THE ROLE OF CREAKY VOICE ATTRIBUTES IN MANDARIN TONAL PERCEPTION

Yaqian Huang

Department of Linguistics, University of California, San Diego, USA, 92093
yah101@ucsd.edu

ABSTRACT

This study tests the linguistic importance of various acoustic properties of creaky voice for Mandarin tone perception. Mandarin native speakers identified tones with resynthesized copies of natural tokens, with four citation tones and with one of four creak manipulations: low spectral tilt, irregular F0, pitch doubling, and extra-low F0. Listeners were most sensitive to extra-low F0, which affected identification of the four tones differently: it improved the identification accuracy of Tone 3 and hindered that of Tone 1 and Tone 4. Irregular F0 also hindered Tone 1 identification. No effect of the creak manipulations was found for Tone 2 accuracy. Thus, creak is used in Mandarin tone identification to the extent that it involves low F0 and irregular F0.

Keywords: creaky voice, tonal perception, F0

1. INTRODUCTION

Acoustic properties of different kinds of creaky voice have been documented by [15], who claimed that low F0, irregular F0, or constricted quality each can be sufficient for generating a percept of creaky voice [10]. Still, it remains to be shown whether the different acoustic attributes of creak are linguistically important. For example, in tone languages like Mandarin, where voice quality covaries with pitch [19], it is unclear how voice source parameters interact with F0 in tone identification.

Production studies of Mandarin utterances show that creak is mostly seen on the lowest dipping Tone 3, but can also occur on tones with a low pitch target (rising Tone 2 and falling Tone 4) [1, 3, 7, 14, 19]. This implies that creaky voice in Mandarin is associated with a low F0 [19]. Results of perceptual studies disagree as to the importance of creak in Mandarin tone identification [2, 6, 9, 23]. For example, [9] found no effect of simulated pitch halving on T3 and T4 identification, and [6] did not show biased T3 responses for tokens with jitter in T2 and T3 perception. However, other studies reported facilitation or improvement of Tone 3 identification in a gating task with naturalistic stimuli with creak [2] and for resynthesized stimuli [23].

This leads to the research question: what specific attributes of creaky voice play a role in tonal identification? Given that creaky voice accompanies the production of Mandarin Tones 2, 3, and 4, we hypothesize that listeners may use all available acoustic cues of creaky voice in tone identification. For Tones 2, 3, and 4 that have creak, creak attributes may have positive effects; for the high level Tone 1, creak attributes may have negative effects. In particular, if creaky voice in Mandarin is mainly due to the occurrence of a low pitch, as documented in the production studies mentioned above, listeners may be more sensitive to the low F0 among all possible creak attributes. In this study, we test these hypotheses in a tone identification task.

2. METHODS

2.1. Stimuli

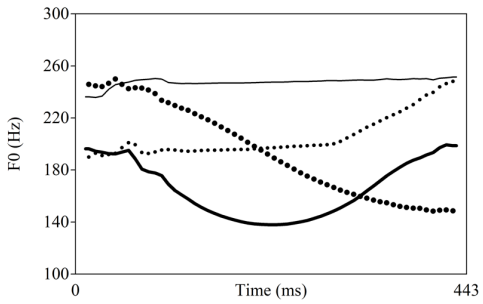
Target monosyllables /jou/, /jau/, /ei/, and /ja/, controlled for orthographic frequency from SUBTLEX-CH [5], were produced by a Beijing female speaker using the four citation tones. Original syllables were resynthesized based on a single Tone 1 token with modal or "creaky" voice using the Klatt synthesizer [17] in Praat [4]. The duration of stimuli differs across target monosyllables, while maintaining a fixed value within each monosyllable. The amplitude of stimuli was scaled to 75 dB after all manipulations described in following subsections.

2.1.1. Modal tokens

After the natural tokens were resynthesized, some preset voice parameters such as *Aspiration*, *Open Phase* (H1–H2), and *Spectral Tilt* (H1–3000Hz) were modified to ensure naturalness of the voice. These were included for all stimuli, but the last two parameters were further manipulated for certain creak attributes, as discussed below.

The pitch contours of the original tones were modeled based on the speaker's production using minimal determinant points. Linear interpolation was added among F0 points, as shown in Fig. 1. The duration of pitch contours was adjusted accordingly in order to fit the duration of each monosyllable.

