

A study on the speaker discriminatory power of the first five formants in six vowels of Mandarin for Forensic Speaker Comparison

LEI Yijing¹ Honglin^{1,2}

(1 Institute of Evidence Law and Forensic Science, China University of Political Science and Law, Beijing 100088, China) (2 Fada Institute of Forensic Medicine & Science, Beijing 100192, China)

1. Introduction: Formants are among the most widely used acoustic features in forensic speaker comparison, as they encode both segmental information and speaker-specific characteristics^[1-3]. However, systematic investigations into the speaker discriminatory power of higher formants, particularly the fifth formant (F5) in Mandarin monophthongs remain scarce. This study aims to address this gap by quantitatively evaluating the speaker discrimination performance of the first five formant center frequencies (F1–F5) across different Mandarin vowels, with particular attention to the potential value and limitations of F5.

2. Methods: 53 native male speakers of Beijing Mandarin (aged 18–23) were recorded twice, four months apart, producing six monophthongs (/a/, /i/, /ɤ/, /u/, /y/, /ɿ/) in both sustained and connected speech. Recordings were conducted in a sound-attenuated booth using a Sennheiser E945 microphone at 44.1 kHz. Stable vowel segments were manually selected, and F1–F5 center frequencies were extracted via Wavesurfer (Snack algorithm). Speaker discrimination performance was evaluated using Linear Discriminant Analysis (LDA), yielding correct classification rates. Spearman's rank correlation assessed formant co-variation, while repeated-measures ANOVA examined inter-vowel differences.

3. Results: Descriptive statistics showed that inter-speaker variability increased with formant order. For the vowel /a/, F5 exhibited the largest variation, whereas F1 showed the smallest. Spearman correlations showed a consistent, highly significant positive relationship between F4 and F5 across all vowels ($p < 0.001$), suggesting shared anatomical influences (Figure 1). LDA results demonstrated that single-formant models performed poorly (4.5%–12.3% accuracy), but multi-formant combinations dramatically improved discrimination. The full F1–F5 model achieved classification rates of 58.5% to 71.7% across vowels. Notably, models including F4 and F5 consistently outperformed those limited to F1–F3 (Figure 2). Among vowels, /ɿ/ yielded the highest discrimination (71.7%), followed by /a/ (66.4%), while rounded vowels (/u/, /y/) showed lower performance. Furthermore, F1 and F2 differed significantly across nearly all vowel pairs ($p < 0.001$), confirming their role in vowel identity, whereas F5 showed no significant differences between most vowels—supporting its speaker-specific rather than phonemic function.

4. Conclusion: This study shows that higher formants, particularly F4 and F5, significantly enhance speaker discrimination in Mandarin beyond traditional F1–F2 analysis. While F5 offers marginal gains in accuracy, its measurement instability in real world recordings limits forensic reliability. In contrast, F4 strikes an optimal balance between discriminatory power and robustness, making it a recommended addition to standard formant-based protocols. However, its discriminative role in vowel classification is relatively limited, and F2 remains the most effective indicator for distinguishing vowel categories.

Keywords: forensic voice comparison, higher formants, Mandarin, monophthong, speaker discriminatory power

