Speech Prosody 2026
26-29 May 2026, Philadelphia, Pennsylvania, USA

Mandarin vowels: Intrinsic properties and interaction with tone

Lorna McGregor-Smith, Irena Yanushevskaya

School of Linguistic, Speech and Communication Sciences
Trinity College Dublin
lmcgrego@tcd.ie, yanushei@tcd.ie

Abstract

This study examines how vowel quality and lexical tone interact with voice source behaviour in Mandarin Chinese. Six monophthongs (/i/, /y/, /u/, /ɤ/, /o/, /a/) were elicited in monosyllabic words under four tone conditions, both in carrier phrases and citation form, from ten female and two male native speakers. Relevant acoustic parameters (f_0, jitter, shimmer, HNR, CPPS) were extracted and compared across vowel and tone conditions using mixed-effect models.

Results suggest that different vowel qualities are associated with differences in voice quality parameters. In general, close vowels (/i/, /u/, /y/) showed higher f_0 and HNR and lower jitter and shimmer than /a/, indicating more regular phonation. /u/ exhibited lower CPPS than other vowels. The effect of different tonal environment (particularly Tones 3 and 4) on phonation was not the same for all vowels. Close vowels (/i/, /u/, /y/) tended to show less tone-related phonatory instability, whereas the opposite was found for close-mid back vowels (/ɤ/, /o/). These findings suggest that articulatory configurations and the realisation of tonal targets tend to co-vary.

The study provides preliminary evidence from a tonal language that source and filter characteristics interact in shaping phonatory outcomes, with potential relevance for tonal phonetics and applied speech research.

Index Terms: source-filter interaction, lexical tone, vowel acoustics, Mandarin Chinese, phonation, speech production

1. Introduction

According to the source-filter theory of speech production [1, pp. 15–21], speech signals originate from an acoustic source (airflow through vibrating vocal folds for voiced sounds or turbulent airflow for voiceless sounds) and are filtered by the vocal tract, where the acoustic energy of the source is shaped by resonances (formants) of the vocal tract. The theory assumes the source and the filter to be independent processes. However, they are known to interact, with vocal tract adjustments influencing properties of the laryngeal system and vice versa [2], [3]. Biomechanical studies show that vocal tract and laryngeal changes can interact to modulate voice quality [4].

Early work by Lieberman [2] proposed that changes in vocal tract configuration can affect the glottal cycle and thus influence source properties. Similarly, Stevens' [3] theory of the "quantal nature of speech" suggested that small articulatory variations can lead to non-linear acoustic effects, implying a coordination between source and filter. Ladefoged and Lindau [5] further demonstrated that such effects require compensatory articulatory strategies, meaning that the vocal tract actively interacts with the glottal source to maintain a desired acoustic output. Biomechanical studies corroborate this view: the larynx and vocal tract are anatomically and functionally interconnected, with alterations in one subsystem influencing the other through shared musculature and pressure dynamics [6], [7], [8]. Garellek [4] also notes that vowel quality, a filter property, can modulate voice quality, a source characteristic, through laryngeal constriction and resonance effects.

This interdependence is reflected in what have been described as intrinsic vowel properties such as fundamental frequency (f_0) and intensity. Close vowels tend to be produced with higher f_0 than open vowels, a pattern observed across languages [9], [10]. Physiologically, raising the tongue increases tension on the larynx and vocal folds, thereby elevating f_0. However, studies have reported exceptions to this pattern. For instance, Shu et al. [11] found no significant f_0 differences between close (/i/, /u/, /y/) and open (/a/) vowels under the level tone (Tone 1) in Mandarin, suggesting that vowel-pitch interactions may vary by language or context. Their study examined vowels produced under Tone 1 and did not consider the potential influence of lexical tone.

