

Attractiveness Judgments of F0-Manipulated Mandarin Voices Across Heterosexual and Homosexual Listener Groups

Chun He ^a, Yingjia Han ^a, Ziwenjun Chen ^b

a. Guangxi University, Nanning, Guangxi, China; b. Dali University, Dali, Yunnan, China.

Abstract: Fundamental frequency (F0) is a salient acoustic cue in voice perception, but how F0-related attractiveness judgments vary across listeners with different sexual orientations remains insufficiently understood. This study examined attractiveness ratings under controlled F0 manipulation in Mandarin voice stimuli produced by one male and one female announcer. Eighty adults were recruited: 20 heterosexual male, 20 heterosexual female, 20 homosexual male, and 20 homosexual female participants. An adaptive three-alternative forced-choice pretest with 12 listeners established a group-level F0 just-noticeable difference (JND) of 0.215 ERB. Seven F0 levels, from -3 to +3 JNDs around each speaker's own baseline, were then presented three times in the main experiment. Because each listener group rated only the target-voice sex that matched its presumed attraction pattern, the study compares two observed listener–target configurations rather than fully crossing listener sex, sexual orientation, and target-voice sex. Linear mixed-effects models showed that attractiveness ratings for the male voice decreased as F0 increased in both heterosexual female and homosexual male listeners, with no reliable difference in slopes. For the female voice, ratings increased with F0 in both heterosexual male and homosexual female listeners, while the homosexual-female group additionally showed a significant nonlinear component, with a fitted maximum near +2.4 JND. These findings indicate that, within these two listener–target configurations and these two controlled source voices, F0-related attractiveness judgments generally followed target-sex-related directions, whereas between-group variation was more evident in response strength and curve shape. The study contributes to experimental research on social voice perception and provides a controlled Mandarin-language baseline for future cross-linguistic comparisons of vocal social evaluation; it does not itself test intercultural variation.

Keywords: vocal attractiveness; fundamental frequency (F0); sexual orientation; social voice perception; Mandarin

1. Introduction

Human voices transmit linguistic information while also carrying socially meaningful cues about speakers. Listeners can extract information about sex, age, affect, body-related characteristics, and attractiveness from relatively limited vocal material, making voice perception an important interface between language and social cognition (Babel et al., 2014; Pisanski & Bryant, 2019). Vocal attractiveness is correspondingly multidimensional: acoustic properties contribute to judgments, but listener characteristics and the social context in which a voice is evaluated also matter (Babel et al., 2014).

Fundamental frequency (F0), the rate of vocal-fold vibration and a major acoustic correlate of perceived pitch, is especially relevant here (Titze, 1989). Puberty changes the larynx and contributes to substantial sex-related differences in adult F0; male voices generally have lower F0 than female voices (Evans et al., 2008; Fitch & Giedd, 1999). Experimental and observational studies have linked lower F0 in male voices to masculinity, dominance-related impressions, and attractiveness, though these links are sensitive to mating context and other social conditions (Puts, 2005; Puts et al., 2006; Puts et al., 2012; Hodges-Simeon et al., 2010). More broadly, vocal attractiveness seems to depend on a set of acoustic cues, not on pitch alone (Babel et al., 2014; Xu et al., 2013).

Evidence for female voices is less uniform. Higher F0 can increase perceived attractiveness in some contexts, but studies also report non-monotonic or context-dependent relationships between female F0 and social evaluation (Borkowska & Pawlowski, 2011; Pérez-Ramírez et al., 2020). These findings suggest that F0 is better seen as a socially interpreted acoustic cue, not as a universally linear index of attractiveness. This point matters for language-oriented research because listeners evaluate voices within

communicative and social expectations, and those expectations can vary across communities and cultural settings.

