employing this creakiness algorithm over previous methods. For example, the creakiness algorithm used here uses multiple acoustic measures in a composite way to assess, in a single quantitative manner, whether an audio frame is creaky or not. The weighting of each acoustic measure is based on the training of the neural network [22]. Previous non-composite methods examine acoustic measures such as H2–H1, H2–A1 or RPP, independently [12, 33, 34]. One could observe the disparate acoustic correlates in this way to determine whether a token is creaky or not, but this work is rather painstaking and prone to errors. While it is valid to apply a particular acoustic measure that is already known to correlate with creakiness in a particular language, exploratory studies on languages where the status of creakiness is unknown should consider all acoustic measures.

The method used here has the additional advantage that it does not require any vowel normalization in order to correct for differences in H2–H1. As such, we believe that the composite method in [20] can readily be extended to less studied languages for further investigation of creakiness; due to the nature of the data collection, these languages do not usually result in a balanced, controlled data set. However, not all acoustic features associated with creakiness are perceptual cues, and so the results here do not necessarily imply that creakiness is contrastive.

After normalizing the duration of each Du'an Zhuang token, the differences in creakiness profiles of each tone were investigated in the same fashion as was done with F0 contours. Fig. 2 illustrates the creakiness contours for each unchecked tone with their corresponding 95% Bayesian confidence intervals. For creakiness, the coefficient of determination was $R^2 = .273$ for the unchecked tone model.

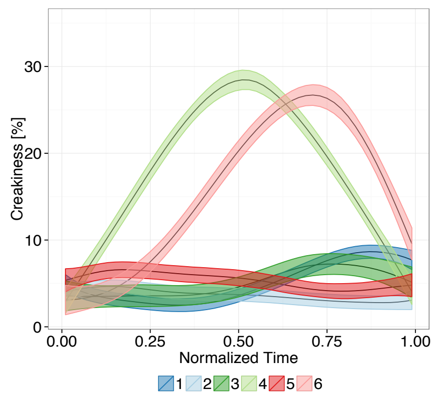

Figure 2: Creakiness probability contour for unchecked tones and their corresponding 95% confidence intervals.

4. Discussion

Four distinct categories of F0 profile in unchecked syllables are exemplified in Fig. 1. These categories are a rising-falling contour (exemplified by tone 1), a lower falling contour (tones 2 and 4), a higher falling contour (tones 3 and 6), and a mid-level contour (tone 5). Notably, tones 2 and 4 as well as tones 3 and 6 have very similar F0 contours. The following discussion will focus on tones 2, 3, 4 and 6 since these tones exemplify the interplay between F0 and phonation.

Tone 4 differs from tone 2 only in that tone 4 falls earlier; while tone 3 and tone 6 have nearly identical falling F0 profiles, tone 3 is on average 5–10 Hz higher throughout. These similarities in F0 profile highlight pairs of tones in which phonation may be contrastive. One member of each pair (namely, tone 4 and tone 6) has significant creakiness, as is seen in Fig. 2. This analysis is consistent with the F0-lowering effect, seen in both creaky tones relative to their non-creaky counterpart. F0-lowering is a common feature of creaky phonation [25, 26]. The fact that the F0 profiles are not identical suggests that the contrasts between tones 2 and 4 and between tones 3 and 6 may involve F0 and creakiness playing a role together.

Another interesting aspect is in the timing of F0 contours. Rose, and Seidle [27] found that creaky phonation correlates with an earlier fall in F0 in Thai-Phake. The same correlation is seen in Du'an Zhuang when comparing tones 2 and 4, for example, as shown in Fig. 3.

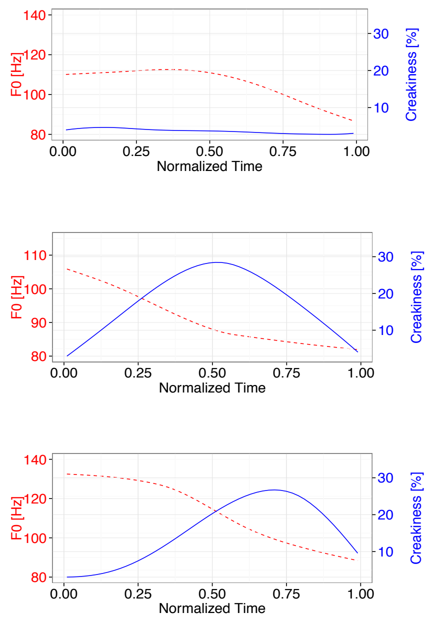

Figure 3: Overlaid temporal contours of F0 (dashed line) and creakiness (solid line) for tone 2 (top panel), tone 4 (middle panel) and tone 6.