Various acoustic measures are used to objectively quantify aspects of vocal function and voice quality. Perturbation measures, namely, jitter (cycle-to-cycle variation in f_0) and shimmer (amplitude perturbation), reflect the stability of the glottal source. However, their relationship with vowel quality remains inconsistent. Some studies report lower jitter and shimmer for back or high vowels [12], [13], whereas others find higher values for low vowels or no systematic pattern at all [11], [14], [15]. Harmonics-to-Noise Ratio (HNR) [16] quantifies the relative amount of periodic (harmonic) energy to aperiodic (noise) energy in the voice signal and is related to the regularity of vocal fold vibration. MacCallum et al. [17] report that open vowels have lower HNR values. Shu et al. [11] found the close back /u/ vowel had significantly higher HNR values than other vowels. Cepstral Peak Prominence (CPP) [18] quantifies the strength of the dominant cepstral peak relative to the overall cepstrum, and was designed to capture the degree of harmonic organisation within a voice signal. CPP has been proposed as a robust measure of dysphonia in clinical applications [19], [20]. CPP has been found to distinguish between phonation types [20] and to vary with vowel height, with open vowels often showing higher CPP values [21], [22]. Shu et al. [11] reported that the close back vowel /u/ had the lowest CPP in Mandarin compared to other vowels.

Tones are primarily realised through modulations in f_0, which is a direct correlate of laryngeal activity (the rate of vibration of the vocal folds). Esling [23] and Kuang [24], [25] demonstrate that tone and phonation are linked: low tones are frequently associated with creaky voice, while high tones often exhibit tenser, more periodic phonation. Mandarin, a tone language, provides a valuable test case for examining such interactions. It has four contrastive tones: Tone 1 (high level), Tone 2 (rising), Tone 3 (low dipping), and Tone 4 (falling). They are

10.21437/SpeechProsody.2026-68

realised through f_0 modulations that involve varying degrees of laryngeal tension. Tone 3, in particular, is often produced with creaky phonation [24], [26]. Previous studies report systematic tone-related differences in vowel articulation in Mandarin and other tone languages, particularly for low-onset tones, which are found to condition a lowered tongue body for open vowels and a raised tongue for close front vowels [27], [28], [29], [30].

Previous research on Mandarin vowels and associated phonatory characteristics provides a useful foundation. In a clinically motivated study, Shu et al. [11] examined sustained Mandarin vowels produced with Tone 1 and found systematic differences in acoustic parameters related to voice quality in different vowels. Their results indicated that /a/ exhibited the highest jitter and shimmer, /u/ the highest HNR, and /a/ the lowest HNR and highest CPP, suggesting correlation between vowel articulation and phonatory characteristics. Their study was limited to one tone and their statistical analyses did not account for speaker variation or tone-vowel interactions.

This study explores whether differences in vowel quality are associated with differences in voice quality under different tone conditions. Specifically, it compares acoustic parameters associated with periodicity (f_0), phonatory stability (jitter, shimmer) and the amount of noise in the voice signal (HNR, CPP) across different vowel qualities under different tone conditions in Mandarin Chinese. It extends the study of Shu et al. [11] by incorporating multiple lexical tones and investigating how vowel quality and tone interact with phonatory characteristics. The same set of acoustic parameters is used here as they allow comparison with previous work [11] and provide a comprehensive profile of voice quality. Formant frequencies (F1, F2) are used to describe vowel qualities (vocal tract filter characteristics), while f_0 reflects pitch variation in lexical tone.

The study addresses the following questions:

1. Do Mandarin vowels of differing quality (height, backness, and lip rounding) exhibit systematically different intrinsic voice and pitch properties?
2. Do different tones co-occur with characteristic patterns in perturbation or spectral/cepstral measures?
3. How do vowel quality and tone interact in their effects on phonatory measures, and what does this reveal about the degree of interdependence between the voice source and the vocal tract filter?

The study aims to clarify the extent to which vowel articulation and tone jointly interact with voice quality in Mandarin. In doing so, it hopes to contribute to broader theoretical discussions of source-filter interaction and provide some empirical data relevant to both phonetic modelling and clinical voice analysis.

2. Methods

This study is loosely based on methodological framework in [11] and includes voice parameters commonly used for clinical voice assessment with adaptations to investigate vowel-tone interactions in connected speech rather than sustained vowels. Pilot testing conducted prior to the current study indicated that sustained vowels were inappropriate for tonal analysis, particularly for the low dipping tone (Tone 3), as the resulting pitch contours sounded unnatural to a native speaker consultant. Consequently, vowels were elicited in monosyllabic words in citation form and in carrier phrases.

The study was approved by the Research Ethics Committee of the authors' institution, with written informed consent obtained from all participants. Twelve native speakers of Mandarin Chinese (10 female, 2 male) participated. Participants were university students, aged 18-30, with no history of voice disorders or smoking.