Listener characteristics are therefore central to the interpretation of F0–attractiveness relationships. Voice responses can differ according to listener sex, target-voice sex, and the social relevance of the target (Hodges-Simeon et al., 2010; Puts et al., 2012). Sexual orientation is especially relevant because it relates to the sex of potential romantic or sexual targets. Research on voice perception has shown that sexual orientation can shape behavioral and neural responses to male and female voices (Smith et al., 2019). Recent work has also directly compared heterosexual and homosexual listeners in vocal-attractiveness judgments, finding group differences in responses to vocal modulation but also substantial agreement on some unmodulated voices (Belyk et al., 2026). Three variables are thus relevant at once — listener sex, listener sexual orientation, and the sex of the target voice — and the present design does not treat them as independent factors. In the configurations studied here, listener sex and target-voice sex are aligned with sexual orientation: heterosexual women and homosexual men rated the male voice, whereas heterosexual men and homosexual women rated the female voice. This alignment is what makes the study informative about real listener groups, but it also means that any difference between two groups can reflect sexual orientation, listener sex, the sex of the voice being rated, or a combination of these. The groups are therefore described throughout as observed listener–target configurations, and group comparisons are interpreted accordingly.

Recent experiments have further shown the value of perceptually calibrated F0 manipulation for studying social voice perception: F0 shifts expressed in JND units can systematically affect attractiveness and other social judgments (He et al., 2026). That work, however, focused on heterosexual listeners and compared F0 with formant dispersion, so it remains unclear whether F0–attractiveness relationships vary systematically across listeners with different sexual orientations. The present study therefore focuses on F0 as a single controlled acoustic dimension. The design compares listener groups within target-voice categories, but it does not fully cross listener sex, sexual orientation, and target-voice sex. It is therefore not designed to estimate a unique causal effect of sexual orientation. Instead, it asks whether controlled F0 variation produces systematic and comparable attractiveness–response functions across the listener groups observed.

A further methodological issue is the distinction between an acoustic difference and its social interpretation. A change in F0 can be detected as a pitch difference without necessarily producing a corresponding change in attractiveness, and a change in attractiveness does not establish which social inference mediates the response. Work on voice attractiveness has emphasized that judgments reflect combinations of acoustic features and listener-specific expectations rather than one universal acoustic recipe (Babel et al., 2014; Pisanski & Bryant, 2019). Studies of experimentally shifted voices also show that listeners may respond to pitch modifications in ways that depend on the task and the social attribute being judged (Feinberg et al., 2005, 2008; Zheng et al., 2020). The present study therefore treats attractiveness as the measured outcome and does not equate it with perceived masculinity, femininity, dominance, or actual mate choice.

The study addresses three questions: (1) whether attractiveness ratings for male voices systematically change as F0 increases; (2) whether heterosexual female and homosexual male listeners show comparable F0-related patterns for male voices, and whether heterosexual male and homosexual female listeners show comparable patterns for female voices; and (3) whether the response functions differ in strength or shape, particularly for female voices. By combining JND-based acoustic manipulation with mixed-effects modeling, the study links experimental phonetic control with the social evaluation of spoken language. Although the sample is limited to Mandarin-speaking listeners, the design provides a controlled baseline against which future cross-linguistic and intercultural studies can examine whether F0-related social

evaluations are stable or culturally contingent. Two qualifications follow from the design. First, all claims concern these two controlled voice stimuli and these four listener groups, and the design supports statements about the shape of the F0-attractiveness function within each observed configuration rather than about the causal contribution of sexual orientation to voice evaluation. Second, because the listeners were Mandarin speakers, the intercultural dimension of the study is prospective: the data offer a baseline against which other speech communities can be compared, not evidence about cultural variation itself.

2. Methods

2.1 Participants

Eighty adults (mean age = 21.3 years, SD = 2.3; range = 18–30 years) participated in the main experiment. The sample comprised 20 heterosexual male, 20 heterosexual female, 20 homosexual male, and 20 homosexual female participants. All were native speakers of Standard Mandarin and reported no hearing or speech disorders. Participants were recruited through the university community and social-media platforms. Sexual orientation was assessed with a seven-point, Chinese-language adaptation of the Kinsey scale (Kinsey et al., 1948), on which 0 denotes exclusively heterosexual and 6 exclusively homosexual experience and attraction. Participants were classified as heterosexual if their score was 0 or 1 and as homosexual if it was 5 or 6, and eligibility additionally required agreement between the scale score and self-identification. Participants whose score fell in the intermediate range (2–4), or whose self-identification did not match their score, were not included in the analyzed sample; the present data therefore contain no bisexual or other intermediate-orientation group, and the heterosexual–homosexual contrast is a contrast between the two extreme groups only. All participants provided written informed consent and received compensation.

