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

Boundary effects on allophonic creaky voice: A case study of Mandarin lexical tones

Hong Zhang[1]

[1]University of Pennsylvania
zhangho@sas.upenn.edu

Abstract

The present study investigates the effects of prosodic boundary on the acoustic realization of allophonic Mandarin lexical tone creak. Speech of twenty speakers from Mandarin Call Home corpus was analyzed. Results show that stronger creaky voice occurs more often at the end of an utterance than the end of an intermediary phrase, and is least likely to occur phrase medially. The observed creak is stronger at boundaries with higher BI levels, regardless of the intrinsic pitch target of lexical tones. Tonal category also appears to be worse at predicting the phonetics of creak. Therefore, compared to lexical tone, prosodic boundary has stronger effects on the acoustic realization of creaky voice. The interaction between prosodic boundary and lexical tone on the acoustics of creaky voice also has implication on the interaction between lexical tone and intonation.

Index Terms: creaky voice, prosody, lexical tone, interaction

1. Introduction

Creaky voice is usually characterized by increased glottal constriction and/or irregular vibration. Except for the function of expressing phonological contrast, studies have shown that creaky voice can be produced allophonically in various phonetic conditions. In American English, creaky voice can be realized as coda glottalization [1],[2],[3] and phrase final creak [3],[4]. In a tonal language, creaky voice can be produced in the lexical tones with low pitch targets [5],[6]. Mandarin Chinese is typical in its predictable tone-creak interaction [7], where tones can become creak whenever pitch target is low. The term "creaky voice" in the traditional three-way distinction of voice quality [8], however, is unable to reflect the variation among its exact acoustic realizations. Studies, for example, [9] and [10], have demonstrated that creaky voice in fact covers a relatively wide range of voice qualities with distinct acoustic features.

The interaction between prosodic boundary and allophonic creak in American English has received much attention [1],[11]. [1] established that stress has a much greater effect on word-inital glottalization than the creak at coda positions in American English. [3] further suggests that all else being equal, coda glottalization is more likely to happen at an intonational phrase boundary. It has also been claimed that boundaries with higher BI level tend to be more likely to have phrase final creak [11]. The coded "creaky voice" at higher BI boundary is also more constricted and can spread to multiple syllables, which is more likely to be driven by low F0 at final positions [4]. However, the rate of production and acoustic characteristics of glottalization are subject to individual preference [12] [11].

Lexical tone creak appears to have similar pragmatic functions as the creak found in English. It has both perceptual and emphatic functions. Tone identification is tested to be facilitated by the presence of creaky voice in low level tones in Cantonese [6] and low tones in Mandarin [7]. At discourse level, creaky voice in a tonal language could function as a marker of discourse structure [13], similar to that of phrase final creak in English.

The interaction between prosody and the creak caused by lexical tones, however, are different in their production mechanism. The target of complete glottal closure is always absent. Given the influence of prosody on the phonetics of t-glottalization [1],[11], it can be expected that the phonetics of lexical tone creak indeed varies depending on the prosodic structure. Recent work has provided evidence that in Cantonese, realizations of lexical tones are deviated from canonical forms to fit the requirement of intonation [14], and such modification poses challenges for tone identification [15].

Given the similarity in pragmatic functions in prosody and the different sources of the two kinds of creak, the present study attempts to better our understanding of lexical tone creak by examining its acoustic realizations at its intersection with prosody. The experiments use the data drawn from a corpus of spontaneous speech. In the first experiment, we first present a careful acoustic classification of creaky voice observed in the data. The acoustics of creak, defined by creak categories, at different prosodic positions and lexical tones are then discussed. The second experiment further suggests that prosody can serve as a better predictor of voice quality in general. Therefore we conclude that prosody has stronger effects on the acoustics of creak than lexical tones.

2. Experiment I

2.1. The data

Data are drawn from the Call Home telephone conversation corpus from the University of Pennsylvania. This corpus contains telephone conversations collected from 240 Mandarin speakers. Each speaker was engaged in a fifteen-minute phone call with their friends or family members discussing a topic of their choice while the conversation being recorded. To remove gender effects on creak production, only female speakers were selected for analysis. Recordings of ten speakers from five conversations are chosen for the analysis. The total number of analyzed syllables is 1082.