2.1. Materials

The stimulus set consisted of monosyllabic Mandarin words, each with one of six target monophthongs (/i/, /y/, /u/, /ɤ/, /o/, /a/) realised with four lexical tones, 24 tokens in total.

The wordlist was designed with guidance from a native Mandarin consultant to ensure naturalness and phonotactic plausibility. A partial wordlist is shown in Table 1. The words are monosyllabic with open syllable (CV) structure to avoid coda effects. Onsets were primarily bilabial or velar stops (/p/, /b/, /k/), providing a consistent consonantal environment while minimising coarticulatory influence. Because of Mandarin phonotactic constraints, the front rounded vowel (/y/) was elicited in natural /l/ or /j/ contexts.

Table 1: *Partial wordlist used to elicit vowels.*

Target Vowel IPA (Pinyin)	Target Tone	Word	Translation
i (i)	1	pī / 批	critics
y (ü)	1	yǖ / 淤	bruise
u (u)	1	pū / 哭	to cry
ɤ (e)	1	kē / 科	discipline of study
o (o)	1	bō / 波	wave/storm
a (a)	1	bā / 巴	to wish

Words were elicited both in citation form and as embedded in a carrier phrase (1).

(1) 再把_说一遍		(Simplified Chinese)
Zài bǎ _ shuō yībiàn		(Pinyin)
Say _ again		(English Translation)

The elicitations were recorded in a semi-anechoic chamber using a Neumann TLM 103 microphone, in WAV format at a 44.1 kHz sampling rate, 16-bit quantisation. Each participant read the word list three times in randomised order, resulting in 1,728 tokens (12 participants x 24 words x 2 elicitations x 3 repetitions).

2.2. Data annotation and extraction

Recorded WAV files were annotated manually using Praat [31]. Each recording was manually segmented using TextGrids, with three annotation tiers marking the steady-state vowel portion, full vowel interval, and vowel midpoints for measuring formant frequencies. Vowel formant frequencies (F1, F2) and interval durations were extracted in Praat. Praat Voice Report with default voice analysis settings (raw cross-correlation) was used to extract f_0, jitter, shimmer, and HNR. Cepstral Peak Prominence smoothed (CPPS) was extracted using Praat's native method with the default settings. Mean and SD were calculated for these measures over the entire duration of each vowel in each tone condition.

To represent dynamic variation of f_0 and voice quality in vowels under different tones, contour data for f_0 and CPPS were extracted at 0.2 s intervals across the full vowel interval, normalised for duration using a custom Praat script. These contours were plotted using LOESS [32] smoothing in RStudio [33].

2.3. Statistical analyses

Statistical analyses were conducted in R [33] using the lme4 [34] and lmerTest [35] packages. Paired-sample t-tests indicated no significant differences between acoustic parameter values elicited in carrier phrase and citation form, so these datasets were combined for analysis. Linear mixed-effects models were fitted to each acoustic parameter, with vowel quality, lexical tone and their interactions as fixed effects and participant and word as random intercepts. Due to the limited number of male participants, analyses focused on female data (n = 10).

3. Results

Figure 1 shows normalized F1-F2 data for the six Mandarin monophthongs based. The clear and well-separated ellipses confirm expected distinctions in vowel height and backness. Vowel /ɤ/ overlaps slightly with the back vowels /u/ and /o/. Overall, vowel categories are distinct across speakers and contexts.

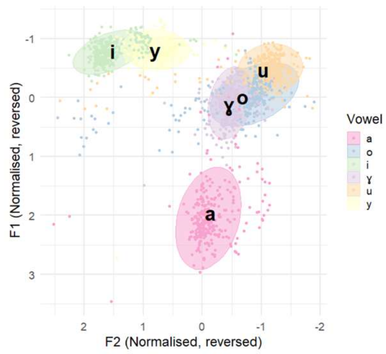

Figure 1. *Formant frequencies (Lobanov-normalised) of Mandarin Chinese vowels in this study.*