Sex and gender terminology followed the SAGER principles (Heidari et al., 2016). The present study recorded participants' self-reported sex and used male/female as the analytical categories; sexual orientation was treated as a separate participant characteristic. Because gender identity and gender roles were not independently measured, the analyses do not make claims about gender identity. Throughout the manuscript, male/female therefore refer to the study's sex-based participant classification and to the sex category of the constructed voice stimuli, whereas heterosexual/homosexual refer to sexual orientation.

2.2 Stimuli and F0 Manipulation

The baseline recordings were obtained from two professional Mandarin-language announcers (one male and one female), both certified at Grade I-B (一级乙等) in the Putonghua Proficiency Test. Both produced the Mandarin greeting “你好” (nǐ hǎo, “hello”). Recording took place in May 2026 in the sound-attenuated phonetics laboratory at Guangxi University. An AKG C544L microphone was connected through a Steinberg UR22C external audio interface to a Huawei MateBook D 14 laptop; recordings were captured with Praat version 6.4.46 (Boersma & Weenink, 2025). The distance between the lips and microphone was held at approximately 10 cm. Signals were recorded at 44,100 Hz with 16-bit quantization. After recording, the male and female tokens were standardized for duration by trimming or zero-padding and normalized in amplitude so that every stimulus was presented at approximately 70 dB SPL; a single playback level was used for all participants and all conditions. The normalized recordings of the two announcers served directly as the male- and female-voice baseline stimuli for subsequent F0 manipulation; no across-speaker averaging or composite-voice synthesis was performed (Bruckert et al., 2010).

Perceptual calibration was performed in a separate pretest involving 12 listeners with normal hearing, none of whom took part in the main experiment. The pretest used an adaptive three-alternative forced-choice (3AFC) discrimination procedure with the same Mandarin greeting to estimate the just-noticeable difference (JND) in F0. Pitch shifts were expressed on the equivalent rectangular bandwidth (ERB-rate) scale. The threshold was estimated at the group level rather than for each listener separately: the adaptive

runs of the 12 listeners were pooled, and the resulting value of 0.215 ERB was then applied to every participant in the main experiment as a fixed step size. For the main experiment, each baseline recording was manipulated in Praat to produce seven F0 conditions: -3, -2, -1, 0, +1, +2, and +3 JNDs. Relative to the unmodified baseline, these levels corresponded to ERB-rate displacements of -0.645, -0.430, -0.215, 0, +0.215, +0.430, and +0.645, respectively. Each F0 condition was repeated three times in the attractiveness task.

The F0 manipulation used an ERB-rate transformation, $\text{ERB-rate} = 21.4 \log_{10}(1 + 0.00437 \times \text{F0[Hz]})$; its inverse is $\text{F0[Hz]} = (10^{(\text{ERB-rate}/21.4)} - 1)/0.00437$ (Glasberg & Moore, 1990). For a condition at k JNDs, the target ERB-rate was calculated by adding $k \times 0.215$ to the baseline F0 expressed on this scale, followed by conversion to hertz. The shift was implemented by modifying the F0 contour of each baseline token and resynthesizing the signal with a PSOLA-based overlap-add method, which changes F0 while leaving the spectral envelope, and hence the formant pattern, essentially intact. Praat-based acoustic checks were then applied to every generated stimulus: the F0 of each token was re-measured and compared with its target on the ERB-rate scale, a tolerance of 0.01 ERB was applied, and any stimulus falling outside that tolerance was regenerated. The non-target acoustic dimension (formant dispersion) was likewise measured, so that unintended changes introduced by the single-parameter manipulation could be identified. Only F0-varied stimuli were included in the present attractiveness experiment; formant dispersion was not an experimental factor. The two baseline voices each yielded seven F0-manipulated conditions (14 distinct stimulus files in total, including the 0-JND condition for each voice category). Because the seven levels are defined relative to each speaker's own baseline, their absolute size depends on that baseline. For F0 values in the 100–250 Hz range spanned by adult male and female voices, 0.215 ERB corresponds to roughly 7–11 Hz, or about 0.8–1.1 semitones, so the full seven-level range (± 0.645 ERB) spans roughly 2.5–3.4 semitones.

