How do listeners identify creak? The effects of tone, pitch range, prosodic position and creak locality in Mandarin

Aini Li[1], Wei Lai[2] and Jianjing Kuang[1]

[1] (liaini, kuangj)@sas.upenn.edu, Department of Linguistics, University of Pennsylvania.
[2] wei.lai@Vanderbilt.edu, Department of Psychology and Human Development, Vanderbilt University

Background

- **Creaky voice**: an aperiodic phonation that is often related to low pitch targets (also known as "creak", "vocal fry" and "glottalization").
- Acoustic cues featuring creak: irregular pulses, low F0, constricted glottis, damped pulses and presence of subharmonics [1].
- As a non-modal phonation, creak has been found to influence the perception of pitch range, prosodic boundary and lexical tones [2, 3, 4, 5]
- However, few studies have examined how these factors could in turn affect listeners' perception of creak (e.g., [6])
- **This study**: examines the effects of pitch range, prosodic position, creak locality and lexical tones on creak identification in Mandarin

Method I

- **Experimental design**

A 8 (Tone) X 2 (Pitch range) X 2 (Prosodic position) X 3 (Creak locality) within-subject design was implemented.

Table 1: *A schema of the experimental design*

Sentence	Prosodic Position	Creak Locality
MMMMMMMCCCCC	SentenceFinal	Global creak
MMMMMMMMMMMC	SentenceFinal	Local creak
CCCCCMMMMMMM	SentenceNonfinal	Global creak
MCMMMMMMMMMM	SentenceNonfinal	Local creak

- **Materials**
  - Stimulus creation
    - 64 simple declarative sentences were constructed.
    - 12-syllable long with the same syntactic structure (NP1-VP-NP2) but varying in terms of the exact content and lexical items.
    - NP1 and NP2 are disyllabic person names and only the tone of the second syllable was different (X Y1 vs. X Y2).
    - Names were all sonorants.
    - Creak-containing syllables differed in terms of *prosodic position* (final vs. non-final), *pitch range* (high-pitched vs. low-pitched) and creak locality (global creak [the surrounding 4-5 syllables of the creak-containing target were creak] vs. local creak [only the creak-containing target was creak]).
    - For each target syllable, another two modal sentences were included to balance items with and without creak.
  - Recording and manipulation
    - 64 sentences were naturally produced by a female native speaker of Mandarin.
    - Recording was conducted in a professional sound booth using a high-quality BlueSnowball iCE microphone.
    - Sentences were produced in equalized speech rates (at 40 bpm using online metronome) with a sentence-final falling intonation.
    - Recordings were digitized at a sampling rate of 44,100kHz and 32 bit sample width.
    - Each sound file lasted for 2-3 seconds in duration.
    - Sentences were then manipulated into low-pitched targets (the mean F0 for the low-pitched recordings was 110 Hz and the mean F0 for the original high-pitched recordings was 225 Hz.

Method II

- **Participants**

A total number of 40 native speakers of Mandarin from the mainland of China participated into this study (self-reported gender: 8 men; 33 women; age: 19-36, mean = 25.12) and were paid 20 RMB for their participation. 24 of them reported that they have never heard the term "creaky voice" prior to the study. No one reported to have hearing deficits.

- **Procedure**
  - Conducted online in Mandarin Chinese using Qualtrics.
  - Consent ⟶ Familiarization phase (participants learned what creaky voice sounds like) ⟶ Practice trials (same as the test trials except with feedback) ⟶ Test trials
  - Task: identify for each sentence, whether and where they think creaky voice occurs.