3.1. Effects of vowel quality

Figure 2 illustrates acoustic parameter values for each vowel under Tone 1. Mean values reflected vowel effect: f_0 ranged from 253 Hz (/a/) to 280 Hz (/u/); jitter and shimmer were lowest for the high/front vowels; and HNR was highest for /y/ (27.2 dB) and lowest for /a/ (19.4 dB). CPPS values were broadly similar across vowels, with lower value for /u/. Linear mixed-effects models were fitted for each acoustic parameter (f_0, jitter, shimmer, HNR, CPPS) with Vowel as the fixed effect and /a/ as the reference. Random intercepts were included for Participant and Repetition. For female speakers (n = 10), all vowels showed significantly higher f_0 than /a/ (β = 14.98-25.24, t = 4.05-6.83, $p < .001$). Jitter was significantly lower for /i/, /y/, and /ɤ/ (β = –0.09 to –0.13, t = –2.65 to –3.93, $p < .01$), while shimmer was also lower for /i/, /y/, /ɤ/, and /o/ (β = –0.80 to –1.87, t = –3.68 to –8.60, $p < .001$). All non-/a/ vowels had significantly higher HNR values (β = 5.42-7.82, t = 12.31-17.74, $p < .001$), indicating a clearer or more periodic voice quality relative to /a/. Only /u/ showed a significant reduction in CPPS (β = –1.80, t = –7.92, $p < .001$), suggesting greater spectral noise for this vowel.

These patterns broadly align with those reported by Shu et al. [11] where /a/ also showed the highest jitter and shimmer and the lowest HNR. Both studies found that close vowels (/i/, /y/, /u/) yield lower jitter and shimmer and higher HNR values than /a/. Our CPPS results also corroborate findings in [11], with /u/ showing lower CPPS compared with all other vowels.

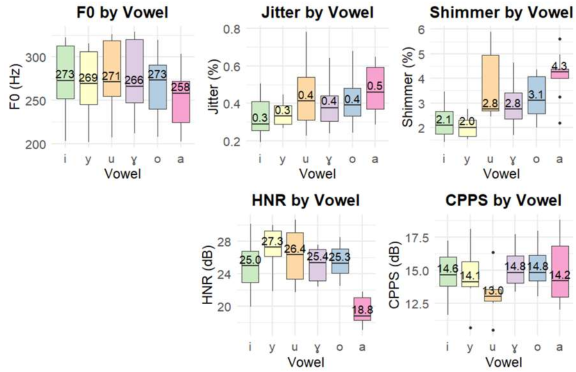

Figure 2. *Acoustic measures (median, IQR, min and max values) of female voice by Vowel (Tone 1).*

3.2. Effects of vowel quality and tone

Figure 3 shows parameter values for each vowel by tone. Linear mixed-effects models were fitted for each acoustic parameter (f_0, jitter, shimmer, HNR, CPPS) with Vowel, Tone, and their interaction as fixed effects, and random intercepts for Participant and Repetition.

For f_0, both vowel and tone had significant effects. High vowels (/i/, /y/, /u/, /ɤ/, /o/) showed significantly higher f_0 relative to /a/ across all tones (β = 21.44, 16.96, 25.24, 21.40, 14.98; *t* = 5.80, 4.59, 6.83, 5.79, 4.05; $p < .001$). Tones 2 and 3 showed significantly lower f_0 relative to Tone 1 (β = –40.5, –51.7; t = –7.6, –9.7; $p < .001$) across all vowels. Several vowel-tone interactions were significant: under Tone 3, for /ɤ/ and /o/ vowels f_0 lowering was smaller (β = 26.8, 17.0, $p < .05$) than under Tone 1, while /i/ and /u/ showed the opposite trend (β = –18.6, –21.2, $p < .05$). Under Tone 4, /i/ showed a greater f_0 reduction (β = –17.3, p = .022).

For jitter, tone had a strong main effect, with higher jitter values under Tones 2, 3 and 4 relative to Tone 1 (β = 0.81, 1.74, 2.79; *t* = 4.8, 10.3, 16.7; $p < .001$), although vowel-tone interactions were also observed. Thus, under Tone 3, /u/ and /o/ showed smaller increases in jitter than /a/ (β = –0.48, –0.63, $p < .05$), whereas /o/ showed higher increase in jitter under Tone 4 (β = 0.65, p = .005).