2.3 Procedure and Design

The experiment used a mixed design with listener sex and sexual orientation as observed between-participant characteristics and F0 level as the within-participant variable. Heterosexual male and homosexual female participants rated female voices, whereas heterosexual female and homosexual male participants rated male voices. Target-voice sex was therefore not fully crossed with listener sex and sexual orientation.

Stimuli were presented through headphones in a sound-attenuated laboratory using E-Prime 3.0 (Psychology Software Tools, 2021). Each trial began with a 500-ms central fixation cross, followed by an approximately 800-ms recording of “你好” and a seven-point attractiveness question (1 = very unattractive, 7 = very attractive). Participants entered ratings using the number keys 1–7 without a response deadline, as specified in the source experiment protocol. Five practice trials preceded the experimental block. Each participant then completed 21 experimental trials (7 F0 levels $\times$ 3 repetitions). Trial order was randomized anew for each participant within the assigned target-voice category, the three repetitions of a given F0 level were interleaved with the other levels rather than presented as a block, and the 500-ms fixation cross that opened each trial served as the inter-trial interval. The design yielded 1,680 ratings across 80 participants. The three repetitions of each stimulus were additionally used to estimate the consistency of repeated ratings: single-rating agreement was moderate ($\text{ICC}(1,1) = .25$ for the male voice and $.47$ for the female voice), and the average of the three repetitions was more reliable ($\text{ICC}(1,3) = .50$ and $.73$, respectively).

2.4 Statistical Analysis

Male- and female-voice ratings were analyzed separately using linear mixed-effects models in R version 4.5.1 (R Core Team, 2025), with lme4 2.0.1 (Bates et al., 2015), lmerTest 3.2.1 (Kuznetsova et al., 2017), and emmeans 2.0.2 (Lenth, 2025). F0 level was entered as a continuous predictor centered at 0 JND,

so that the intercept represents the fitted rating at the unmanipulated baseline, and listener group was entered as a between-participant factor within each target-voice category, dummy-coded with the heterosexual group as the reference. The $F0 \times \text{listener-group}$ coefficient therefore indexes the difference in slope between the heterosexual and the homosexual group of that target-voice set. Participant identity entered the model as a random intercept, allowing each listener's overall use of the rating scale to vary, and models were fitted by restricted maximum likelihood. Because the design is balanced (three ratings per participant and $F0$ level), the fixed-effect estimates coincide with those of an equivalent within-participant analysis, and the random intercept accounts for the dependence among the 21 ratings contributed by each listener. The primary model was:

$$\text{Attractiveness} \sim F0 \times \text{sexual orientation} + (1 \mid \text{participant})$$

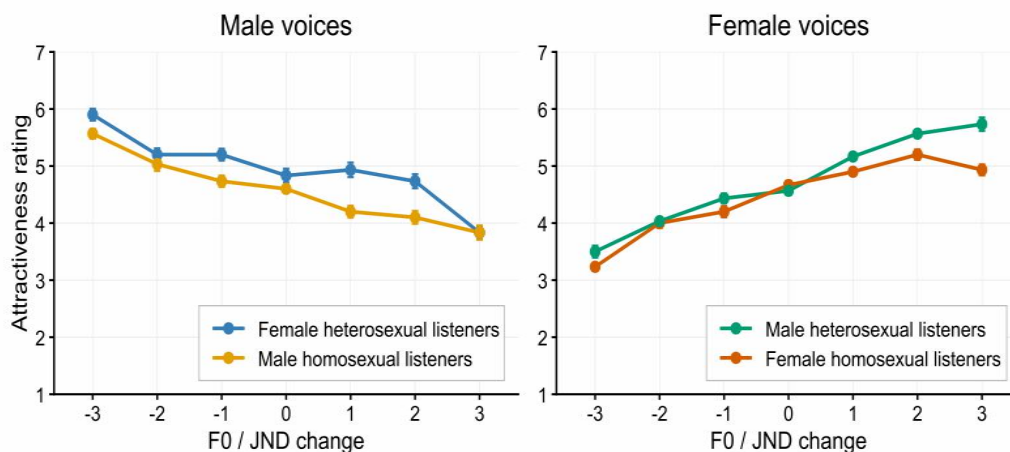
For female voices, $F0^2$ and its interaction with listener group were additionally examined. Inspection of the cell means showed that ratings in the homosexual-female group rose up to +2 JND and then fell at +3 JND, a departure from linearity that a linear model cannot capture; the quadratic term was therefore added to test whether this curvature differed between listener groups, and it was retained only for female voices, where the descriptive downturn was present. Group-specific linear and quadratic coefficients were obtained as simple effects with emmeans and are reported with Bonferroni adjustment where multiple simple effects are compared. Wald 95% confidence intervals are reported for all fixed effects, and denominator degrees of freedom were approximated with Satterthwaite's method; because each model contained 840 ratings from 40 participants and at most six fixed effects, the $F0$ -related terms had at least 790 denominator degrees of freedom, so their t statistics can be read against the normal approximation. The estimated vertex of the quadratic function was interpreted as a fitted response peak rather than a directly tested stimulus condition. Inferences about listener-group differences were restricted to the observed listener–target combinations, because listener sex, sexual orientation, and target-voice sex were not fully crossed.

