THE SOURCE OF CREAK IN MANDARIN UTTERANCES

Yuan Chai

University of California, San Diego
yuc521@ucsd.edu

ABSTRACT

In Mandarin, the lower F0 a segment has, the creakier it is likely to be [11]. In addition, creak is more common in utterance-final position than in non-final positions [12, 22]. In this study, we determine whether creak in Mandarin is motivated by the utterance-final position or by the relatively low F0 in utterance-final position, and whether the effects of utterance position and F0 differ by sentence type. An acoustic study was conducted to test for the presence and degree of creakiness in utterance-initial, medial, and final positions in statements and questions. The results confirm that, in Mandarin, low F0 motivates creak. Controlling for F0, utterance-final position in statements was creakier than non-final positions. In questions, utterance-final position was more periodic than non-final positions, but not creakier. This study shows that in Mandarin, utterance position, F0, and sentence type all have independent effects on creak.

Keywords: creak, F0, position, sentence type.

1. INTRODUCTION

Prototypical creaky voice is characterized by low F0, glottal constriction, and aperiodic pulses [10]. On the one hand, in tonal languages, creaky voice frequently co-occurs with low tones (Cantonese [20]; Mandarin [5, 11, 23]; Hmong [6]). On the other hand, creaky voice can be independent of low F0; Jalapa Mazatec [8] and Mpi [3] have creaky high tones, realized by constricted glottis. Aperiodic voice does not require low F0 either: the vocal folds vibration is so irregular that there is no F0.

In Mandarin, utterance-final position was found to be creakier than non-final positions [12, 22]. We ask whether the creak in utterance-final position depends on low F0. In Mandarin, F0 declines as the phrase proceeds [17, 21]. Thus, utterance-final position is likely to have a lower F0, which should also result in a creakier quality than non-final positions. Alternatively, creak could be independent of F0, and may be associated with final position specifically to indicate finality. This study seeks to determine whether controlling for F0, utterance-final position is still more constricted and irregular than non-final positions.

Additionally, the sentence type (e.g. a statement vs. a question) may also have an independent effect on the creak patterns across the utterances. Smith [19] found that, in French, the voice quality in phrase-final positions differs between statements and questions. The question-final vowels were more periodic than statement-final vowels, which was true for both questions with a pitch fall and those with a pitch rise. The author suggested that sentence type had an effect on voice quality independent from F0. We propose that the difference in voice quality between French statements and questions can extend to Mandarin. In [19], the author compared the voice quality of statements with questions in utterance-final position. However, the author did not compare the voice quality of utterance-final position with non-final positions within each sentence type while at the same time controlling for F0. This study thus examines whether the effect of utterance position differs between different types of sentence controlling when F0 is controlled for.

In sum, the research question of this study is: is it the case that F0, utterance position, and sentence type all have independent effects on the presence and the degree of creak in Mandarin utterances? We hypothesize that utterance position will have a significant effect on the presence/degree of creak independent from F0. Moreover, we test whether that effect differs between statements and questions.

2. EXPERIMENT

2.1. Participants and Stimuli

We recruited 32 speakers who are native speakers of Mandarin (mean age = 20.7, 25 women, 7 men). The speakers did not report any listening or speech disorder. They were asked to read a list of sentences with target words /da/ in Mandarin falling Tone 4 at the initial, medial, and final positions of sentences. Tone 4 was chosen because, according to Kuang [11], among the four tones of Mandarin, Tone 1 and Tone 2 seldom have creak. Tone 3 is almost always creaky, while Tone 4 is creaky half of the time. So to increase the variability of creak in the production, Tone 4 was chosen.

The stimuli consisted of 16 target sentences varying in length (short vs. long) and sentence type (statement vs. question). The short sentences contained 10-11 syllables while the long sentences contained 17-20 syllables. We varied the length of sentences in order to increase the variability of the independent variable F0, which should lead to more statistical power. Sentence length should affect F0 because the slope of F0 declination is steeper in shorter than in longer sentences [15, 21]. It is likely then that shorter sentences will have a lower final F0 than longer sentences.