Figure 1: Resynthesized F0 tracks of token /jou/. (T1: thin solid, 232.7~246.5~250.9 Hz; T2: thin dotted, 190~200~255 Hz; T3: thick solid, 196.5~138~200 Hz; T4: thick dotted, 245~148 Hz)



2.1.2. Creaky tokens

Creaky stimuli were resynthesized from the modal stimuli described above by adding a "creak" portion in different positions which correspond to the loci of a low pitch target; namely, at the beginning of Tone 2, in the middle of Tone 3, and at the end of Tone 4. The splice points for creak were chosen from the onset, middle, and the end of the target vowel, each spanning a one-third window of the entire duration. The values of the creak parameters linearly interpolated between the critical "creak" region and the modal portion of the vowel. For Tone 1, the creak attributes were added throughout the entire duration, because no portion of that tone is associated with a low F0 or creak. The "creak" portion had one of four manipulations: low spectral tilt, irregular F0, pitch doubling, and extra-low F0. The values of parameters used in Klatt synthesizer are shown in Table 1.

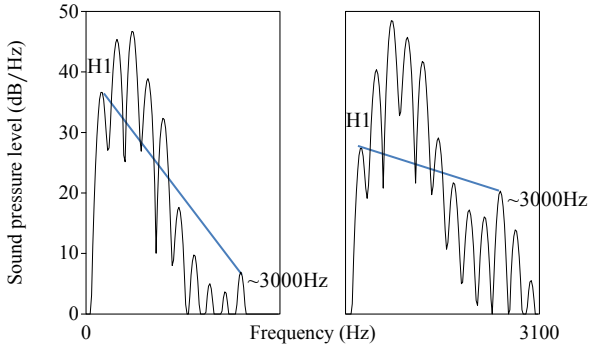
Table 1: Klatt parameters for creaky stimuli.

Parameter	Value
Open Phase (low H1–H2)	0.2
Spectral tilt (low H1–3000Hz)	0
Flutter (irregular F0)	0.95
Double pulsing (pitch doubling)	0.15
Extra-low F0	by 3 semitones

Stimuli with low spectral tilt were created by lowering two parameters: *Open Phase* (H1–H2) and *Spectral Tilt* (H1–3000 Hz), resulting in less energy in the first harmonic relative to the second harmonic and higher frequencies (Fig. 2). Note that, although more energy in higher frequencies is associated with creak, [20] also showed that low spectral tilt could bias English listeners to perceive a higher pitch.

To illustrate an example: in Tone 2 the parameter of *Open Phase* was set to the low value of 0.2 throughout the first third of the vowel duration, after which it rose linearly to a modal preset value of 0.5 at the end of the vowel. Consequently, this manipulation spanned the first ⅓ portion of the vowel and linearly interpolated to the modal value at the end of the vowel.

Figure 2: Spectrum for low spectral tilt of T1 /jou/. (Left: modal; Right: low spectral tilt) Note the H1−3000 Hz slope.



Irregular F0 was manipulated using the *Flutter* parameter to create jitter, or pulse-to-pulse variation in F0 (Fig. 3). Pitch doubling, indicating an alternating pitch cycle, was created using the *Double Pulsing* parameter (Fig. 4). The specific parameter values were chosen so as to create a percept of creak while still ensuring a natural voice quality.

Stimuli with extra-low F0 were created by lowering the minimum F0 value in the modal stimuli by three semitones for each tone.

Figure 3: Spectrogram for irregular F0 of T1 /jou/.

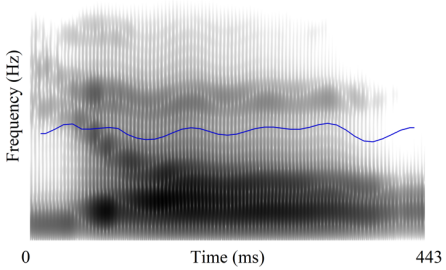
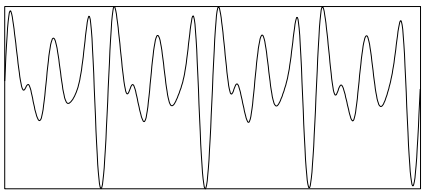


Figure 4: Waveform for pitch doubling of T1 /jou/. Note the alternation of strong and weak pulses.



2.2. Participants and procedure

Thirty-two (6 males, 26 females; mean age: 20.16; range: 18-24) native Mandarin Chinese speakers were recruited from the undergraduate population at UCSD. Participants had no prior exposure to auditory training and no hearing or language disorders reported.

