

# Acoustic and Perceptual Characteristics of Mandarin Speech in Homosexual and Heterosexual Male Speakers

Puyang Geng<sup>1</sup>, Wentao Gu<sup>1</sup> and Hiroya Fujisaki<sup>2</sup>

<sup>1</sup>School of Chinese Language and Culture, Nanjing Normal University, Nanjing, China <sup>2</sup>The University of Tokyo, Tokyo, Japan gengpuyang6@163.com, wtgu@njnu.edu.cn, fujisaki@alum.mit.edu

## Abstract

The present study investigated both acoustic and perceptual characteristics of Mandarin speech in homosexual and heterosexual male speakers. Acoustic analyses of monosyllabic words showed significant differences between the two groups in F0 features (including the mean, the max, and the range), F1 and F2 of vowels, aspiration/frication duration of consonants, and center of gravity as well as skewness for /s/. Especially, the patterns were found to be opposite between Mandarin and American English speakers, which might be due to social psychological differences between the two societies. The perceptual experiment showed that the perceived score of gayness differed significantly between the speeches of the two groups. Among those acoustic parameters showing significant differences, fricative duration may be the most salient cue for sexual orientation of Mandarin male speakers.

**Index Terms**: homosexual male, heterosexual male, Mandarin, acoustic analysis, speech perception

# 1. Introduction

Speech conveys not only linguistic information, but also speaker's information such as age, gender, health condition, and social status. There have been a few studies on the effects of sexual orientation on speech production of American English. In comparison with heterosexual males, homosexual males produced /s/ with longer duration, higher peak frequency, and more negative spectral skewness [1, 2]; much larger VOT for consonants [3]; larger acoustic vowel space [3, 4], especially with higher F1 for  $\langle \epsilon \rangle$  and  $\langle a \rangle$  [2]. However, no significant difference was found in fundamental frequency (F0) parameters between homosexual and heterosexual males, neither in isolated words [2] nor in continuous speech [5].

Previous perceptual studies showed that listeners were able to identify male speakers' sexual orientation from their speech. Some reported the identification from continuous speech [1, 3, 5], and others claimed that reliable judgement could be made even from isolated word stimuli [2, 6], depending highly on such segments as /s/, /a/, /c/, /ou/, and /ei/ [2, 7].

Similar studies have been conducted on Cantonese [16] and Spanish speakers [18], but there is no study on Mandarin yet. Therefore, the present study aimed to test Mandarin speakers, and to make a crosslinguistic and crosscultural comparison to find whether the homosexual male speakers of a tone language in the oriental culture will show any specific characteristics.

We conducted both acoustic and perceptual experiments. The acoustic experiment investigated whether homo- and heterosexual males differed in acoustic parameters of monosyllabic words. The perceptual experiment investigated how well the sexual orientation of male speakers could be identified from their monosyllabic words. Moreover, the relationship between acoustic and perceptual results was examined to find the acoustic cues of sexual orientation.

# 2. Acoustic analysis

## 2.1. Speech Data

Twenty-four participants (12 homosexual males and 12 heterosexual males) took part in the acoustic experiment, and all homosexual male speakers were recruited from a semi-open volunteer service center in which all members were self and mutual identified as homosexual. To minimize dialectal differences, all participants were native speakers of Northern Mandarin and spoke Standard Mandarin very fluently. None of them had a reported history of speech or hearing disorders. Each participant was reasonably paid for the experiment.

The mean age was 23.8 years (SD=3) for the homosexual group, and 25.8 years (SD=1.9) for the heterosexual group, with no significant difference between the two groups. The mean height was 177.7 cm (SD=5.1) for the homosexual group, and 175.6 cm (SD=6.6) for the heterosexual group, with no significant difference between the two groups. Therefore, the two groups were well matched in age and height. The match on height is to minimize the potential F0 differences caused by body size difference.

The Mandarin text materials consist of 64 monosyllabic words which are composed of nine consonants (/p/, /p<sup>h</sup>/, /m/, /s/, / $(\xi$ /, /ts<sup>h</sup>/, /ts<sup>h</sup>/, /ts<sup>h</sup>/, /tc<sup>h</sup>/, /l/), three vowels (/a/, /i/, /u/), and four tones (T1, T2, T3, T4). Pseudo words are excluded.

After the participants got familiar with the text materials, speech recording was conducted in a sound-proof booth, using a portable digital recorder Zoom H4n, digitized at 44.1 kHz with 16-bit precision. Each participant read aloud the materials with his normal speaking rate, pitch and loudness.

Among all 64 monosyllabic words, 32 words composed of three consonants (/p/, /p<sup>h</sup>/, /m/), three vowels (/a/, /i/, /u/) and four tones were selected to examine the F0 and formants of vowels, and 41 words composed of six consonants (fricatives /s/, / $\xi$ /; aspirated affricates /ts<sup>h</sup>/, /tg<sup>h</sup>/, /te<sup>h</sup>/; and aspirated stop /p<sup>h</sup>/), three vowels (/a/, /u/, /i/, excluding the apical allophones of /i/) and four tones were selected to examine the durations of consonants. Acoustic analyses were conducted using Praat [8].