Both tones begin and end at similar F0, however while the creaky tone 4 has an initial drop in F0, tone 2 F0 is flat over the first half of the vowel, only dropping over the latter half of the vowel. In addition to F0, tones 4 and 6 exhibit notable differences in the timing of the creakiness peak. While tone 4 has a peak at the midpoint, tone 6 has a creakiness peak at around 70% of the vowel duration. The early creakiness peak in tone 4 corresponds to an earlier fall in F0. Therefore, location of creakiness peak and timing in the falling F0 contours appear to be correlated with each other. The degree to which creakiness, F0, and their relative timing are involved in tone category discrimination is another interesting direction for future research (cf. Andruski [11] who did such a perception study in Green Mong).

While the tonal system described by [1] for Wuming Zhuang has six contrastive tones in unchecked syllables based only on distinct F0 profiles, acoustic analysis for one Du'an Zhuang speaker suggests only four distinct F0 profiles, with an additional phonation contrast overlaid, resulting in 6 contrastive tone categories, as shown in Table 2. While the same number of contrastive tones (six) is found in each dialect, Du'an Zhuang differs from the description of Wuming Zhuang in that F0 alone does not fully capture the six-way tonal contrast in Du'an Zhuang; phonation plays a significant role in this contrast in Du'an Zhuang. The results here suggest that creakiness may also be present but overlooked in Wuming Zhuang, a possibility that can be verified in future work.

Table 2. *Comparison of Chao Tones for Wuming and Proposed Chao Tones for Du'an Zhuang*

Tone	WUMING ZHUANG AS IN [1]	DU'AN ZHUANG		
		Our Proposal	Pitch	Phonation
1	24	453	rising falling	modal
2	31	31 modal	lower falling	modal
3	55	51 modal	falling	modal
4	42	31 creaky	lower falling	creaky
5	35	33	mid level	modal
6	33	51 creaky	falling	creaky

5. Conclusions

The tonal system of Du'an Zhuang involves four distinct F0 contours in unchecked syllables, two of which are further divided by phonation differences. The two falling tones, [51] and [31], have creaky and modal counterparts, resulting in six distinct tonal categories. While the standard dialect, Wuming Zhuang also has six distinct tones, those tones reportedly are distinguished based only on F0 contours, with no mention of phonation playing a contrastive role.

This acoustic study highlights the fact that tonal contrast may not be entirely fulfilled by a single acoustic dimension; in the case of one Du'an Zhuang speaker, it appears that creakiness plays an important role, in addition to F0. While these roles have been established here based on a production study of one speaker, it remains to be shown whether and to what extent speakers of Du'an Zhuang use creakiness and/or F0 as cues in identifying tones in general. Additionally, future research on the perception side would complement the current results from the production side. This new research could also address whether speakers use timing information to establish tonal categories as well, as it was noted that the location of maximum creakiness correlated with timing of falls in F0.

6. Acknowledgements

We have particularly benefited from comments of Bruce Connell, John Pham, Keren Rice, and Shigeto Kawahara. Additional thanks are due to Will Bennett and Sophie Salffner for their comments as well as anonymous reviewers. Previous versions of this work were presented at the CUNY Phonology Conference on Endangered Languages and the 2013 International Conference on Phonetics and Phonology held at NINJAL. Of course, all errors are our own responsibility. This research was also possible thanks to the research reassigned time awarded to Seunghun Lee.