3. Results

3.1 Descriptive patterns

The 1,680 rating observations showed clear $F0$ -related patterns. For male voices, attractiveness ratings declined across the continuum in both groups: heterosexual female listeners decreased from 5.90 at –3 JND to 3.83 at +3 JND, while homosexual male listeners decreased from 5.53 to 3.80. For female voices, heterosexual male listeners increased from 3.50 at –3 JND to 5.73 at +3 JND, whereas homosexual female listeners increased from 3.23 at –3 JND to 5.20 at +2 JND and then declined slightly to 4.93 at +3 JND. Each cell mean is based on 60 ratings (20 participants $\times$ 3 repetitions).

Figure 1. Attractiveness ratings across $F0$ levels. Error bars represent ± 1 SE.



3.2 Male voices

Both heterosexual female and homosexual male listeners showed significant negative F0 slopes. The mixed-effects model yielded a significant overall F0 effect ($\beta = -0.264$, $SE = 0.021$, $t = -12.66$, $p < .001$, 95% CI $[-0.305, -0.223]$), whereas the $F0 \times$ sexual-orientation interaction was not significant ($\beta = -0.007$, $SE = 0.027$, $t = -0.27$, $p = .792$, 95% CI $[-0.060, 0.046]$). The two observed groups therefore showed closely aligned downward response functions. This result supports relative consistency in the present sample but should not be interpreted as evidence that sexual orientation has no effect. The model-implied rating at the unmanipulated baseline (0 JND) was 4.95. Homosexual male listeners also gave somewhat lower ratings overall than heterosexual female listeners, so the two fitted functions were closely parallel rather than identical.

3.3 Female voices

Both heterosexual male and homosexual female listeners generally gave higher ratings to female voices as F0 increased. The model showed a significant positive F0 effect ($\beta = 0.375$, $SE = 0.024$, $t = 15.72$, $p < .001$, 95% CI $[0.328, 0.422]$), together with a significant $F0 \times$ sexual-orientation interaction ($\beta = -0.082$, $SE = 0.027$, $t = -3.05$, $p = .004$, 95% CI $[-0.135, -0.029]$). The quadratic interaction was also significant ($\beta = -0.050$, $SE = 0.016$, $t = -3.16$, $p = .003$, 95% CI $[-0.081, -0.019]$). In the homosexual-female group, the linear component was positive ($\beta = 0.293$, $p < .001$) and the quadratic component was negative ($\beta = -0.061$, $p < .001$), yielding a fitted maximum near +2.4 JND. This value is the vertex of the fitted function rather than a directly sampled F0 level. The main effect of $F0^2$ in the same model was small and did not reach significance ($\beta = -0.011$), indicating that the fitted function of the heterosexual-male group was essentially linear over this range, and the group difference therefore lies in the curvature rather than in the overall slope alone.