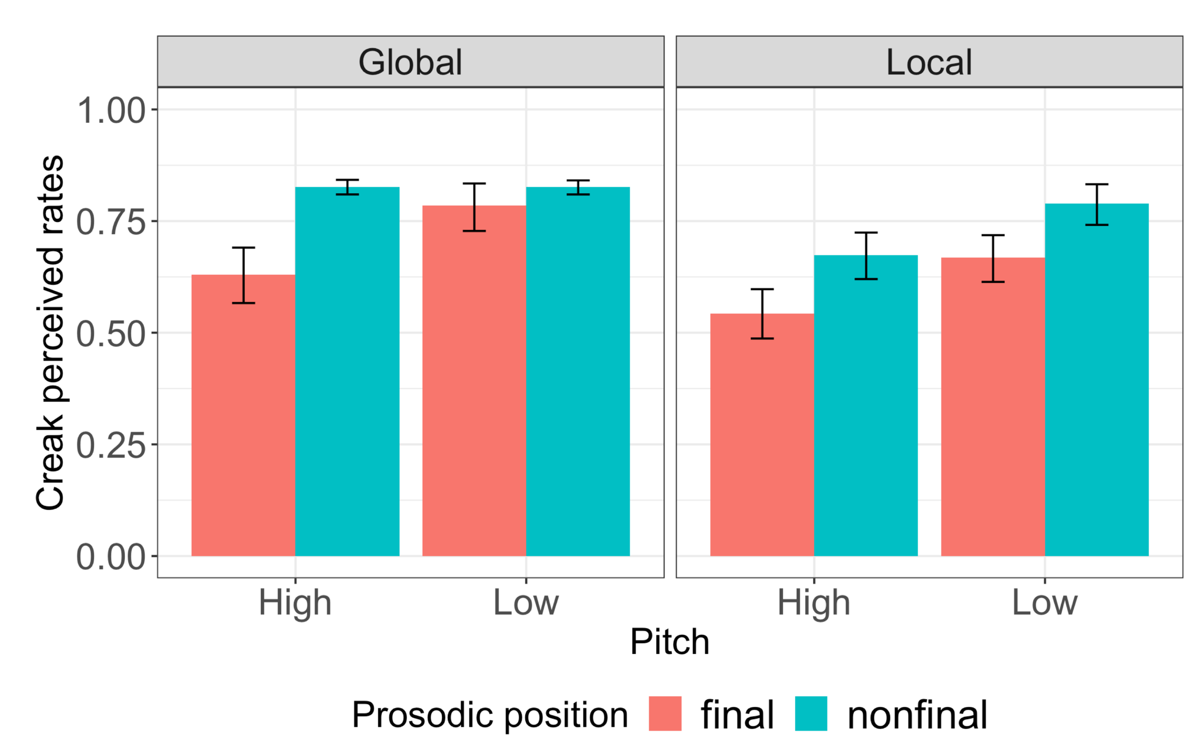
  - The experiment took around 25 minutes to finish.

Results I

- **Confusion matrix of creak identification**

Table 1: overall	Production-Creak	Production-Modal
Perception-Creak	Hit: 0.73	False Alarm: 0.05
Perception-Modal	Miss: 0.27	Hit: 0.95
Table 2: High | Low pitch	Production-Creak	Production-Modal
Perception-Creak	Hit: 0.69 | 0.76	False Alarm: 0.03 | 0.17
Perception-Modal	Miss: 0.31| 0.24	Hit: 0.97 | 0.83
Table 3: Final | Non-final	Production-Creak	Production-Modal
Perception-Creak	Hit: 0.65 | 0.81	False Alarm: 0.07 | 0.03
Perception-Modal	Miss: 0.35 | 0.19	Hit: 0.93 | 0.97
Table 4: Global | Local	Production-Creak	Production-Modal
Perception-Creak	Hit: 0.88 | 0.69	False Alarm: 0.01 | 0.07
Perception-Modal	Miss: 0.12 | 0.31	Hit: 0.99 | 0.93
Table 5: T1 | T2 | T3 | T4	Production-Creak	Production-Modal
Perception-Creak	Hit: 0.85 | 0.55 | 0.91 | 0.59	False Alarm: 0.08 | 0.13 | 0.10 | 0.09
Perception-Modal	Miss: 0.15 | 0.45 | 0.09 | 0.41	Hit: 0.92 | 0.87 | 0.90 | 0.91
Table 6: N1 | N2 | N3 | N4	Production-Creak	Production-Modal
Perception-Creak	Hit: 0.64 | 0.77 | 0.62 | 0.75	False Alarm: 0.05 | 0.09 | 0.07 | 0.08
Perception-Modal	Miss: 0.36 | 0.23 | 0.38 | 0.25	Hit: 0.95 | 0.91 | 0.93 | 0.93