2.2. Labeling

Creaky voice is first visually identified. The corresponding syllable type, tone type and prosodic position are then labeled by a trained phonetician who is also a native speaker of Mandarin. Prosodic positions are labeled according to the C-ToBI system [16]. Two prosodic boundaries are identified: Interme-

http://dx.doi.org/10.21437/TAL.2016-20

diary phrase boundary and utterance boundary (corresponding to Break Index 3 and 4 in C-ToBI system). Phrase initial and phrase medial positions are also labeled. Figure 1 is an example of the labeled segment.

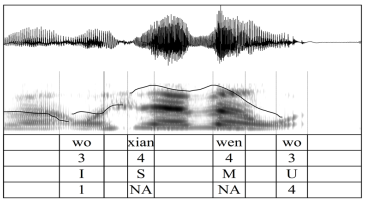


Figure 1: *An example of labeled segment. The four tiers from top to bottom are: syllable, tone, prosodic position and creak type. I: intermediary phrase final; U: utterance final; M: phrase medial; S: phrase initial.*

Classification of different types of creaky voice was performed manually. Cues for the presence of creak are summarized in Figure 2 [9]. Four types of creak [9] were classified. More periodic voicing is assigned a smaller identification number. Manual classification is preferred over automated procedure based on acoustic analysis due to limited explanatory power of current acoustic measurements [10] and highly reliable diagnostic of creak using visual inspection [11].

2.3. Acoustic validation

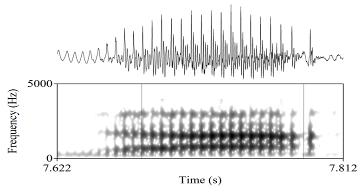
Visual classification of creaky voice is then validated by six acoustic parameters. Previous studies have identified several acoustic parameters (AP) that correlate with different aspects of voice quality. Following [17]'s proposal, F0, H1*-H2*[1], H2*-H4*, H4*-2kHz*, CPP are thought to be sufficient in identifying creak variation. However, Subharmonic-to-harmonic ratio (SHR) should also be included as an indicator of multiply pulsed voice [9]. Extreme high F0's are corrected in *post hoc* examination. Measurements of spectral tilt was not applied to quantify aperiodic voicing due to missing F0. Measurements of the AP's were carried out in VoiceSauce [18], in which F0 was estimated using the Straight algorithm [19]. Raw measurements of each AP were then normalized to a 0 to 1 scale by using the range of that AP observed across speakers from the data.

The AP's of each type of creaky voice with periodic voicing (type I to III) are summarized in Figure 3. The results are mostly significant across the three types of creak. The weak significance of H4*-H2KHz* and H2*-H4* is due to the sampling frequency of the signal. SHR is also marginally significant, suggesting multiply pulsed voicing may not be a prominent characteristic in the data.

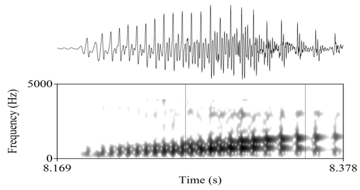
2.4. The distribution of prosodic creak and tone creak

The distribution of four types of creaky voice at intermediary final, utterance final, phrase initial and phrase medial positions are summarized in Figure 4. Although type IV creak tended to occur more often at all prosodic positions, there is still a bias toward less periodicity (stronger creak) at boundary positions. Final positions are in favor of stronger creak, and higher BI levels

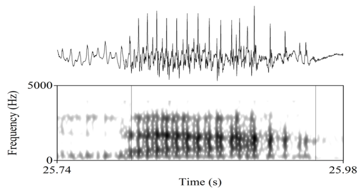
[1]The * indicates correction for formant effects.



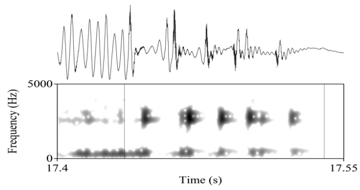
(a) Type I creak, which has relatively high regular F0, but with dampened pulses.



(b) Type II creak. Dampened pulses, but still regular F0, is the major cue for this kind of creaky voice.



(c) Type III creak. F0 becomes extremely low or untraceable. Apparent alternations of pulses can be observed.



(d) Type IV creak. Complete aperiodicity is the primary cue for this kind of voice.