Table 1. Mixed-effects model results for male and female voices

Target voice / effect	β	SE	95% CI	t	p
Male: F0	-0.264	0.021	$[-0.305, -0.223]$	-12.66	< .001
Male: $F0 \times$ orientation	-0.007	0.027	$[-0.060, 0.046]$	-0.27	.792
Female: F0	0.375	0.024	$[0.328, 0.422]$	15.72	< .001
Female: $F0 \times$ orientation	-0.082	0.027	$[-0.135, -0.029]$	-3.05	.004
Female: $F0^2 \times$ orientation	-0.050	0.016	$[-0.081, -0.019]$	-3.16	.003

Note. Models were fitted separately for male and female voices: Attractiveness $\sim F0 \times$ listener group + (1 | participant) for male voices, with $F0^2$ and its interaction with listener group added for female voices. F0 was centered at 0 JND and listener group was dummy-coded with the heterosexual group as the reference. N = 840 ratings from 40 participants per target voice; denominator degrees of freedom (Satterthwaite) were at least 790 for all F0-related terms. Both models also contained an intercept (4.95 for male voices and 4.76 for female voices) and, for female voices, a main effect of $F0^2$ ($\beta = -0.011$, ns); the intercept is the fitted rating at 0 JND, and the $F0^2$ main effect describes the curvature of the heterosexual-male group.

4. Discussion

The principal finding is a systematic relationship between F0 and vocal-attractiveness ratings across the observed listener–target configurations. Lower F0 was associated with higher attractiveness ratings for male voices, whereas higher F0 was associated with higher ratings for female voices. These directions are broadly consistent with previous work linking male F0 to masculinity and attractiveness and female F0 to social and sexual evaluation (Puts, 2005; Puts et al., 2006; Puts et al., 2012; Feinberg et al., 2005; Xu et al., 2013; Pérez-Ramírez et al., 2020). The present experiment extends this literature by expressing the manipulation in perceptually calibrated JND units and by examining heterosexual and homosexual listeners under the same controlled F0 continuum. These claims are correspondingly specific: the patterns describe responses to the two controlled voice stimuli used here, within the four observed listener groups.

For male voices, heterosexual female and homosexual male listeners showed highly similar

downward response functions, and the $F0 \times$ orientation interaction was not significant. The appropriate interpretation is relative consistency in the current sample rather than equivalence between orientations. Because listener sex and sexual orientation were not independently crossed, the design cannot determine whether the similarity reflects sexual orientation, listener sex, target relevance, or their combination. This limitation is important when comparing the present results with studies showing that sexual orientation can influence basic responses to voices (Smith et al., 2019). Homosexual male listeners also gave lower ratings to the male voice overall (cell means were on average about 0.4 points lower than those of heterosexual female listeners), so the two functions were closely parallel rather than identical. Because these two groups differ in both listener sex and sexual orientation, however, neither the parallel slopes nor the level difference can be assigned to sexual orientation alone.

For female voices, the overall positive $F0$ effect was accompanied by significant linear and quadratic orientation interactions. The homosexual-female group showed a rise in attractiveness ratings up to the upper-middle range of the continuum, followed by a decline at the highest sampled level; the fitted maximum near +2.4 JND should therefore be understood as an estimated response optimum rather than evidence for an inherently optimal $F0$. Because the study did not measure perceived femininity, age, maturity, or other social attributes, the psychological basis of the downturn cannot be determined. The finding is better interpreted as evidence of response-shape variation than as proof of a specific mechanism.

Taken together, the data suggest that $F0$ can function as a socially meaningful acoustic cue in attractiveness judgments while allowing both shared and listener-specific response patterns. This is compatible with the broader view that vocal attractiveness depends on multiple acoustic and social dimensions rather than on a single universal cue (Babel et al., 2014). The fact that the direction of the $F0$ effect generally corresponded to the sex category of the target voice is also consistent with research showing that listeners use voice to process socially relevant information. At the same time, the present findings do not establish that listeners consciously interpret $F0$ as a marker of masculinity or femininity, because those perceptual dimensions were not directly measured. A further constraint is that each voice category was represented by a single professional announcer, so the functions reported here describe responses to these two controlled stimuli rather than to male and female voices in general; speaker-specific properties such as voice quality, speaking style, or the particular vowel context of the token may interact with $F0$.