In order to test the effect of sentence type controlling for all other factors, questions were nearly identical to the statements. A second-person subject and/or an adverb were added to some of the questions in order to make the questions more natural. Questions also had an additional question particle "ma" at the end of the sentences. Example 1 illustrates two sample sentences:

Example 1.1 Statement-short
大家觉得大礼堂不够大。
da4 jia1 jue3 de **da4** li3 tang2 bu4 gou4 **da4**.
People think ballroom not enough big
"People think that the ballroom is not big enough."

Example 1.2 Question-short
大家觉得大礼堂不够大吗？
da4 jia1 jue3 de **da4** li3 tang2 bu4 gou4 **da4** ma?
People think ballroom not enough big Q
"Do people think the ballroom is not big enough?"

Each speaker produced 72 sentences in total (16 target sentences+8 fillers*3 repetitions). The stimuli were grouped into three blocks. The order of the sentences was randomized within each block for each participant. The stimuli were displayed on a computer screen using Psychopy [16]. The recording took place in a sound booth. The productions were recorded at a 44.1 kHz sampling rate and a 32-bit quantization rate using a Yeti microphone and Audacity [2].

2.2. Segmentation Criteria

All target words consisted of monosyllables beginning with an unaspirated stop. The vowel of each target word was segmented from its onset to offset. The vowel onset was marked from the release burst of the onset stop. When the target word was followed by a word with an obstruent onset, the vowel offset was segmented right before the stop closure or the onset of frication of the following syllable. When the target word was followed by a word with a sonorant onset, the vowel offset was segmented when the amplitude of the formants decreased. When the target word was followed by a word without an onset, the target vowel offset was labelled at end of formant transitions for the next vowel. When the target word occurred at the end of the utterance, the offset of the vowel was marked where the voice bar or the formant ended, whichever came first.

2.3. Measuring Creak

The probability of creak was measured using a creaky voice detector [4]. The creaky voice detector uses various acoustic properties including H1*–H2*, residual peak prominence, and pulse irregularities to determine whether creak is present. It makes a binary decision as to whether the sound is creaky every 10 ms. Kuang [12] showed that the detector is accurate for Mandarin roughly 80% of the time. Following [12], the probability of creak equalled the number of "creaky" decisions divided by the number of 10-ms intervals in the sound. For instance, if a given segment was 100 ms long and was marked as creaky in 10 positions, its probability of creak was 0.1.

The output from the creaky voice detector marks the presence of creak, but does not indicate whether the creak is characterized by low F0, glottal constriction, aperiodicity, or a combination of those features. Thus, F0, H1*–H2*, and HNR (harmonics-to-noise-ratio) of the vowel of each target word were calculated automatically using VoiceSauce [18], which outputs a value every millisecond over a target interval. H1*–H2*, the difference in amplitude between the first and second harmonics (corrected for formant frequencies and bandwidths), is related to the degree of vocal fold constriction: a lower value of H1*–H2* is associated with a higher degree of constriction [7]. HNR measures the degree of noise: a lower HNR value represents a higher degree of noise. Both low H1*–H2* and low HNR are associated with creaky voice [7].

3. RESULTS

The results of each acoustic measure and the probability of the creak were analysed using separate linear mixed-effects models, implemented with the lmer() function in the lmerTest package in R. All the slopes and intercepts of the models were random at the subject and the sentence level. The values of F0, H1*–H2*, and HNR were transformed to z-scores to eliminate differences in means among speakers and avoid individuals with larger variance outweighing those with smaller variance [1].

3.1. F0 Trends Within Utterances

To describe the F0 trend over the course of utterance, we compared the F0 in different utterance positions separately for statements and questions. Tokens with an F0 larger than three standard deviations from the speaker's mean were excluded from the analyses. F0 was regressed on the utterance position with utterance position as the fixed effect. For statements, F0 generally declined linearly from the utterance-initial to the final position ($F[1, 1136] = 108.764, p <. 0001$). In questions, utterance-initial position generally had a higher F0 than the non-initial positions ($F[1, 1143] = 22.316$, $p = .003$), but F0 did not differ between utterance-medial and -final positions ($F[1, 1143] = .039$, $p = .845$). Those results conform to the F0 pattern found in previous studies: in Mandarin, F0 falls in the final position in statements, but rises in the final position of questions [9]. The rising contour in the final position of questions counters declination, so that the F0 remained steady from the medial to final position. Figure 1 shows the F0 track of a statement and a question produced by a male speaker. Figure 2 shows the average F0 by position in utterance for both sentence types.