Figure 2: *Waveforms and spectrograms of each type of creaky voice.*

also tend to be less periodic (with more type III and IV creak) among final positions. Chi-squared test of variance suggested that the variance of creak distribution across prosodic positions was significant ($\chi^2 = 92.8198$, $df = 9$, $p<0.001$).

Figure 5 summarizes the distribution of types of creaky voice across lexical tones. The plot shows a similar tendency of producing Type IV creak regardless of lexical tone category (when tones become creak). However, unlike in the previous plot, a bias toward certain types of creak in certain types of tone is absent. Chi-squared test confirms this observation that the variance of creak realization across tones was not significant ($\chi^2 = 19.9866$, $df = 12$, $p = 0.0673$). Therefore lexical tone does not appear to be a good predictor of the acoustics of creak.

The distribution of creak type across prosodic positions and lexical tones suggests a possible correlation between the phonetics of creak and prosodic position, but not lexical tone. This speculation is further supported by the periodicity, measured

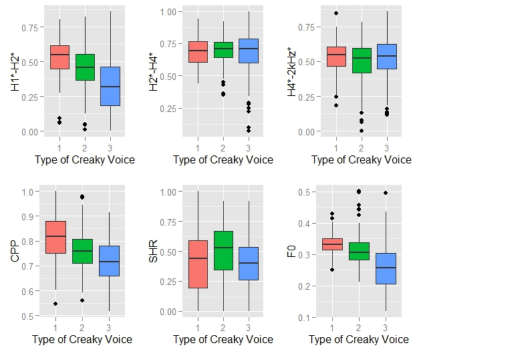


Figure 3: *The boxplots of the AP's of four classified types of creaky voice. H1*-H2*: F(1,438)=151.093,* $p<0.001$*; H2*-H4*: F(1,438)=6.136,* $p<0.05$*; H4*-H2KHz*: F(1,438)=6.337,* $p<0.05$*; F0: F(1,438)=27.414,* $p<0.05$*; CPP: F(1,438)=134.828,* $p<0.001$*, SHR: F(1,438)=6.5,* $p<0.05$*.*

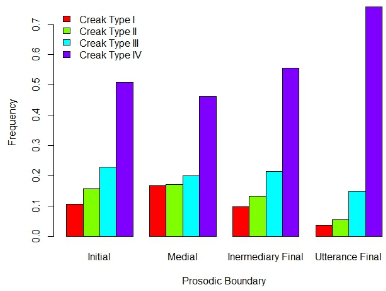


Figure 4: *The barplot of the distribution of creaky voice at four prosodic positions.*

with CPP, among all the creaky segments.

2.5. The phonetics of creak is a function of prosody

Among the AP's measured during validation, CPP (Cepstral Peak Prominence) appears to be a better indicator of creak due to its relative independence from harmonic magnitude in frequency domain. This property entails it the ability to quantify signal's periodicity even with missing F0. Therefore given the source of aperiodicity[2], we can reduce the multidimensional quantification of creak to a single dimension (i.e., periodicity).

Figure 6 plots the average CPP measured at each prosodic position grouped by lexical tones. Since lower CPP values indicate less periodicity, the plot shows a clear trend toward stronger creak at final positions regardless of lexical tones. One-way ANOVA analysis of variance confirms this observation: CPP variation by lexical tones is not significant at utterance final($p > 0.1$) and phrase initial positions ($p = 0.078$), only marginally significant at intermediary phrase final ($p = 0.048$) position, but more significant at phrase medial ($p = 0.014$) and positions. Therefore lexical tone creak is better differentiated by tone categories at the non-boundary position.

A mixed-effect analysis is performed to examine the effects

[2]Since other sources of aperiodicity is also possible, CPP is less telling when the source is not specified