For shimmer, both vowel and tone were significant, significant interactions were also observed. Vowels /i/, /y/, and /ɤ/ showed lower shimmer than /a/ (β –1.59, –1.87, –1.02; *t* = –7.29, –8.60, –4.69; $p < .05$). Shimmer values were higher under Tones 2 and 3 (β = 2.6, 5.9, t = 6.1, 14.0, $p < .001$). /i/, /y/, /u/, and /o/ showed smaller shimmer increases under Tones 2 and 3 (β = −1.6, −2.0, −2.1, −2.6, $p < .01$) compared to Tone 1, whereas /ɤ/ and /o/ showed greater shimmer increases under Tone 4 (β = 2.7, 3.1, $p < .001$).

For HNR, all vowels (/i/, /y/, /u/, /ɤ/, /o/) had higher HNR than /a/ (β = 5.42, 7.82, 6.40, 5.55, 6.15; *t* = 12.31, 17.74, 14.52, 12.60, 13.95; $p < .001$). HNR values were lower under Tones 2, 3 and 4 relative to Tone 1 (β = −5.3, −8.2, −10.7; *t* = −10.7, −16.1, −21.8; $p < .001$). Under Tone 3, vowel /ɤ/ showed further reduction in HNR (β = −1.9, p = .007), while /i/ and /u/ partially offset it (β = 2.4, 1.4, $p < .05$).

For CPPS, only /u/ had significantly lower CPPS than /a/ across all tones (β = −1.7, t = −5.5, p < .001). Under Tone 3 (β = −4.7, t = −14.8, p < .001) and Tone 4 (β = −2.0, t = –6.2, p < .001), all vowels showed reduced CPPS values compared to Tone 1. In terms of significant vowel-tone interactions, /y/ and /u/ had higher CPPS under Tones 3 and 4 compared to /a/ under Tone 1(β = 1.0, 1.7; p < .05), while /o/ under Tone 4 had lower CPPS than /a/ under Tone 1 (β = –1.0, p = .020).

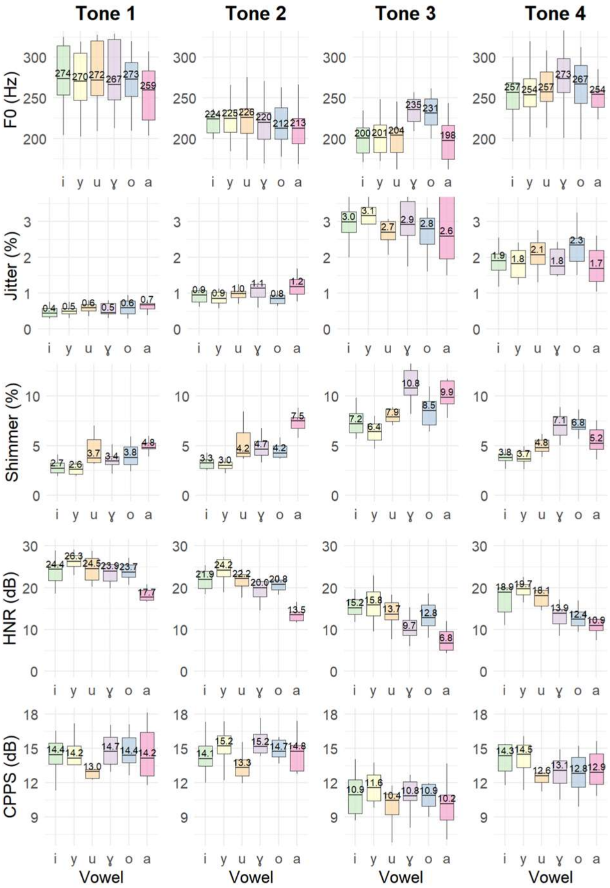

Figure 3. *Acoustic measures (median, IQR, min and max values) of female voice by vowel and tone.*

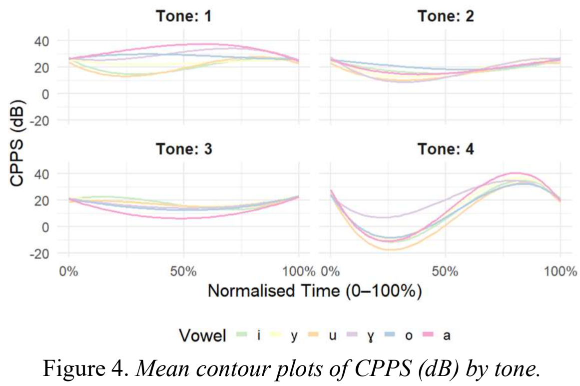

Figure 4. *Mean contour plots of CPPS (dB) by tone.*

The time-normalised contour plots for CPPS across vowel duration (Figure 4) show relatively stable spectral periodicity for Tones 1, 2 and 3. In contrast, for Tone 4, mean CPPS rises with the f_0 increase. Quantitative modelling of these trajectories could offer deeper insight into these dynamic effects. Note that qualitative inspection of male data (n = 2) showed similar trends, but the small sample limits generalisation.