Figure 1: F0 contours of a sample statement (top) and question (bottom).

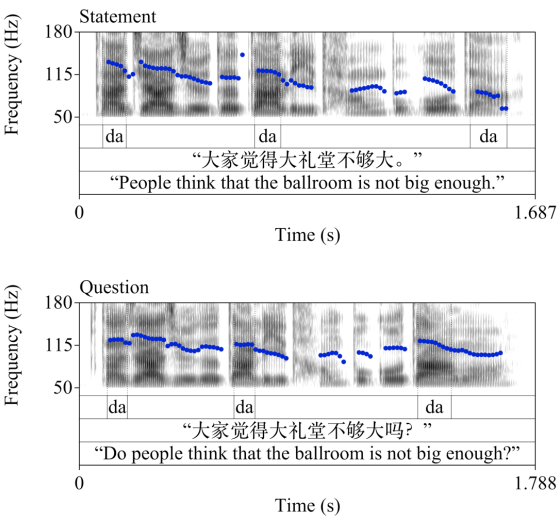
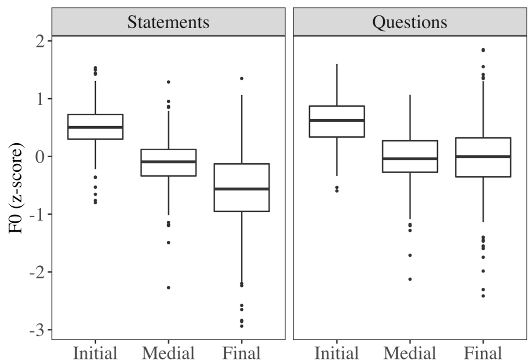


Figure 2: F0 in different utterance positions

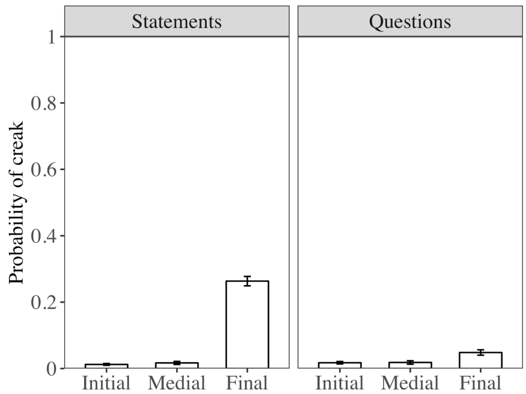


3.2. Probability of Creak

The probability of creak was regressed on F0 and utterance position for statements and questions separately, with F0 and utterance position as fixed effects. The probability of creak was adjusted to its arcsine value in order to satisfy the normality distribution assumption of linear regression. Within each model, utterance-final position was compared with utterance-initial and -medial positions. The alpha level was adjusted to .025 using a Bonferroni adjustment.

In statements, F0 was not significantly related to the probability of creak ($F[1,1133] = 2.996, p = .094$). Controlling for F0, utterance-final position had a significantly higher probability of creak than initial position ($F[1,1133] = 73.668$, $p < .0001$) and medial position ($F[1,1133] = 139.098$, $p < .0001$). In questions, F0 was significantly negatively-related to the probability of creak ($F[1,1140] = 16.484$, $p = .002$). The higher the F0 was, the less likely it was for the segment to have creak. Controlling for F0, the probability of creak in utterance-final position did not differ significantly from the initial position ($F[1,1140] = .063$, $p = .808$) or from the medial position ($F[1,1140] = 4.482$, $p = .086$). The probability of creak in different utterance positions is shown in Figure 3.

Figure 3: Probability of creak in different positions.