- **Creaky syllables**



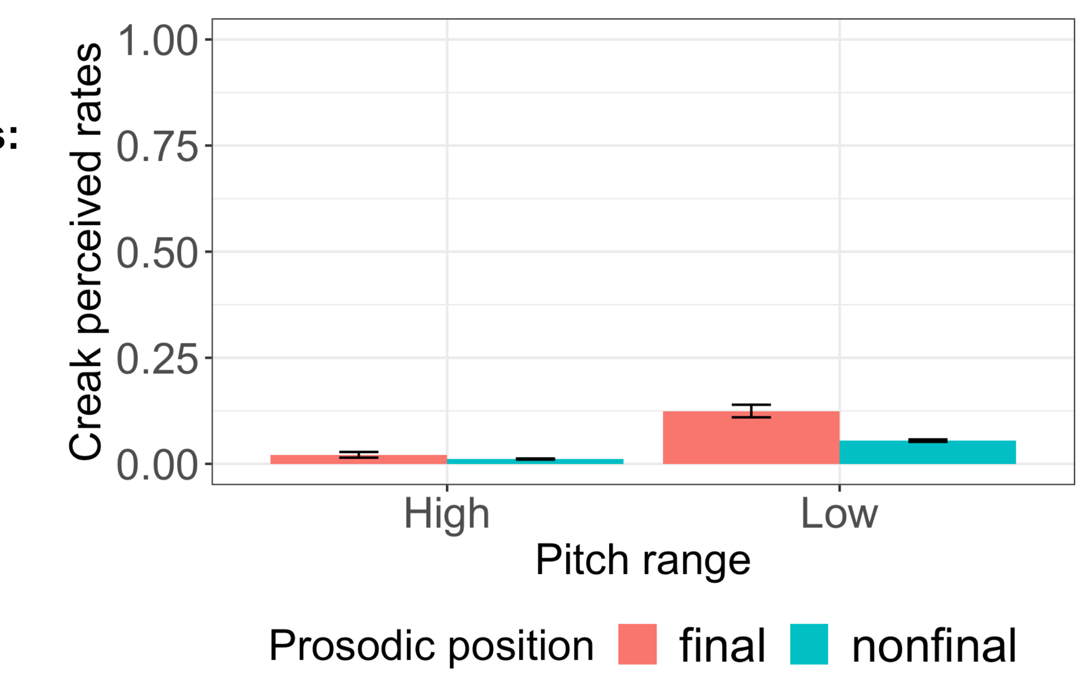
Results II

- Mixed-effects modeling of creaky syllables

Estimate	Std.	Error	z value	Pr(>|z|)
(Intercept)	1.19	0.26	4.51	<0.001 ***
Pitch range (High)	-0.29	0.05	-6.44	<0.001 ***
Prosodic position (Final)	-0.28	0.06	-4.55	<0.001 ***
Creak locality (Global)	0.32	0.05	5.89	<0.001 ***
Pitch range (High) : Prosodic position (Final)	-0.10	0.05	-2.23	0.03 *
Pitch range (High) : Creak locality (Global)	0.08	0.05	1.74	0.08 .
Prosodic position (Final) : Creak locality (Global)	-0.22	0.05	-3.96	<0.001 ***
Pitch range (High) : Prosodic position (Final) : Creak locality (Global)	-0.13	0.05	-2.80	<0.01 **

Formula: Response $\sim$ Pitch range $\times$ Prosodic position $\times$ Creak locality + (1|Participant) + (1|Syllable) + (1|Tone)

- **Modal syllables:**



- Mixed-effects modeling of modal syllables

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-4.68	0.19	-24.10	<0.001 ***
Pitch range (High)	-0.94	0.05	-20.00	<0.001 ***
Prosodic Position (Final)	0.06	0.06	1.10	0.27
Pitch range (High) : Prosodic Position (Final)	-0.07	0.05	-1.46	0.14

Formula: Response $\sim$ Pitch range $\times$ Prosodic position + (1|Participant) + (1|Syllable) + (1|Tone)

Discussion & Conclusion

- How Mandarin listeners identify creak is influenced by pitch range, prosodic position, creak locality as well as lexical tones
- For creaky syllables, identification would be more accurate if the target syllables were in a context where its surrounding syllables were also creaky; creaky syllables at sentence non-final positions are easier to be identified; high-pitched syllables are less likely to be perceived as creaky
- For modal syllables, when they are in a creaky environment, identification is easier when it is high-pitched; whereas when they are in a modal environment, low-pitched targets are more likely to trigger false alarm.
- Taken together, the perception of creak is highly context-dependent and modulated by different acoustic and linguistic cues.

References

[1] P. A. Keating, M. Garellek, and J. Kreiman, "Acoustic propertiesof different kinds of creaky voice." in CPhS, 2015. [2] 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, pp. 1320–1333,2014. [3] J. Kuang and M. Liberman, "Pitch-range perception: The dynamic interaction between voice quality and fundamental frequency." in InterSpeech, 2016, pp. 1350–1354. [4] G. Kuo, "Perceived prosodic boundaries in taiwanese and their acoustic correlates," in Thirteenth Annual Conference of the International Speech Communication Association, 2012. [5] L. Davidson, "The effects of pitch, gender, and prosodic context on the identification of creaky voice,"Phonetica, vol. 76, no. 4,pp. 235–262, 2019.

Acknowledgement: We thank members of the Upenn Phonetics Lab and Mark Garellek for their feedback.