The effects of vocal fry and period doubling on the perceived naturalness of Mandarin tones

Yaqian Huang

Acoustics Research Institute, Austrian Academy of Sciences

Voice quality cues often contribute to tone perception. For Mandarin, creaky voice was found to facilitate the identification of the low-dipping Tone 3 in naturalistic [1] or resynthesized stimuli [2]. However, previous studies have not probed the role of finer-grained phonetic details of creak in tonal representation. Here we focus on two commonly-occurring types of creak, period doubling, and vocal fry, that are allophonic to Mandarin tones [3]. Vocal fry is often perceived as low pitched [4], while period doubling has ambiguous pitches depending on amplitude or frequency modulation [5]. This study investigates how naturally-occurring period doubling and vocal fry, compared to modal voice, contribute to the perception of Mandarin tones in terms of naturalness and representativeness. It is expected that lower tones with vocal fry and higher tones with modal voice will sound more natural. As for period doubling, the effects will depend on the modulation type. We expect that lower tones with modulations involving frequency will sound more natural because a low pitch is more probable, and higher tones with modulations only involving amplitude will be more natural because a higher pitch is more probable. Alternatively, roughness or indeterminate pitch resulting from concurrent frequencies of period doubling may hinder tone identification for all tones, resulting in a less natural or representative percept.

Methods. The stimuli were 80 tokens of naturalistic Mandarin tones (20 per tone) with varying voice qualities extracted from a read corpus [6]. A balanced set of tokens of modal voice, vocal fry, and period doubling with amplitude and/or frequency modulation across five female speakers were selected. The portions of the voice quality especially the creaky ones were similarly distributed throughout the syllable across different tokens. All tokens were silence-padded to be 500 ms long and scaled to 70 dB. 36 native Mandarin speakers (24 F) in college participated. They passed a hearing prescreening test to ensure normal hearing. A visual sort-and-rate task was implemented (following [7,8]). Each tone category was presented in a separate trial. At the beginning of a trial, stimuli were represented as sound icons at the top of a computer screen in a random order. Listeners played the sounds by clicking the icons, and then dragged each icon onto a horizontal scale that stood for naturalness or representativeness. If the participants placed a token more toward the right corner rather than the left, they found the token more natural or representative of the specified tone category. What counts for naturalness or representativeness is subject to participants' own interpretation. **Fig. 1** shows a sample trial of Tone 1 with responses.

Results. The resulting positions of all stimuli (in numerical values) were min-max normalized within each participant according to individual usages of the horizontal scale. A linear mix-effects model was used to evaluate the fixed effect of *tone*, *voice quality*, and their interactions on the naturalness ratings while taking into account the random effects of participants and tokens. The high-level Tone 1 and modal voice were used as the baseline for comparison. The low-dipping Tone 3 with modal voice had lower ratings (β=-.43, p<.001), and vocal fry in general also had lower ratings (β=-.36, p<.01). However, Tone 3 was rated more natural when produced with vocal fry (β=.52, p<.001) or period doubling with combined frequency and amplitude modulation (β=.41, p<.01). All the other tones were rated most natural with modal voice and least with vocal fry. The rising and falling Tones 2 and 4 produced with amplitude-modulated period doubling were also rated lower (βs<-.30, p<.05). **Fig. 2** shows the patterns.