4. Discussion

Although admittedly limited in scope, this study provides preliminary evidence that both vowel articulation and tone shape voice quality measures in Mandarin, and that these effects are not purely additive. The significant interactions of vowel and tone effects indicate that their influence on source behaviour must be interpreted jointly.

Vowel-related patterns broadly aligned with findings reported in earlier studies, with non-low vowels exhibiting higher f_0 than /a/. Jitter, shimmer, and HNR also varied systematically, with /a/ showing higher perturbation and lower HNR, in line with previous research on Mandarin vowels [11] and English vowels [17]. These results suggest that vowel articulation affects laryngeal behaviour rather than passively filtering it. It is possible that the configuration of the vocal tract during close vowel production creates more favourable biomechanical conditions for regular vocal fold oscillation. The wider pharyngeal opening and lowered larynx associated with open vowels may cause breathy or irregular phonation patterns [36, pp. 92–7]. Interestingly, /u/ shows a different acoustic profile. Despite being a close vowel, it exhibits significantly lower CPPS values across all tones, which suggests less periodic spectral structure. This may be attributable to articulatory features unique to /u/, such as lip rounding.

Tonal effects were also evident, although they differed with vowel context. Tone 1 showed comparatively stable phonation, while Tone 3 yielded the greatest perturbation and the lowest periodicity, consistent with previous descriptions of creaky or irregular voice associated with this lexical tone [24], [26]. Tones 2 and 4 showed intermediate values indicating moderate instability compared to Tone 1. In terms of vowel-tone interactions, tone-induced instability was attenuated in close vowels, while the reductions in periodicity associated with Tone 3 and, for some parameters, Tone 4 were more pronounced in close-mid, back vowels (/ɤ/, /o/). These findings indicate that articulatory configuration can modulate how tonal pitch movements are implemented at the laryngeal level. Articulatory studies find that Tone 3 involves systematic retraction and lowering gestures, suggesting a broader coordination of supralaryngeal and laryngeal adjustments [28].

This study provides some empirical support for models of speech production [37] that treat source and filter as dynamically interconnected rather than independent. The findings support intrinsically higher f_0 for close vowels [10], [38], [39] and creaky phonation under Tone 3 [24], [26] in female voice. However, the current dataset is small, and quantitative analyses were limited to female data. Future work with larger, balanced samples is needed to corroborate these findings and explore sex-related differences.

5. Conclusions

The findings suggest that vowel quality and lexical tone jointly shape phonatory characteristics in Mandarin, with tone-related phonatory instability varying with vowel quality. Open vowels tend to be less stable, and Tone 3 exerts a strong destabilising effect. Close or rounded vowels can buffer against tonal instability. Trends were consistent with established descriptions of intrinsic vowel effects and tonal phonation. The observed vowel-tone interactions support the need for integrated models of articulatory-laryngeal coordination. Future larger-scale studies should explore sex-related differences and cross-dialectal variation to assess the universality of these patterns.

6. References

[1] G. Fant, *Acoustic Theory of Speech Production : With Calculations Based on X-Ray Studies of Russian Articulations*. De Gruter Mouton, 1960.

[2] P. Lieberman, 'Some acoustic measures of the fundamental periodicity of normal and pathologic larynges', *J. Acoust. Soc. Am.*, vol. 35, no. 3, pp. 344–353, Mar. 1963, doi: 10.1121/1.1918465.

[3] K. N. Stevens, 'On the quantal nature of speech', *J. Phon.*, vol. 17, no. 1–2, pp. 3–45, Jan. 1989, doi: 10.1016/s0095-4470(19)31520-7.

[4] M. Garellek, 'Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality', *J. Phon.*, vol. 94, p. 101155, Sep. 2022, doi: 10.1016/j.wocn.2022.101155.