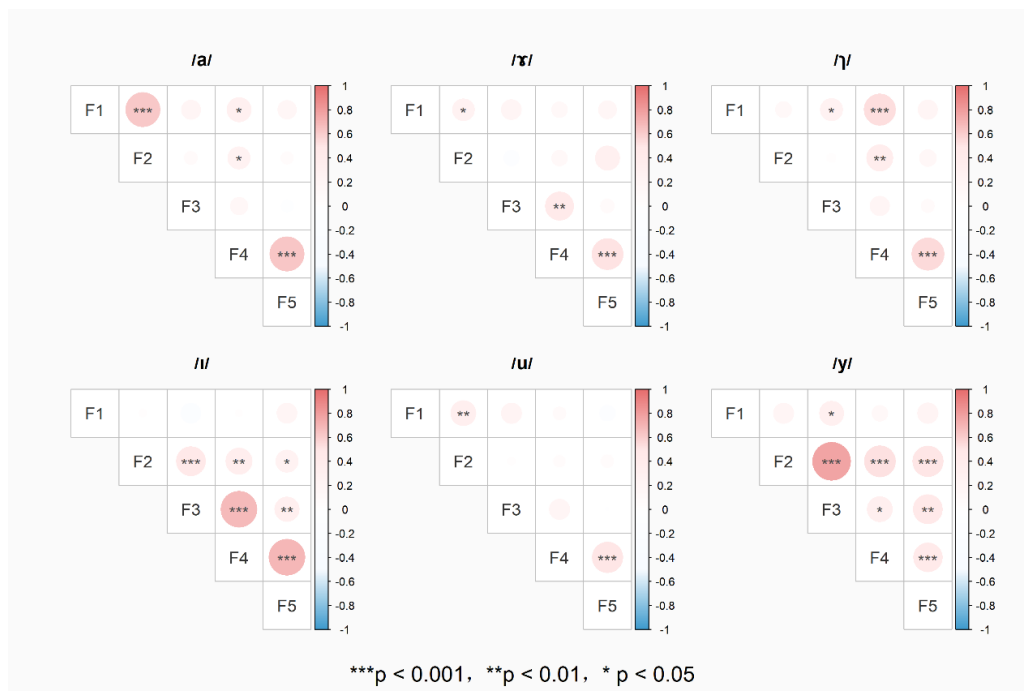


Fig. 1. The correlation matrix of the formants (F1-F5) for all the six vowels

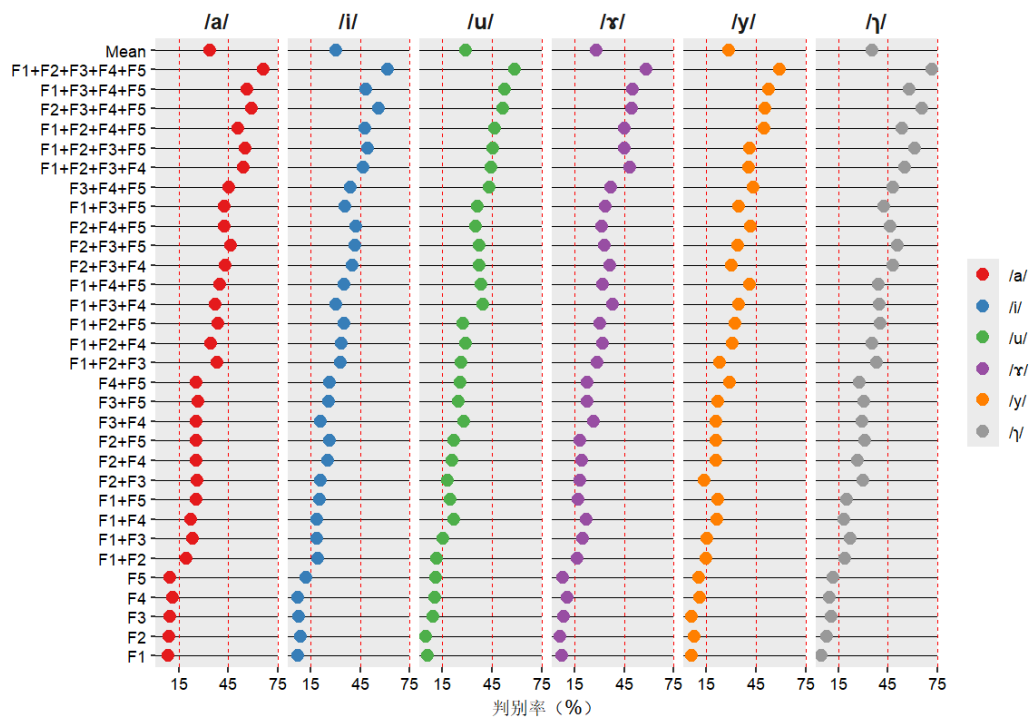


Fig. 2. Classification rates (CR) of the formants combinations across six vowels

References

- [1] Gold E, French P. (2019). International practices in forensic speaker comparisons. *The International Journal of Speech, Language and the Law*, 26(1), 1-20.
- [2] McDougall K. (2006). Dynamic features of speech and the characterization of speakers: toward a new approach using formant frequencies. *International Journal of Speech Language and the Law*, 13(1), 89-126.
- [3] Nolan F, Grigoras.(2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech Language and the Law*, 12(2), 143-173.