3.3. Acoustic Parameters

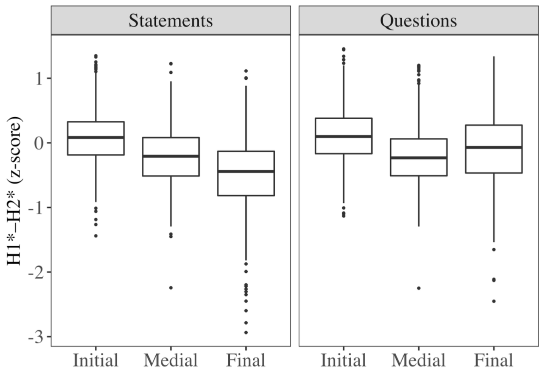
H1*–H2* and HNR were regressed on F0 and the utterance position for statements and questions separately. F0 and utterance position were included as fixed effects. Tokens with H1*–H2* or HNR larger than three standard deviations from the speaker's mean were excluded from the analysis of H1*–H2* or HNR. Since H1*–H2* was calculated based on F0, F1, and F2, tokens with F0, F1, or F2 larger than three standard deviations from the speaker's mean were also excluded from the analysis of H1*–H2*. Within each model, utterance-final position was compared to utterance-initial and -medial positions (alpha = .025).

3.3.1. H1–H2**

In statements, a lower H1*–H2* was associated with lower F0 values ($F[1,1133] = 47.142$, $p < .0001$). The

H1*–H2* of utterance-final positions was not distinctively different from initial ($F[1,1133] = 2.039$, $p = .179$) or medial positions ($F[1,1133] = 2.381$, $p = .132$). In questions, a lower H1*–H2* was associated with lower F0 values as well ($F[1,1139] = 36.409$, $p < .0001$). H1*–H2* of utterance-final position was not significantly different from initial position either ($F[1,1139] = .854$, $p = .362$). Utterance-final position had a marginally higher H1*–H2* than the medial position ($F[1,1139] = 4.657$, $p = .056$). Figure 4 shows the average H1*–H2* by position in utterance for both sentence types.

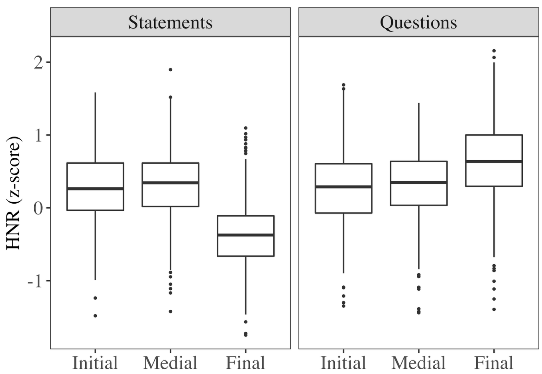
Figure 4: H1*–H2* in different utterance positions.



3.3.2. *HNR*

In statements, a lower HNR was associated with lower F0 values ($F[1,1133] = 17.733$, $p = .007$). Utterance-final position had lower HNR than utterance-initial position ($F[1,1133] = 34.492$, $p < .0001$) and medial position ($F[1,1133] = 69.723$, $p < .0001$). In questions, a lower HNR was also associated with lower F0 values ($F[1,1140] = 19.483$, $p = .0001$). Conversely from statements, in questions, utterance-final position had higher HNR than both utterance-initial position ($F[1,1140] = 37.161$, $p < .0001$) and utterance-medial position ($F[1,1140] = 34.881$, $p < .0001$). Figure 5 shows the average HNR by position in utterance for statements and questions.

Figure 5: HNR in different utterance positions.



4. DISCUSSION

Overall, the results demonstrate that, in statements and questions, a lower F0 is associated with a higher degree of vocal fold constriction (as indexed by lower H1*–H2*) and a noisier quality (as indexed by lower HNR). Together, this suggests that lower F0 is associated with creak, in accordance with [11].

Moreover, the effect of utterance position differs between statements and questions. Controlling for F0, final position of statements is likelier to have creak. However, the probability of creak does not significantly increase at the end of questions. The acoustic results of creak accord with the creak probability. In statements, utterance-final position is noisier than non-final positions, even controlling for F0. Conversely, utterance-final position in questions is less noisy (more periodic) than non-final positions, controlling for F0.