#### 2.2. Results

### 2.2.1. Fundamental frequency

Figure 1 shows the F0 contours (measured in semitone with reference to 100 Hz) on the vowels in all four tones. Table 1



Figure 1:  $F_0$  contours of the syllables in four tones.

Table 1: Comparison of F0 parameters (St).

	mean F0	F0 max	F0 min	F0 range
Homosexual	3.94*	8.65*	-1.27	9.92*
Heterosexual	$5.96^{*}$	12.01*	-0.77	$12.78^{*}$

shows the results of Mann-Whitney U tests between the two groups on four F0 parameters including the mean, the max, the min, and the range of F0. Hereinafter, unless specified, the asterisks in tables stand for significant differences between the two groups (p < 0.05). The homosexual group shows lower F0 (especially lower at high pitch) in all four tones.

To further compare the F0 contours, we conducted growth curve analysis [9], an efficient method to analyze nonlinear contours based on multi-level regression techniques. Since F0s of Mandarin tones are no more complex than convex/concave curves, we adopted quadratic polynomials to approximate the F0 contours, i.e.,  $F_0(t) = a+bt+ct^2$ . Here, *t* stands for normalized time, and *a*, *b*, and *c* indicate the mean, the slope, and the U-shape curvature of the curve. Then, we conducted linear mixed-effects model using the R package *lme4* [10, 15] to find the differences in *a*, *|b|* (absolute slope), and *c* between the two groups. Sexual orientation and tone type were fixed effects, while subjects and word items were random effects.

The results of growth curve analysis shows that the main effect of sexual orientation is significant on *a* [F(1, 24.0) = 6.00, p < 0.05], lower for the homosexual group; significant on *b* [F(1, 24.0) = 3.30, p < 0.05], lower for the homosexual group – hence a narrower F0 range; but not significant on *c* [F(1, 24.1) = 0.29, p = 0.59]. The main effects of tone type on *a*, |*b*|, and *c* are all significant [F(3, 31.1) = 251.59, p < 0.001; F(3, 31.9) = 1003.72, p < 0.001; F(3, 53.0) = 195.55, p < 0.001, respectively]. Moreover, there are significant interactions between sexual orientation and tone type on *a*, |*b*|, and *c* [F(3, 7525.9) = 102.06, p < 0.001; F(3, 7525.2) = 148.96, p < 0.001; F(3, 7526.0) = 10.86, p < 0.001, respectively]. Simple effect analysis using the R package *lsmeans* [11] shows that the homosexual group has significantly lower *a* (for T1, T2, T4), |*b*| (for T2, T3, T4) and *c* (for T3, T4) than the heterosexual group (p < 0.05). These

results coincide with the patterns shown in Fig. 1. The difference in *a* is associated with the average pitch, while the differences in absolute slope |b| and curvature *c* are meaningful only for contour tones T2/T3/T4.

#### 2.2.2. Acoustic vowel space

Table 2 shows the mean values of F1 and F2 for three nominal vowels. There are significant differences in F1 of /a/ and /i/, and in F2 of /i/ and /u/. Especially, the homosexual group shows a much lower F1 of /a/ than the heterosexual group, indicating that homosexual males open their mouths to a lesser degree than heterosexual males in producing low vowels.

Table 2:	Formants	of three	nominal	vowels	(Hz).
					· · · ·

	Group	/a/	/i/	/u/	
E1	Heterosexual	838.9*	$296.6^{*}$	390.1	
ΓI	Homosexual	690.1*	$279.7^{*}$	380.0	
F2	Heterosexual	1272.3	$2279.1^{*}$	$709.9^{*}$	
	Homosexual	1260.7	$2191.9^{*}$	$667.3^{*}$	



Figure 2: Acoustic vowel space.

Figure 2 shows the mean acoustic vowel space constituted by three nominal vowels, with both formants scaled in the base-10 logarithmic domain. The homosexual group has a smaller space than the heterosexual group. Based on the equation below [12], the triangular vowel space area (tVSA) was calculated:

$$VSA = ABS\{[(F1[i] * (F2[a] - F2[u]) + F1[a] * (F2[u] - F2[i]) + F1[u] * (F2[i] - F2[a])]/2\}.$$
 (1)

The mean tVSA was 0.099  $(\log Hz)^2$  for the homosexual group and 0.085  $(\log Hz)^2$  for the heterosexual group. Mann-Whitney U test showed that tVSA was significantly smaller in the homosexual group than in the heterosexual group (p < 0.05).

#### 2.2.3. Durations of vowels

ť

We used the R package *lme4* [10] to conduct a linear mixedeffects model on the durations of three nominal vowels, taking sexual orientation, consonant (/p/, /p<sup>h</sup>/, /m/), and tone type as fixed effects, and subject as random effect.