7. References

[1] Wei, Q. and Qin, G., "[Zhuang yu jianzhi] (A brief description of the Zhuang language)", Minzu chubanshe, Beijing. In Chinese, 1980.
[2] Hansen, B. and Castro, A., "Hongshui he zhuang dialect intelligibility survey", SIL Electronic Survey Reports 2010-025, 2010.
[3] Kullavanijaya, P. "A study of lexical variation in seven Zhuang dialects", in M. R. K. Tingsabadh and A. S. Abramson [Eds], Essays in Tai Linguistics, 229-258, Chulalongkorn University Press, Bangkok, 2001.
[4] Abramson, A. S. and Luangthongkum, T., "A fuzzy boundary between tone languages and voice-register languages", in Fant, G., Fujisaki, H., and Shen, J., eds., Frontiers in Phonetics and Speech Science, 149–155, The Commercial Press, 2009.
[5] Duanmu, S., "A formal study of syllable, tone, stress and domain in Chinese languages", PhD thesis, MIT, 1990.
[6] Bao, Z., "The structure of tone", Oxford University Press, New York, 1999.
[7] Zhu, X., "Multiregisters and four levels: A new tonal model", J. of Chinese Linguistics 40(1):1, 2012.
[8] Henderson, E. J., "The main features of Cambodian pronunciation", Bulletin of the School of Oriental and African Studies 14(1):149–174, 1952.
[9] DiCanio, C. T., "The phonetics of register in Takhian Thong Chong", J. of the Int. Phonetic Association 39(2):162–188, 2009.
[10] Gruber, J. F., "An articulatory, acoustic, and auditory study of Burmese tone", PhD thesis, Georgetown University, 2011.
[11] Andruski, J. E., "Tone clarity in mixed pitch/phonation-type tones", J. of Phonetics 34:388–404, 2006.
[12] Kuang, J., "Phonation in tonal contrasts", PhD thesis, UCLA, 2013.
[13] Chao, Y., "A system of tone letters", Le maître phonétique, 45, 1980.
[14] Boersma, P. and Weenink, D. "Praat: doing phonetics by computer", Available [Sep. 2016] from www.praat.org, 2016.
[15] Boersma, P. "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound", in Proc. of the Institute of Phonetic Sciences 17:97–110, Amsterdam, 1993.
[16] Gu, C., "Smoothing spline ANOVA models: R Package GSS", J. of Statistical Software, 58(5):1–25, 2014.
[17] Lee-Kim, S. I., Kawahara, S. and Lee, S. J., "The 'whistled' fricative in Xitsonga: its articulation and acoustics", Phonetica, 71(1):50–81, 2014.
[18] Moisik, S., Lin, H. and Esling, J., "Larynx height and constriction in Mandarin tones", in Eastward Flows the Great River: Festschrift in Honor of Professor William S. Y. Wang on his 80th Birthday, 187–205, City University of HK Press, 2013.
[19] Kane, J., Drugman, T., and Gobl, C., "Improved automatic detection of creak", Computer Speech & Language 27(4):1028–1047, 2013.
[20] Drugman, T., Kane, J. and Gobl, C., "Data-driven detection and analysis of the patterns of creaky voice", Computer Speech & Language 28(5):1233–1253, 2014.
[21] Ishi, C. T., Sakakibara, K.-I., Ishiguro, H. and Hagita, N., "A method for automatic detection of vocal fry", IEEE Trans. on Audio, Speech, and Language Processing 16(1):47–56, 2008.
[22] Degottex, G., Kane, J., Drugman, T., Raitio, T., and Scherer, S., "COVAREP – A collaborative voice analysis repository for speech technologies", in IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP 960–964, 2014.
[23] Gordon, M. and Ladefoged, P., "Phonation types: a cross-linguistic overview", J. of Phonetics 29:383–406, 2001.
[24] Garellek, M. and Keating, P., "The acoustic consequences of phonation and tone interactions in Jalapa Mazatec", J. of the Int. Phonetic Association 41(2):185–205, 2011.
[25] Esling, J. H., Harris, J. G., Hardcastle, W. and Beck, J., "States of the glottis: An articulatory phonetic model based on laryngoscopic observations", A figure of speech: A Festschrift for John Laver 347–383, 2005.
[26] Edmondson, J. A. and Esling, J. H., "The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies", Phonology 23(2):157–191, 2006.
[27] Rose, P. and Seidle, R., "Thai-Phake tones: Acoustic and perceptual data on a Tai dialect with contrastive creak", in Proc. 3rd Australian Int. Conf. on Speech Science and Technology 394–400, Canberra, 1990.