The experiment, which was implemented in PsychoPy, consisted of two blocks of forced-choice

tonal identification tasks among the four Mandarin tones. The experiment took place in a sound-attenuated booth and was presented over headphones. The order of the stimulus presentation with corresponding Chinese characters and tone labels was randomized across participants. Listeners were asked to choose among four characters which word they heard by pressing keyboard buttons. Stimuli were played at a fixed volume with an interstimulus interval of 500 ms following the previous response.

The stimulus set contains 4 *target words* x 4 *tones* (T1, T2, T3, T4) x (4 *creak attributes* (low spectral tilt, irregular F0, pitch doubling, extra-low F0) + 4 repetitions of *modal* voice) x 2 block repetitions, resulting in 256 stimuli in total. Creaky and modal tokens were balanced across trials.

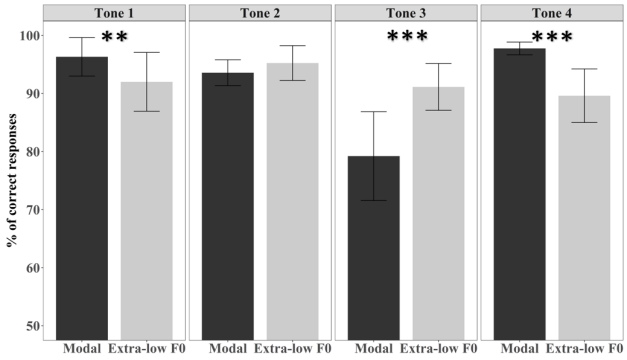
3. RESULTS

We used generalized linear and linear mixed-effects model comparisons (*lme4* package in R) for response accuracy and log-transformed response time, and investigated the roles of tone and manipulation of the acoustic attributes (modal, low spectral tilt, irregular F0, pitch doubling, extra-low F0) as fixed factors, with random intercepts for subjects, words, and repetitions. Trials in which participants took less than 50 ms or longer than 2000 ms to respond after the stimulus were excluded; after z-score normalization within participants, outliers greater or less than 2.5 standard deviations from the individual means were also excluded (in total, 5.9% trials of the overall data).

Accuracy varied as a function of tone (χ^2 (3) = 391.1, p < .001), as well as the interaction between tone and manipulation (χ^2 (12) = 71.15, p < .001). There was no main effect of manipulation (χ^2 (4) = 0.60, p = .96). All the creak attributes in the manipulation were compared to the mean-centered modal tokens. These results indicate that single attributes of creaky voice did not affect accuracy independently of tone; instead, they differently affected the identification of each tone. Extra-low F0 was the only creak attribute that played a significant role in tone identification (β = -0.86, p < .01).

Next, we review the results for each tone. Manipulation affected Tone 3 accuracy (χ^2 (4) = 27.32, p < .001), with extra-low F0 showing improvement in tone identification (β = 1.28, p < .001). Manipulation also affected Tone 4 accuracy (χ^2 (4) = 38.35, p < .001); extra-low F0 showed worsening (β = -1.61, p < .001), and a marginal effect of low spectral tilt showed improvement (β = 1.72, p = .087). The effect of manipulation was marginally significant for Tone 1 accuracy (χ^2 (4) = 8.83, p = .065). Extra-low F0 decreased Tone 1 accuracy (β = -1.01, p < .01). No effect of manipulation was found for Tone 2 accuracy (χ^2 (4) = 2.06, p = .72). The different effects of extra-low F0 on identification accuracy are illustrated in Figure 5.

Figure 5: Accuracy under modal and extra-low F0 manipulations across tones. (Error bars: 95% CI ± mean)



Overall, the four tones were all identified well above chance (25%), but Tone 3 showed the lowest identification accuracy (T1: 95.71%, T2: 93.99%, T3: 81.17%, T4: 97.03%). Specifically, the tone confusion patterns with the extra-low F0 stimuli are shown in Table 2.

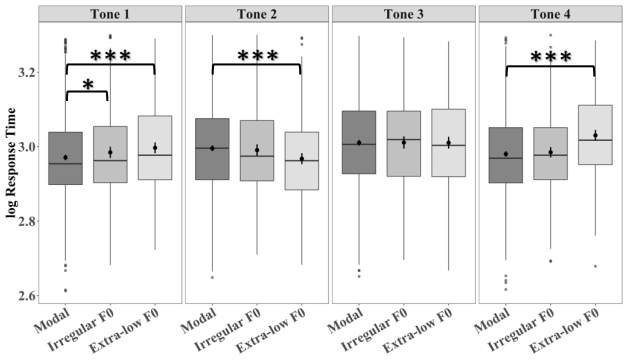
Table 2: Confusion matrix across extra-low F0 stimuli. (Responses are in percentage.)

