

Pitch relationship and phonation cues in Mandarin tone perception

Ok Joo Lee, Kyungmin Lee
Seoul National University, Korea

<https://doi.org/10.36505/TheLinguisticProceedings/2025/16/01/015/000675>

Abstract

This study investigates how pitch relationship and phonation cues shape Mandarin tone perception among native and non-native listeners. Focusing on high and low tones (Tones 1 and 3) in Mandarin, we examined the effects of pitch relationships between syllables and phonation type (modal vs. creaky) across five groups: native Mandarin listeners, and Cantonese and Korean listeners with low or high Mandarin proficiency. Results from a tone identification task with 140 participants show that the pitch relationship was the primary cue, while creaky voice notably facilitated Tone 3 responses. Native listeners were most sensitive to pitch, whereas Cantonese and Korean listeners relied differently on phonation cues depending on L1 background and L2 proficiency, revealing distinct perceptual adaptation strategies.

Keywords: pitch relationship, phonation, Mandarin, tone, perceptual strategies

Introduction

Mandarin has four lexical tones, ‘55’, ‘35’, ‘214’, and ‘51’, in Chao’s five-number scale (Chao 1930), conventionally labeled Tones 1 to 4. Tone 3 surfaces as a low tone when followed by Tones 1, 2, or 4, and as Tone 2 before another Tone 3. The low-tone realization predominates in the lexicon and natural speech (Duanmu 2000/2007; Zhang 2010). While pitch height and movement are crucial for tone identification, pitch relationships between syllables and phonation also affect perception: creaky voice, often produced with Tone 3, biases listeners toward Tone 3 identification (Kuang 2017; Huang 2020; Lee and Lee 2022). The goal of this study is to examine how pitch relationship and phonation cues, which are not inherent to the target tone itself, interact in native and non-native perception of Mandarin Tones 1 and 3. A tone identification task was conducted with 140 participants, including native Mandarin listeners as well as Cantonese and Korean listeners with differing levels of Mandarin proficiency, to assess the effects of L1 background (tone vs. non-tone) and L2 proficiency (low vs. high).

Methods

A tone identification experiment was conducted with five listener groups: 32 native Mandarin listeners (Man), 22 high-proficiency Cantonese listeners (CanH), 26 low-proficiency Cantonese listeners (CanL), 30 high-proficiency Korean listeners (KorH), and 30 low-proficiency Korean listeners (KorL). Mandarin proficiency was classified primarily based on *Putonghua Shuiping Ceshi* or *Hanyu Shuiping Kaoshi* scores and the length of Mandarin study.

The stimuli were disyllabic expressions with a level tone on the first syllable and a rising tone on the second, perceived as either T1+T2 or T3+T2 depending on first-syllable pitch height. Nine segmentally identical pairs, each with a plosive onset and high, mid, or low vowel, were used (e.g., *baohan* [pao.xan], “to contain; to be filled with”). Recordings by one male and one female Mandarin speaker were resynthesized, manipulating F0 along an 11-level continuum and first-syllable phonation type. The second-syllable onset pitch was fixed at midrange to control contrast effects. A total of 360 stimuli were presented individually via *Labvanced* in eight counterbalanced blocks, and participants identified each as T1+T2 or T3+T2.

The tone identification responses were analyzed using the R package *Hmisc* (Harrell 2022). Fixed factors included listener group (Man, Can, Kor), first-syllable phonation (modal, creaky), $\Delta F0$ (Syllable 1 height minus Syllable 2 onset), onset type, pitch relationship (Syllable 1 < Syllable 2 or Syllable 1 > Syllable 2), and speaker gender (M, F), with participants treated as random effects. A mixed-effects logistic regression tested main and interaction effects of listener group with other variables. *Bonferroni* post hoc analyses were conducted, with a 95% confidence interval applied.

Results

Figure 1 summarizes the effects of F0 and phonation on tone perception. All groups except KorL showed categorical perception, with the F0 boundary increasing in the order KorH < Man < CanL < CanH. KorH exhibited a broader Tone 1 category, whereas CanL and CanH showed broader Tone 3 categories. Creaky voice enhanced Tone 3 identification across groups, particularly among Korean listeners. For KorL, however, it did not yield categorical perception but instead enhanced low-pitch perception acoustically, indicating limited reliance on phonation for tonal contrast.

