

The role of phonation type in Chinese Jin tones: a study using acoustic metrics

Xiaoqing Du, Francis Nolan, Brechtje Post

Tonal contrasts in the Huoji¹ dialect classified as a Jin Chinese variety in the *Language Atlas of China* (2nd ed.) [1] within the Han-Xin subgroup, are undergoing convergence that compresses f₀ space and brings the historical checked category (T5) close to the level category (T1). Checked syllables are frequently produced with glottal constriction and irregular excitation at the end of the rhyme, in the final portion of the syllable (offset-adjacent to any coda closure). This raises a question: if f₀ space is compressed and rhyme offsets are least periodic, which acoustic cues remain reliable indicators of a T5–T1 distinction? 17 native Huoji speakers (23–79 years) from four localities near Mandarin-speaking areas produced 70 monosyllabic citation forms (14 per tone, T1–T5) in five repetitions, yielding 5,950 tokens. Tone-bearing units were segmented as vowel nuclei plus voicing-carrying codas, enabling time-aligned extraction of TBU duration and time-normalised f₀ (nine points). To diagnose algorithm dependence under weak periodicity, f₀ was estimated with three waveform-based tracker families under harmonised low-f₀ settings: Praat [2], YIN [3], and REAPER [4] [5]. Voice-quality measures were extracted in VoiceSauce [6], focusing on low-frequency HNR05 (0–500 Hz) and the formant-corrected harmonic amplitude difference (H1*–H2*) [7]; phonation type (modal, creaky, whispery) was also coded auditorily. Across trackers, f₀-only representations yield a stable tonal geometry in which T5 remains closest to T1, while tracker divergences concentrate at rhyme offsets. Metric multidimensional scaling shows that adding HNR05 and H1*–H2* expands the tonal space and robustly separates T5 from T1, supporting a multi-cue account of checked-tone identity in which phonation contributes an interpretable dimension where periodicity degrades. This acoustic pattern motivates a perception link: when tonal categories are supported by multiple co-varying cues, f₀ may be necessary but not sufficient for robust categorisation, especially under syllable-final rhyme irregularity [8], [9]. Perceptual testing should therefore manipulate f₀ and phonation jointly to assess whether listeners rely on voice-quality cues when offset-adjacent f₀ becomes unreliable.

[1] Chinese Academy of Social Sciences and City University of Hong Kong, Eds., *Language Atlas of China*, 2nd ed. Beijing: Commercial Press, 2012, vol. Chinese Dialects.

[2] P. Boersma, “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,” in *Proc. Institute of Phonetic Sciences, Univ. of Amsterdam*, 1993, pp. 97–110.

[3] A. de Cheveigné and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930, 2002.

[4] D. Talkin, “A robust algorithm for pitch tracking (RAPT),” in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Amsterdam, The Netherlands: Elsevier, 1995, pp. 495–518.

[5] D. Talkin, “Reaper: Robust epoch and pitch estimator,” <https://github.com/google/REAPER>, 2015, google Research, GitHub repository, archived Feb. 3, 2023.

[6] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, “VoiceSauce: A program for voice analysis,” in *Proc. ICPhS*, 2011, pp. 1846–1849.

[7] M. Iseli, Y.-L. Shue, and A. Alwan, “Age, sex, and vowel dependencies of acoustic measures related to the voice source,” *J. Acoust. Soc. Am.*, vol. 121, no. 4, pp. 2283–2295, 2007.

[8] J. Kuang and M. Liberman, “Integrating voice quality cues in the pitch perception of speech and non-speech utterances,” *Frontiers in Psychology*, vol. 9, Art. no. 2147, 2018.

[9] R. Kreiman, J. Gerratt, and C. M. Garellek, “Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research,” *J. Speech Lang. Hear. Res.*, vol. 57, no. 3, pp. 925–950, 2014.

¹ In the *Language Atlas of China* (2nd ed.), Jin is treated as a major northern Sinitic group distinct from Mandarin. Within Jin, Huoji is classified as a cluster of the Han–Xin subgroup, spoken across southeastern Shanxi, southern Hebei (including Handan), and northern Henan (including Xinxiang). Huoji is not independently encoded in ISO 639-3 or Glottolog; the identifiers reported here refer to Jin Chinese (ISO 639-3: cji; Glottocode: jiny1235).