Actual / Response	T1	T2	T3	T4
T1	**92.24**	0.83	0	0.42
T2	7.35	**95.04**	8.4	0.42
T3	0	3.72	**91.18**	9.28
T4	0.41	0.41	0.42	**89.87**

Log-transformed response time (log RT) varied as a function of tone (χ^2 (3) = 129.33, p < .001), manipulation (χ^2 (4) = 10.57, p < .05), and their interaction (χ^2 (12) = 78.27, p < .001). Irregular F0 (β = 0.016, p < .05) and extra-low F0 (β = 0.025, p < .001) delayed the response time overall.

Modal Tone 1 had the fastest responses compared to other modal tones (970.36 ms). Manipulation affected the log RT for Tone 1 (χ^2 (4) = 19.94, p < .001), with both irregular F0 (β = 0.017, p < .05) and extra-low F0 (β = 0.025, p < .001) slowing the response time. Manipulation affected the log RT for Tone 2 (χ^2 (4) = 22.55, p < .001); extra-low F0 *accelerated* the response time (β = -0.027, p < .001). Manipulation also had a significant effect on the log RT in Tone 4 identification (χ^2 (4) = 61.28, p < .001), with extra-low F0 slowing responses (β = 0.051, p < .001). No effect of manipulation was found for Tone 3 log RT (χ^2 (4) = 0.90, p = .92). Figure 6 shows the different effects of irregular F0 and extra-low F0 on log RT for each tone.

Figure 6: Log response time under modal, irregular F0, and extra-low F0 manipulations across tones. (Error bars: 95% CI ± mean)



4. DISCUSSION

Overall the results indicate that extra-low F0 and irregular F0 affected Mandarin tone identification accuracy and/or speed. The hindering effect of extra-low F0 and irregular F0 for Tone 1 identification confirms that T1 is perceived as a modal tone with a high pitch. Deviation from this F0 pattern by a low or irregular F0 could lead to lower Tone 1 identification. But in this study, the creak attributes were synthesized over the entire duration of T1, which may have induced stronger effects of creak (cf. Cantonese [24]). For Tone 2, extra-low F0 resulted in faster responses. We suspect this speaks to the nature of the F0 contour of the resynthesized stimuli. The F0 shape of T2 sloped gently during the first two thirds and rose up steeply thereafter. This could make T2 at the beginning sound more confusable with T1, so that listeners would need to wait until the late rise occurred to identify the tone; with an extra-low F0 added at the beginning, listeners might have responded faster given the initial rise as a cue to T2. For Tone 4, extra-low F0 always had negative effects on identification. The F0 contour in T4 spanned the largest pitch range, and adding an extra-low F0 at the end resulted in a dramatic-sounding pitch fall, which may have sounded more similar to T3.

Interestingly, though extra-low F0 improved Tone 3 identification, it did not affect response time. In fact, none of the other attributes of creak improved Tone 3 identification accuracy or speed, even though this tone is generally produced with creak (see similar findings for White Hmong in [12]).

The results speak to the fact that creak attributes are inherently different from one another. Low spectral tilt does not directly depend on F0, and can cue a higher pitch [20], which is at odds with creak induced by a low F0. Still, this attribute did not improve Tone 1 identification either. In production, low spectral tilt also often appears together with other attributes such as low and irregular F0, and as such might not be able to serve as an independent acoustic cue to creaky voice. In contrast, extra-low F0, pitch doubling, and F0 irregularity might affect pitch and tone perception more strongly because they directly change the F0 contour. But pitch doubling did not affect tone perception in the current study, possibly due to a percept of rough voice quality created by alternating pulses [13].