Discussion. As expected, vocal fry with its low-pitch percept improves the perceived naturalness of the low-dipping Tone 3, and period doubling with frequency modulation had a similar effect, consistent with previous studies. Interestingly, the effect of period doubling with amplitude modulation was ambiguous: it patterned with modal voice for the high Tone 1 and low Tone 3 but with vocal fry for the rising Tone 2 and falling Tone 4. This connects to the findings on the pitch percept during amplitude-modulated period doubling in [5] where a low opposed to high tone was perceived for about 50% percent of the time, which implies a bitonal percept. It is possible that for more 'unambiguous' tones like Tones 1 and 3, amplitude modulation conforms to and contributes to the dominant high or low pitch; for Tones 2 and 4 which are intermediate, amplitude modulation affected the contours unfavorably. Altogether, the results suggest that the use of period doubling is versatile in different tone contours, and that creaky voice is integrated in Tone 3.
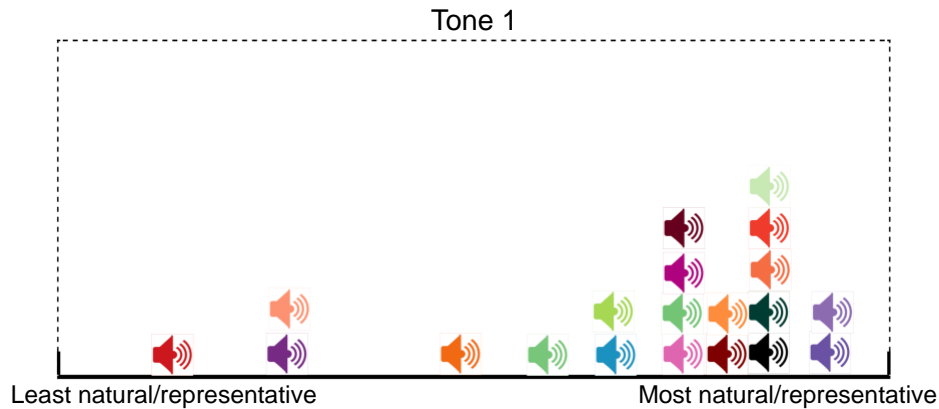
Fig. 1. A sample illustration of the visual sort-and-rate task for Tone 1 with one participant's responses.
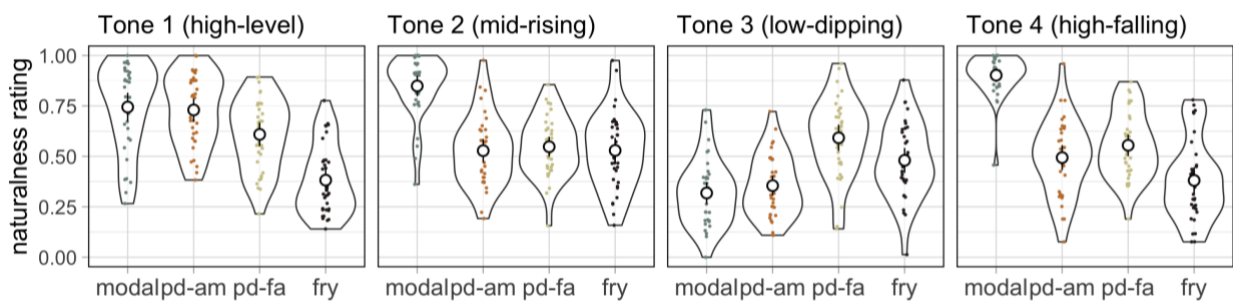

Fig. 2. Naturalness ratings conditioned by voice quality in Mandarin tones. pd-am: period doubling with amplitude modulation; pd-fa: period doubling with frequency and amplitude modulation.

References
[1] Belotel-Grenié, A., & Grenié, M. (1997). Types de phonation et tons en chinois standard. *Cahiers de linguistique-Asie orientale*, *26*(2), 249-279.
[2] Huang, Y. (2020). Different attributes of creaky voice distinctly affect Mandarin tonal perception. *The Journal of the Acoustical Society of America*, *147*(3), 1441-1458.
[3] Yu, K. M. (2010). Laryngealization and features for Chinese tonal recognition. *INTERSPEECH-2010*, pp. 1529-1532.
[4] Kuang, J. & Liberman, M. (2016). The effect of vocal fry on pitch perception. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5260-5264.
[5] Huang, Y. (2023). Perception and imitation of period doubling. *Proceedings of the 20th International Congress of Phonetic Sciences*. Prague, pp. 327-331.
[6] Huang, Y. (2022) Articulatory properties of period-doubled voice in Mandarin. *Proc. Speech Prosody 2022*, pp. 545-549.
[7] Granqvist, S. (2003). The visual sort and rate method for perceptual evaluation in listening tests. *Logopedics Phoniatrics Vocology*, *28*(3), 109-116.
[8] Kreiman, J., Garellek, M., Chen, G., Alwan, A., & Gerratt, B. R. (2015). Perceptual evaluation of voice source models. *The Journal of the Acoustical Society of America*, *138*(1), 1-10.