From the perspective of language and intercultural studies, the findings have a more limited but potentially useful implication. The stimuli were Mandarin lexical items and the listeners were Mandarin speakers, so this is not a cross-cultural comparison. Instead, it offers a controlled Mandarin-language baseline for examining how a phonetic property becomes socially meaningful in voice evaluation. Future research can test whether the same JND-scaled $F0$ functions appear across languages and speech communities, and whether language-specific phonological systems, cultural norms, or familiarity with particular voice styles change how pitch is interpreted. Such comparisons would help distinguish relatively general properties of auditory-social perception from culturally situated voice meanings.

A few limitations should shape how the results are read. First, listener sex, sexual orientation, and target-voice sex were not fully crossed, because each group rated only the target-voice sex that matched its presumed attraction pattern. Second, each observed group had 20 participants; repeated ratings increase measurement precision, but they are not a substitute for independent participants. Third, using one source speaker per voice category limited the stimulus sampling. It improved acoustic control, but it also limits generalization to individual speakers. Acoustic measures such as formant dispersion need careful interpretation, since extraction settings and vowel context can affect estimates (Anikin et al., 2024). Fourth, the experiment manipulated $F0$ alone and used a single lexical item, so the results identify controlled $F0$ –attractiveness associations rather than a unique biological or social mechanism. Finally, the study does not

include a cross-linguistic or intercultural comparison, which future work should address.

5. Conclusion

This study shows that controlled F0 variation can systematically shape vocal-attractiveness ratings in Mandarin-speaking listeners with different sexual orientations. For the male voice, lower F0 was associated with higher ratings in both the heterosexual female and the homosexual male group; for the female voice, higher F0 was associated with higher ratings in both the heterosexual male and the homosexual female group. The two groups evaluating the male voice showed closely aligned response functions, whereas the female-voice groups differed more clearly in response shape, including a nonlinear pattern among homosexual female listeners. Because listener sex, sexual orientation, and target-voice sex were aligned rather than crossed, and because each voice category was represented by a single source speaker, these findings describe these two controlled voice stimuli and the four observed listener groups; they do not establish a causal effect of sexual orientation or a general preference for male and female voices. They support an account in which F0 contributes to socially meaningful voice evaluation without implying a single universal preference. More broadly, the study illustrates how experimental phonetic manipulation can connect acoustic structure with social interpretation and provides a controlled Mandarin-language baseline for future cross-linguistic and intercultural research on the social meaning of voice.

References

- Anikin, A., Barreda, S., & Reby, D. (2024). A practical guide to calculating vocal tract length and scale-invariant formant patterns. *Behavior Research Methods*, 56, 5588–5604.
<https://doi.org/10.3758/s13428-023-02288-x>
- Babel, M., McGuire, G., & King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PLOS ONE*, 9(2), e88616. <https://doi.org/10.1371/journal.pone.0088616>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Belyk, M., Horak, L., Beone, C., & McGettigan, C. (2026). Vocal attractiveness in homosexual and heterosexual listeners. *Royal Society Open Science*, 13(2), 1–16. <https://doi.org/10.1098/rsos.251815>
- Boersma, P., & Weenink, D. (2025). *Praat: Doing phonetics by computer* (Version 6.4.46) [Computer software]. <https://www.praat.org/>
- Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1), 55–59.
<https://doi.org/10.1016/j.anbehav.2011.03.024>
- Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., Kawahara, H., & Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology*, 20(2), 116–120.
<https://doi.org/10.1016/j.cub.2009.11.034>
- Evans, S., Neave, N., Wakelin, D., & Hamilton, C. (2008). The relationship between testosterone and vocal frequencies in human males. *Physiology & Behavior*, 93(4–5), 783–788.
<https://doi.org/10.1016/j.physbeh.2007.11.033>
- Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M., & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, 69(3), 561–568. <https://doi.org/10.1016/j.anbehav.2004.06.012>
- Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, 37(4), 615–623.
<https://doi.org/10.1068/p5514>
- Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 106(3), 1511–1522.