The results also suggest that, in Mandarin statements, creak may signify finality. However, in questions, the final positions are not creakier than non-final positions. Thus, creak may not act as a general marker of utterance finality. Those findings are in accordance with the findings in French [19].

There are several limitations to the current study; it tests a single segment /da/ in Mandarin. In addition, it tests a single type of question—namely, confirmation questions with a “ma” question particle. Yet there are several additional types of questions in Mandarin [13, 14]. Follow-up studies are currently underway to test more segments and look into whether the voice quality pattern of “ma” questions also applies to other types of questions. More specifically, we seek to determine whether other types of questions are also less creaky at final positions than non-final positions.

5. CONCLUSION

Low F0 induces creak in Mandarin for both statements and questions. But when controlling for F0, final positions have a creakier quality than non-final positions in statements, suggesting that the statement-final creak serves a pragmatic motivation. In contrast, controlling for F0, final position in questions is not creakier than non-final positions. F0 is still the major source of creak in Mandarin questions, at least for those that employ the particle “ma.” In summary, F0, utterance position, and sentence type all have independent effects on creak in Mandarin.

6. REFERENCES

[1] Al Shalabi, L., Shaaban, Z., Kasasbeh, B. 2006. Data mining: A preprocessing engine. *Journal of Computer Science*, *2*, 735-739.

[2] Audacity computer software. 2018. https://audacityteam.org/.

[3] Blankenship, B. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics*, *30*, 163-191.

[4] Drugman, T., Kane, J., Gobl, C. 2014. Data-driven detection and analysis of the patterns of creaky voice. *Computer Speech & Language*, *28*, 1233-1253.

[5] Davison, D. 1991. An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics, 78*, 50-57.

[6] Garellek, M. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics*, *40*, 152-161.

[7] Garellek, M. 2019. The phonetics of voice. In: Katz, W., Assmann P. (eds), *The Routledge Handbook of Phonetics.* In press.

[8] Garellek, M., Keating, P. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *JIPA,* 41, 185-205.

[9] Ho, A. 1977. Intonation variation in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative. *Phonetica*. 34, 446-457.

[10] Keating, P., Garellek, M., Kreiman, J. (2015, August). Acoustic properties of different kinds of creaky voice. Proc. 18th *ICPhS* Glasgow.

[11] Kuang, J. 2017a. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *J. Acoust. Soc. Am*. 142, 1693-1706.

[12] Kuang, J. 2017b. Creaky voice as a function of tonal categories and prosodic boundaries. *Proc. Interspeech 2017*, 3216-3220.

[13] Li, C., Thompson, S. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley and Los Angeles: University of California Press.

[14] Liing, W. 2014. *How to ask questions in Mandarin Chinese.* New York: CUNY Academic Works.

[15] Maeda, S. 1976. *A characterization of American English intonation* Cambridge, MA: MIT.

[16] Peirce, J. 2007. PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*, 8-13.

[17] Shih, C. 1997. Declination in mandarin. *Proc. Intonation: Theory, Models and Applications 1997* Athens, 293-296.

[18] Shue, Y.-L., P. Keating, C. Vicenik, K. Yu (2011) VoiceSauce: A program for voice analysis, Proc. 17th *ICPhS* Hong Kong, 1846-1849.

[19] Smith, C. 2002. Prosodic finality and sentence type in French. *Language and Speech*, *45*, 141-178.

[20] Yu, K., Lam, H. 2014. The role of creaky voice in Cantonese tonal perception. *J. Acoust. Soc. Am.*, *136*, 1320-1333.

[21] Yuan, J., Liberman, M. 2014. F0 declination in English and Mandarin broadcast news speech. *Speech Communication*, *65*, 67-74.

[22] Zhang, H. 2016. Boundary effects on allophonic creaky voice: A case study of Mandarin lexical tones. *Proc. 5th Tonal Aspects of Languages* Buffalo, 94-98.

[23] Zheng, X. 2006. Voice quality variation with tone and focus in mandarin. *Proc. 2nd Tonal Aspects of Languages* La Rochelle, 132-136.