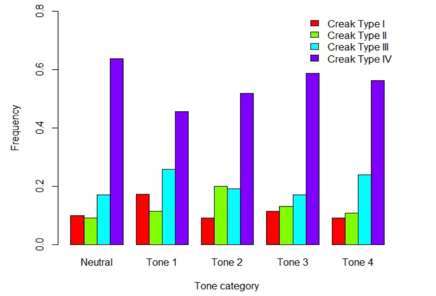


Figure 5: *The distribution of creaky voice across lexical tones.*

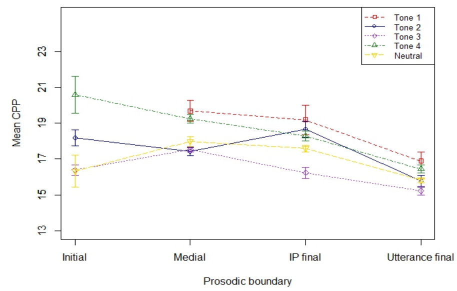


Figure 6: *Periodicity of creak, measured by CPP, of five tones at different prosodic positions. Lower CPP means less periodicity.*

of lexical tones and prosodic positions on the acoustic realization of creaky voice. The model includes prosodic position and lexical tone as fixed effects, and speaker as the random effect factor. CPP is the model's response. Regression result indicates that lexical tone is not a significant predictor of the periodicity of creak (β=0.004, $t = 1.185$, $\chi^2 = 1.196$, $p = 0.1613$), but prosodic position is (β=-0.17, $t = -3.12$, $\chi^2 = 9.827$, $p<0.01$). This result further supports that prosody has stronger effects on the phonetics of the creak which is also presumably a function of lexical tones.

3. Experiment II

The effect of prosodic boundary on the phonetics of creaky voice could be a special case of prosody's effect on voice quality. The second experiment therefore tries to shed some light on the origin of the effect discussed above.

3.1. Data and methods

Recordings of 10 other female speakers from five conversations are drawn from the same corpus. The new data are passed through Penn forced aligner and get segmented at syllable level. The corresponding lexical tone and prosodic position of each syllable are annotated, following similar convention in experiment I. However, in addition to phrase medial positions, prominent positions are also annotated. Therefore the strengthening effect of prosody can be distinguished. Prominence is determined by pure phonetic criterion: raised fundamental frequency and increased intensity in time domain are the two major cues for phonetic prominence. The total number of syllables being analyzed is 5046.

CPP is chosen as the indicator of the aspect of periodicity of voice quality. Unlike in experiment I, measured CPP does not necessarily entail the degree of creak. Rather, it is interpreted as the periodicity of the signal without a specific source. Measurements are also carried out using VoiceSauce. Mean values of each segment are recorded and further analyzed.

3.2. Results

Figure 7 plots the CPP values at each prosodic position grouped by lexical tones. Measured CPP values appear to have a clear pattern over different prosodic positions across all tones but the neutral tone at phrase initial position. The neutral tone at this position is likely to be the result of disfluency or speech errors in running speech, which is not of the interest in this study. One-way ANOVA analyses of variance show that at utterance final and phrase initial positions, CPP is not significantly different across lexical tones ($p > 0.1$). At intermediary phrase final position, CPP is only marginally significant across lexical tones ($p = 0.039$). The difference is only significant at phrase medial position ($p < 0.01$), and at prominent positions ($p < 0.01$).

This result further indicates that voice quality is potentially more dependent upon prosody than lexical tones in Mandarin at least at prosodic boundaries. However, at non-boundary positions, lexical tones do appear to have their own characteristic voice qualities. The parallel between the variation in voice quality and the phonetics of creaky voice provides an evidence that the strong effect of prosodic boundary on the phonetics of creaky voice is part of prosody's effect on voice quality.

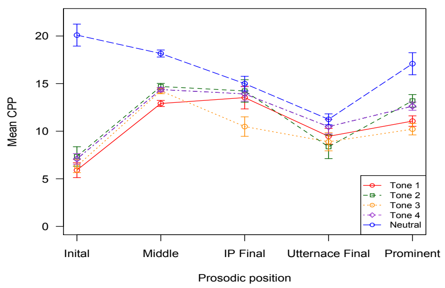


Figure 7: *CPP values of five tones at different prosodic positions*

4. General discussion

This study offers a preliminary view of the interaction between prosodic creaky voice and lexical tone creak. This interaction is different from the one between prosody and coda glottalization, as seen in American English, in that the target of glottal closure is missing. A corpus of conversational Mandarin Chinese has been used. The first experiment investigates the acoustic realization of creak across prosodic positions and lexical tone categories. Among segments produced with creaky voice, a strong effect of prosodic boundary has been found. However, lexical tone category fails to be a significant predictor of the phonetics of creak. The second experiment further indicates that the interaction between prosodic boundary and creak may be case of the interaction between prosody and voice quality.