Thus, it is possible that individual acoustic correlates of creaky voice are not independently sufficient to affect tonal identification, and that low F0 must accompany the changes in voice quality. This suggests that the cues may integrate perceptually with low F0 in tonal contrasts, similar to the cues to voicing contrasts (see e.g., [16]). Furthermore, the current results (cf. [6, 9]) do not completely accord with studies that show effects of creak on T3 perception [2, 23]. This is likely due to different task designs. In previous studies, creak likely had co-occurring acoustic attributes, whereas here only single attributes were tested at a time. Thus, we see weaker effects of creak attributes in isolation. Future work will explore which combinations of the correlates play a role in pitch and tone perception, as well as stimuli with stretches of actual creaky voice. Indeed, how these correlates contribute to pitch/tone perception is still unclear, as seen from the various relations between pitch and voice quality in tone and register languages (such as Mandarin, Cantonese [24], Mpi [22], Mazatec [11], White Hmong [8], Black Miao [18], Zapotec [21]).

Nonetheless, the fact that cues to creaky voice other than extra-low F0 did not play a strong role corroborates the claim that creaky voice is caused by low-F0 targets in production, in accordance with previous production studies in Mandarin. It is also possible that the F0 cues were so robust that the creak attributes did not have a chance to affect tonal perception (see e.g., [9]). Follow-up experiments will look into this possibility by mitigating F0 robustness and strengthening creaky voice.

5. CONCLUSIONS

Mandarin listeners were most sensitive to an extra-low F0 among other creak attributes in a citation tone signal. This confirms a mapping from production to perception that creaky voice in Mandarin is caused by a low F0, which in turn is used by listeners for tone identification. The present study shows that probing different attributes of creak furthers our understanding of the relation between F0 and creaky voice, and between tone perception and production, which could have implications for practical applications such as tone recognition and simulation.

6. REFERENCES

[1] Belotel-Grenié, A., Grenié, M. 1994. Phonation types analysis in Standard Chinese. *Proc. 3rd ICSLP* Yokohama, 343-346.

[2] Belotel-Grenié, A., Grenié, M. 1997. "Types de phonation et tons en chinois standard" ("Phonation types and tones in standard Chinese"), *Cahiers de linguistique-Asie orientale*, 26, 249–279.

[3] Belotel-Grenié, A., Grenié, M. 2004. The creaky voice phonation and the organisation of Chinese discourse. *Proc. TAL-2004* Beijing, 5-8.

[4] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.37, retrieved 14 March 2018 from http://www.praat.org/

[5] Cai, Q., Brysbaert, M. 2010. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS one*, 5, e10729.

[6] Cao, R. 2012. *Perception of Mandarin Chinese Tone 2/ Tone 3 and the role of creaky voice*. Doctoral dissertation. University of Florida.

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

[8] Esposito, C. M. 2012. An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics*, 40, 466-476.

[9] Gårding, E., Kratochvil, P., Svantesson, J. O., Zhang, J. 1986. Tone 4 and Tone 3 discrimination in modern standard Chinese. *Language and Speech*, 29, 281-293.

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

[11] Garellek, M., and Keating, P. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, 41, 185-205.

[12] Garellek, M., Keating, P., Esposito, C. M., Kreiman, J. 2013. Voice quality and tone identification in White Hmong. *J. Acoust. Soc. Am.* 133, 1078-1089.

[13] Gerratt, B. R., Kreiman, J. 2001. Toward a taxonomy of nonmodal phonation. *Journal of Phonetics* 29, 365-381.

[14] Huang, Y., Athanasopoulou, A., Vogel, I. 2018. The Effect of Focus on Creaky Phonation in Mandarin Chinese Tones. *University of Pennsylvania Working Papers in Linguistics*, 24, 12.

[15] Keating, P., Garellek, M., Kreiman, J. 2015. Acoustic properties of different kinds of creaky voice. *Proc. 18th ICPhS* Glasgow, 0821-1.

[16] Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. 2008. On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, 36, 28-54.

[17] Klatt, D., Klatt, L. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J. Acoust. Soc. Am.* 87, 820-857.

[18] Kuang, J. 2013. *Phonation in Tonal Contrasts.* Doctoral dissertation. University of California, Los Angeles.

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

[20] Kuang, J., Liberman, M. 2015. Influence of spectral cues on the perception of pitch height. *Proc. 18th ICPhS* Glasgow.

[21] Pickett, V. B., Villalobos, M. V., Marlett, S. A. 2010. Isthmus (Juchitán) Zapotec. *Journal of the International Phonetic Association*, 40, 365-372.

[22] Silverman, D. 1997. Laryngeal complexity in Otomanguean vowels. *Phonology*, 14, 235–261.

[23] Yang, R. X. 2011. The Phonation factor in the categorical perception of Mandarin tones. *Proc. 17th ICPhS* Hong Kong, 2204-2207.

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