<https://doi.org/10.1121/1.427148>

- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47(1–2), 103–138. [https://doi.org/10.1016/0378-5955\(90\)90170-T](https://doi.org/10.1016/0378-5955(90)90170-T)
- He, C., Tang, Q., & Xian, W. (2026). Acoustic and perceptual evidence relevant to sexual selection in human voice dimorphism. *Frontiers in Psychology*, 17, 1938866. <https://doi.org/10.3389/fpsyg.2026.1938866>
- Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and Gender Equity in Research: Rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review*, 1, 2. <https://doi.org/10.1186/s41073-016-0007-6>
- Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, 21(4), 406–427. <https://doi.org/10.1007/s12110-010-9101-5>
- Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Saunders.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lenth, R. V. (2025). *emmeans: Estimated marginal means, aka least-squares means* (Version 2.0.2) [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- Pérez-Ramírez, M., Reby, D., & Puts, D. A. (2020). Fundamental frequency of the female’s voice: A cross-country empirical study on its influence on social and sexual selection. *Personality and Individual Differences*, 160, 109937. <https://doi.org/10.1016/j.paid.2020.109937>
- Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. In N. S. Eidsheim & K. Meizel (Eds.), *The Oxford handbook of voice studies* (pp. 269–300). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199982295.013.29>
- Psychology Software Tools. (2021). *E-Prime* (Version 3.0) [Computer software]. <https://pstnet.com/products/e-prime/>
- Puts, D. A. (2005). Mating context and menstrual phase affect women’s preferences for male voice pitch. *Evolution and Human Behavior*, 26(5), 388–397. <https://doi.org/10.1016/j.evolhumbehav.2005.03.001>
- Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, 27(4), 283–296. <https://doi.org/10.1016/j.evolhumbehav.2005.11.003>
- Puts, D. A., Jones, B. C., & DeBruine, L. M. (2012). Sexual selection on human faces and voices. *Journal of Sex Research*, 49(2–3), 227–243. <https://doi.org/10.1080/00224499.2012.658924>
- R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Smith, E., Junger, J., Pauly, K., Kellermann, T., Dernt, B., & Habel, U. (2019). Cerebral and behavioural response to human voices is mediated by sex and sexual orientation. *Behavioural Brain Research*, 356, 89–97. <https://doi.org/10.1016/j.bbr.2018.07.029>
- Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, 85(4), 1699–1707. <https://doi.org/10.1121/1.397959>
- Xu, Y., Lee, A., Wu, W.-L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PLOS ONE*, 8(4), e62397. <https://doi.org/10.1371/journal.pone.0062397>
- Zheng, Y., Compton, B. J., Heyman, G. D., & Jiang, Z. (2020). Vocal attractiveness and voluntarily pitch-shifted voices. *Evolution and Human Behavior*, 41(2), 170–175. <https://doi.org/10.1016/j.evolhumbehav.2020.01.002>

Funding Statement

The authors declared that financial support was not received for this work and/or its publication.

Conflict of Interest Statement

The authors declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contribution Statement

He Chun was responsible for experimental design, conducting the experiments, and methodology. Han Yingjia was responsible for conducting the experiments, statistical analysis, and manuscript formatting. Chen Ziwenjun was responsible for statistical analysis, data visualization, manuscript translation, and language polishing. All three authors contributed to manuscript writing. All authors reviewed and approved the final manuscript and assume responsibility for their respective contributions.

Data Availability Statement

The study data comprise attractiveness ratings from adult Mandarin-speaking listeners, listener-group characteristics, and F0 stimulus-condition identifiers. De-identified trial-level data, R scripts for data processing and statistical analysis, and the manipulated audio stimuli are available from the corresponding author upon reasonable request. The original voice recordings contain identifiable voice information and are not publicly available.

Ethical Approval and Informed Consent Statement

Written informed consent was obtained from all participants prior to the experiment, and participation was voluntary. The study protocol was approved by the Institutional Review Board of Guangxi University (Approval No. 2026–060).

Author Bio

Chun He (b. 2001), Master's degree, School of Liberal Arts, Guangxi University; research interests: evolutionary psychology and experimental phonetics; E-mail: ryswtte@163.com.

Yingjia Han (b. 2002), Master's degree, School of Liberal Arts, Guangxi University; research interests: experimental phonetics and dialectology; E-mail: 13036879718@163.com.

Ziwenjun Chen (b. 2002), Master's degree, School of Foreign Languages, Dali University; research interests: phonology and translation studies; E-mail: obcheese@163.com.