Prosody's effects on voice quality weigh heavily at boundary positions. In both experiments, lexical tones lose their traits of voice quality at boundary positions. At final positions, in particular, higher BI level entails both stronger creak and less distinction across tones. However, at phrase medial or prominent positions, the realized creaky voice and voice quality remain distinct across tone categories.

The results of this study have two implications. The effects found in the interaction between prosody and coda or initial glottalization [1] can be extended to the creaky voice presumably resulted from low tones. Although creaky voice maintains different sources in the two scenarios, it appears to be under similar influence of prosodic boundary. The second is that the voice quality of lexical tones may be modified to fit the requirement imposed by prosody. This aspect is similar to the interaction between tone and intonation as discussed in [15].

In this study, however, lexical tones are not categorized based on their realized contour. The boundary effects discussed in this study may as well be an aspect of neutralization of F0 contour at boundary positions.

5. Conclusion

The results of this study suggest that although prosody and lexical tones are both capable of inducing creak, prosodic boundary overrides tone in determining the phonetics of creaky voice, possibly even the overall voice quality. The interaction between prosodic creak and tone creak also has implications on tone-intonation interaction.

6. References

[1] J. Pierrehumbert, "Prosodic effects on glottal allophones. vocal fold physiology: voice quality control, ed. by osamu fujimura and m. hirano, 39-60," 1995.

[2] L. Dilley, S. Shattuck-Hufnagel, and M. Ostendorf, "Glottalization of word-initial vowels as a function of prosodic structure," *Journal of Phonetics*, vol. 24, no. 4, pp. 423–444, 1996.

[3] M. K. Huffman, "Segmental and prosodic effects on coda glottalization," *Journal of Phonetics*, vol. 33, no. 3, pp. 335–362, 2005.

[4] M. Garellek, "Perception of glottalization and phrase-final creak," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 822–831, 2015.

[5] C. M. Esposito, "An acoustic and electroglottographic study of white hmong tone and phonation," *Journal of Phonetics*, vol. 40, no. 3, pp. 466–476, 2012.

[6] K. M. Yu and H. W. Lam, "The role of creaky voice in cantonese tonal perception," *The Journal of the Acoustical Society of America*, vol. 136, no. 3, 2014.

[7] M. Y. Kristine, "Laryngealization and features for chinese tonal recognition." in *INTERSPEECH*, 2010, pp. 1529–1532.

[8] P. Ladefoged, *Preliminaries to linguistic phonetics*. University of Chicago Press, 1980.

[9] P. Keating, M. Garellek, and J. Kreiman, "Acoustic properties of different kinds of creaky voice," 2015.

[10] P. Keating, C. Esposito, M. Garellek, S. Khan, and J. Kuang, "Phonation contrasts across languages," *UCLA Working Papers in Phonetics*, vol. 108, pp. 188–202, 2010.

[11] L. Redi and S. Shattuck-Hufnagel, "Variation in the realization of glottalization in normal speakers," *Journal of Phonetics*, vol. 29, no. 4, pp. 407–429, 2001.

[12] D. Huber, "Aspects of the communicative function of voice in text intonation," *Department of Linguistics and Phonetics, University of Lund, Lund*, 1988.

[13] A. Belotel-Grenié and M. Grenié, "The creaky voice phonation and the organisation of chinese discourse," in *International symposium on tonal aspects of languages: With emphasis on tone languages*, 2004.

[14] A. Fox, K.-K. Luke, O. Nancarrow, , , and , "Aspects of intonation in cantonese/," *Journal of Chinese Linguistics*, pp. 321–367, 2008.

[15] J. K. Ma, V. Ciocca, and T. L. Whitehill, "Effect of intonation on cantonese lexical tonesa)," *The Journal of the Acoustical Society of America*, vol. 120, no. 6, pp. 3978–3987, 2006.

[16] A. Li, "Chinese prosody and prosodic labeling of spontaneous speech," in *Speech Prosody 2002, International Conference*, 2002.

[17] J. Kreiman, S. J. Park, P. A. Keating, and A. Alwan, "The relationship between acoustic and perceived intraspeaker variability in voice quality," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

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

[19] H. Kawahara, A. de Cheveigné, and R. D. Patterson, "An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised tempo in the straight-suite." in *ICSLP*, 1998.