[5] P. Ladefoged and M. Lindau, 'Modeling articulatory–acoustics relations: a comment on Stevens' "On the quantal nature of speech"', *J. Phon.*, vol. 17, no. 1–2, pp. 99–106, Jan. 1989, doi: 10.1016/s0095-4470(19)31515-3.

[6] K. Johnson, P. Ladefoged, and M. Lindau, 'Individual differences in vowel production', *J. Acoust. Soc. Am.*, vol. 94, no. 2, pp. 701–714, Aug. 1993, doi: 10.1121/1.406887.

[7] I. R. Titze, 'Workshop on acoustic voice analysis: Summary statement', presented at the National Center for Voice and Speech, Denver, CO, 1995, pp. 1–36. [Online]. Available: https://ncvs.org/archive/freebooks/summary-statement.pdf

[8] E. Lin, J. Jiang, S. D. Noon, and D. G. Hanson, 'Effects of head extension and tongue protrusion on voice perturbation measures', *J. Voice*, vol. 14, no. 1, pp. 8–16, Mar. 2000, doi: 10.1016/s0892-1997(00)80090-9.

[9] W. G. Ewan and J. J. Ohala, 'Can intrinsic vowel *F* be explained by source/tract coupling?', *J. Acoust. Soc. Am.*, vol. 66, no. 2, pp. 358–362, Aug. 1979, doi: 10.1121/1.383669.

[10] D. H. Whalen and A. G. Levitt, 'The universality of intrinsic F0 of vowels', *J. Phon.*, vol. 23, no. 3, pp. 349–366, Jan. 1995, doi: 10.1016/s0095-4470(95)80165-0.

[11] M. Shu, Y. Zhang, and J. J. Jiang, 'The effect of Mandarin vowels on acoustic analysis: A prospective observational study', *J. Voice*, vol. 38, no. 6, pp. 1296–1301, Nov. 2024, doi: 10.1016/j.jvoice.2022.03.028.

[12] K. A. Wilcox and Y. Horii, 'Age and Changes in Vocal Jitter', *J. Gerontol.*, vol. 35, no. 2, pp. 194–198, Mar. 1980, doi: 10.1093/geronj/35.2.194.

[13] M. C. Franca, 'Acoustic comparison of vowel sounds among adult females', *J. Voice*, vol. 26, no. 5, p. 671.e9-671.e17, Sep. 2012, doi: 10.1016/j.jvoice.2011.11.010.

[14] S. E. Linville and E. W. Korabic, 'Fundamental frequency stability characteristics of elderly women's voices', *J. Acoust. Soc. Am.*, vol. 81, no. 4, pp. 1196–1199, Apr. 1987, doi: 10.1121/1.394642.

[15] R. F. Orlikoff, 'Vocal stability and vocal tract configuration: An acoustic and electroglottographic investigation', *J. Voice*, vol. 9, no. 2, pp. 173–181, Jun. 1995, doi: 10.1016/s0892-1997(05)80251-6.

[16] E. Yumoto, W. J. Gould, and T. Baer, 'Harmonics-to-noise ratio as an index of the degree of hoarseness', *J. Acoust. Soc. Am.*, vol. 71, no. 6, pp. 1544–1550, Jun. 1982, doi: 10.1121/1.387808.

[17] J. K. MacCallum, Y. Zhang, and J. J. Jiang, 'Vowel selection and its effects on perturbation and nonlinear dynamic measures', *Folia Phoniatr. Logop.*, vol. 63, no. 2, pp. 88–97, 2011, doi: 10.1159/000319786.

[18] R. R. Patel *et al.*, 'Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function', *Am. J. Speech Lang. Pathol.*, vol. 27, no. 3, pp. 887–905, Aug. 2018, doi: 10.1044/2018_AJSLP-17-0009.

[19] O. Murton, R. Hillman, and D. Mehta, 'Cepstral peak prominence values for clinical voice evaluation', *Am. J. Speech Lang. Pathol.*, vol. 29, no. 3, pp. 1596–1607, Aug. 2020, doi: 10.1044/2020_AJSLP-20-00001.

[20] C. M. Esposito, 'The effects of linguistic experience on the perception of phonation', *J. Phon.*, vol. 38, no. 2, pp. 306–316, Apr. 2010, doi: 10.1016/j.wocn.2010.02.002.