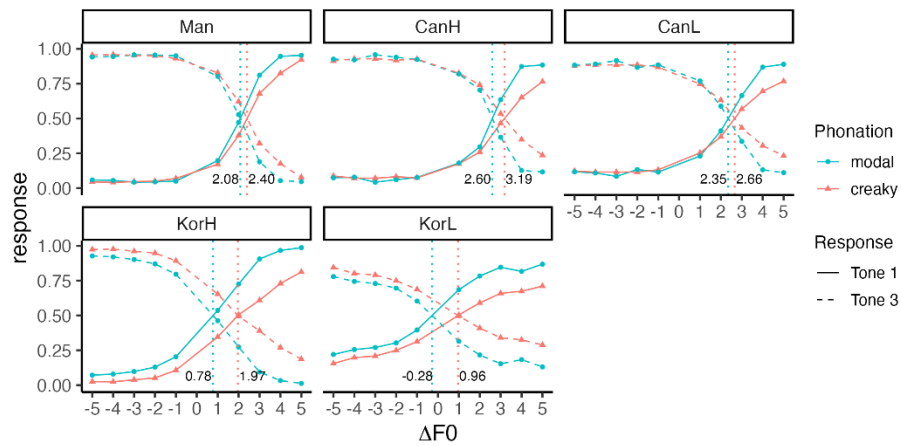


Figure 1. $\Delta F0$ and phonation effects across listener groups

The mixed-effects logistic regression model examined the effects of pitch, phonation, and language experience factors. Significant main effects were found for the S1–S2 pitch relationship, $\Delta F0$, phonation type, and listener group (all $p < .0001$), with pitch relationship and $\Delta F0$ as the strongest predictors ($\chi^2 = 8,079.04$ and $7,659.97$). Creaky voice elicited more Tone 3 responses than modal voice, and the S1 < S2 condition produced more Tone 3 responses than S1 > S2. KorL yielded the fewest Tone 3 responses. Significant pitch–phonation and group–cue interactions indicate that pitch effects are modulated by phonation and language experience. *Bonferroni* post-hoc analyses further demonstrated the effects of L1 background and L2 proficiency on both the S1–S2 pitch relationship and phonation cues. In the S1 < S2 condition, Tone 3 identification ranked Man, CanH, CanL, KorH > KorL ($p < .0001$), indicating weaker Tone 3 perception for KorL. In the S1 > S2 condition, Tone 1 identification ranked KorH, KorL > Man > CanH, CanL, reflecting L1 background effects. Under modal voice, Tone 1 identification ranked KorL > KorH > Man, CanH, CanL, whereas under creaky voice, Tone 3 identification ranked Man, CanH, CanL, KorH > KorL, indicating that creaky phonation facilitated Tone 3 identification for KorH. Post-hoc analyses of the pitch–phonation–group interaction revealed distinct Tone 3 identification patterns: in the S1 < S2 condition, KorL showed the lowest Tone 3 perception across phonation types (all $p < .0001$), whereas KorH approached native levels under creaky voice. Mandarin listeners outperformed CanL and KorH under modal but not creaky voice, suggesting their stable reliance on pitch cues. In S1 > S2 conditions, Korean listeners showed the highest Tone 1 identification, which decreased with creaky voice, while Cantonese listeners consistently favored Tone 3 identification.

Conclusion

The findings of the present study highlight the impact of language experience on tone perception. Mandarin and Cantonese listeners, unlike Korean listeners, relied more on contextual pitch relationship cues, interpreting tones relative to adjacent syllables. By contrast, Korean listeners attended mainly to target pitch height, though high-proficiency learners showed greater use of contextual cues, resembling native tone-language patterns. Tone-language listeners also differed notably. Compared with Mandarin listeners, Cantonese listeners showed a stronger bias toward Tone 3, likely due to their L1 tone system, which includes two low-level tones ('22', '33'). Phonation also significantly influenced tone perception across listener groups; creaky voice enhanced Tone 3 identification, especially among high-proficiency Korean listeners, indicating adaptive cue weighting with experience. Unlike tone-language listeners, they relied more on phonation, reflecting the perceptual salience of creaky voice.

Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government [NRF-2022S1A5A2A01043267].

References

- Chao, Y. R. 1930. A system of tone letters. *Le Maître Phonétique* 45, 24-27.
- Duanmu, S. 2000/2007. *The Phonology of Standard Chinese*. Oxford, Oxford University Press.
- Harrell, F. 2022. Hmisc: Harrell Miscellaneous. R package version 4.7-1. <https://CRAN.R-project.org/package=Hmisc>
- Huang, Y. 2020. Different attributes of creaky voice distinctly affect Mandarin tonal perception. *The Journal of the Acoustical Society of America* 147(3), 1441.
- Kuang, J. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America* 142, 1693.
- Lee, K., Lee, O.J. 2022. Native and non-native perception of Mandarin level tones. *Linguistic Research* 39(3), 567-601.
- Zhang, J. 2010. Issues in the analysis of Chinese tone. *Language and Linguistics Compass* 4(12), 1137-1153.