[21] S. N. Awan, A. Giovinco, and J. Owens, 'Effects of vocal intensity and vowel type on cepstral analysis of voice', *J. Voice*, vol. 26, no. 5, p. 670.e15-670.e20, Sep. 2012, doi: 10.1016/j.jvoice.2011.12.001.

[22] A. R. Meireles, 'Vowel quality and voice quality effects on spectral parameters: a pilot study', in *Proceedings of the 13th International Conference of Nordic Prosody*, Dec. 2023, pp. 222–231. doi: 10.2478/9788366675728-019.

[23] J. H. Esling, 'There are no back vowels: The laryngeal articulator model', *Can. J. Linguist. Rev. Can. Linguist.*, vol. 50, no. 1, pp. 13–44, 2005, doi: 10.1353/cjl.2007.0007.

[24] J. Kuang, 'Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice', *J. Acoust. Soc. Am.*, vol. 142, no. 3, pp. 1693–1706, Sep. 2017, doi: 10.1121/1.5003649.

[25] J. Kuang, 'The tonal space of contrastive five level tones', *Phonetica*, vol. 70, no. 1–2, pp. 1–23, Oct. 2013, doi: 10.1159/000353853.

[26] D. S. Davison, 'An acoustic study of so-called creaky voice in Tianjin Mandarin', *UCLA Work. Pap. Phon.*, no. 78, pp. 50–57, Feb. 1991.

[27] J. A. Shaw, W. Chen, M. I. Proctor, and D. Derrick, 'Influences of tone on vowel articulation in Mandarin Chinese', *J. Speech Lang. Hear. Res.*, vol. 59, no. 6, Dec. 2016, doi: 10.1044/2015_JSLHR-S-15-0031.

[28] P. Hoole and F. Hu, 'Tone-vowel interaction in Standard Chinese', *Int. Symp. Tonal Asp. Lang. Emphas. Tone Lang.*, pp. 28–31, Mar. 2004.

[29] F. Hu, 'Tonal effect on vowel articulation in a tone language', *Int. Symp. Tonal Asp. Lang. Emphas. Tone Lang.*, pp. 28–31, Mar. 2004.

[30] J. Kuang, 'Phonation contrast in two register contrast languages and its influence on vowel quality and tone', in *17th International Congress of Phonetic Sciences*, Hong Kong, China: ICPhS, Aug. 2011, pp. 1146–1149. [Online]. Available: https://www.sas.upenn.edu/~kuangj/Papers_files/Kuang_icphs2011.pdf

[31] P. Boersma and D. Weenik, *Praat: doing phonetics by computer*. (2025). [Online]. Available: https://praat.org

[32] H. Wickham, *ggplot2: elegant graphics for data analysis*, Second edition. in Use R! Cham: Springer international publishing, 2016.

[33] R Core Team, *R*. (Jun. 13, 2025). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org/

[34] D. Bates, M. Mächler, B. Bolker, and S. Walker, 'Fitting Linear Mixed-Effects Models Using **lme4**', *J. Stat. Softw.*, vol. 67, no. 1, 2015, doi: 10.18637/jss.v067.i01.

[35] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, 'lmerTest Package: Tests in Linear Mixed Effects Models', *J. Stat. Softw.*, vol. 82, no. 13, 2017, doi: 10.18637/jss.v082.i13.

[36] K. N. Stevens, *Acoustic phonetics*. MIT Press, 1998.

[37] J. H. Esling, S. R. Moisik, A. Benner, and L. Crevier-Buchman, *Voice Quality: The Laryngeal Articulator Model*, 1st edn. Cambridge University Press, 2019. doi: 10.1017/9781108696555.

[38] J. H. Esling, 'The laryngeal articulator's influence on voice quality and vowel quality', in *FATTORI SOCIALI E BIOLOGICI NELLA VARIAZIONE FONETICA*, vol. 3, C. Bertini, C. Celata, G. Lenoci, C. Meluzzi, and I. Ricci, Eds, in Studi AISV, vol. 3. , IT: Officinaventuno, 2017, pp. 13–26. [Online]. Available: https://doi.org/10.17469/O2103AISV000001

[39] K. Honda, T. Baer, H. Hirose, and M. Sawashima, 'Relationship between vowel articulation and pitch control', *J. Acoust. Soc. Am.*, vol. 69, no. S1, pp. S67–S67, May 1981, doi: 10.1121/1.386169.