No main effect of sexual orientation was found for any vowel, but there were interaction effects. For /a/, significant interaction was found between sexual orientation and tone type [F(3, 24.1) = 4.48, p < 0.05], and simple effect analysis showed that the difference between T2 and T3 was significant in the heterosexual group (T3 was longer, p < 0.05), but was not significant in the homosexual group. For /i/ and /u/, interactions between sexual orientation and consonant were significant [F(2, 28.1) = 3.45, p < 0.05; F(2, 27.7) = 4.04, p < 0.05, respectively].

Especially, the duration of /u/ was significantly longer after /p/ than after /p<sup>h</sup>/ in the homosexual group, but no significant difference was found in the heterosexual group.

#### 2.2.4. Durations of consonants

Linville [1] found that in English speech, /s/ in homosexual males was much longer than in heterosexual males, so the duration of /s/ could be an important cue of sexual orientation. We aim to investigate whether similar results exist in Mandarin speech and also apply to other consonants with aspiration/ frication. Therefore, the durations of /s/, /g/, /ts<sup>h</sup>/, /tg<sup>h</sup>/, /tc<sup>h</sup>/, /p<sup>h</sup>/ were measured (for affricates/stops, the closure portions could not be detected in isolated syllables and thus were excluded). Figure 3 plots the mean durations of six consonants. Mann-Whitney U test shows that all six consonants are significantly longer in the homosexual group (p < 0.05).



Figure 3: Mean durations of six consonants.

### 2.2.5. Center of gravity and skewness for /s/

We analyzed center of gravity and skewness at the midpoint of /s/ in six tokens (i.e., /sa/ in T1, T3, T4, and /su/ in T1, T2, T4). Mann-Whitney U test showed that there was no significant difference, suggesting that these two are not distinctive features. This contradicts with the finding on American English speakers, among whom homosexual males produced /s/ with more negative skewness than heterosexual males [1, 2, 14].

#### 2.3. Discussion

The acoustic results have shown that the homosexual group has lower mean F0, narrower F0 range, and less steep F0 slope than the heterosexual group. This contradicts with the previous report that no significant F0 difference was found between homosexual and heterosexual male speakers of American English [2, 5]. We will propose social psychological explanation for this phenomenon.

There is a stereotype that homosexual males tend to have a feminine voice which is associated with higher pitch and wider pitch range [13]. Unlike in USA, homosexual males are far from accepted in China. So, they may try to hide their sexual orientations in speech communication purposely by speaking in a way opposite to the expected pattern, hence resulting in lower pitch. In fact, a similar result (i.e., no significant difference in pitch between homosexual and heterosexual male speakers) was found among Cantonese speakers in Hong Kong, a city also with low acceptance of homosexual males [16].

Also, the Mandarin homosexual group shows smaller tVSA than the heterosexual group; especially, F1 of /a/ is lower in the homosexual group. In contrast, American homosexual males have lager vowel space and higher F1 of low vowels /æ/ and / $\epsilon$ /

than heterosexual males, as shown in Fig. 4 which is cited from [2]. Since females generally have higher F1 (larger mouth opening) of low vowels than males, it can be assumed that the formant pattern of American homosexual male speakers is a result of a selective adoption of the speech characteristics of heterosexual females [2]. So, the opposition between Mandarin and American English speakers may again be explained by the very low acceptance of homosexual males in China – they may try to hide their sexual orientations purposely by speaking in a way opposite to the expected feminine pattern, resulting in lower F1 of low vowels than heterosexual males.

Moreover, our finding that the aspiration/frication portions of consonants are longer in the homosexual group coincides with the previous reports for American English and Cantonese [3, 16]. Therefore, the lengthened fricatives might be universal speech style of homosexual males. Besides, it is reported that homosexual males tend to articulate more carefully when producing fricatives, hence resulting in longer durations of fricatives [16]. No gender difference in fricative duration has been reported in the literature, and homosexual males are generally not aware of this 'hidden' cue for sexual orientation. Thus, the characteristics of longer fricatives (than heterosexual males) are preserved in Mandarin homosexual males.

For center of gravity and skewness for /s/, no significant difference was found between Mandarin homosexual and heterosexual males. In contrast, previous studies on American English reported that homosexual males showed more negative skewness of /s/ than heterosexual males, giving a clear-speech style [2]. It is to be noted that females tend to produce /s/ with more negative skewness (i.e., more acoustic energy in the frequencies above the mean) than males [17]. Therefore, American English homosexual males produce /s/ in a feminine pattern, whereas Mandarin homosexual males seem to avoid the feminine pattern intentionally.

By applying linear mixed-effects model to vowel durations, interaction effects were also observed. For duration of /a/, there was a significant interaction between sexual orientation and tone type. For duration of /u/, there was a significant interaction between sexual orientation and consonant. Whether these interaction effects are meaningful or not, needs to be further tested on larger speech corpora.

In summary, a set of acoustic features in Mandarin speech produced by homosexual males show the pattern opposite to the expected feminine voice. However, longer aspiration/frication of consonants is observed, which is consistent with the finding for American English.



Figure 4: Acoustic vowel space of American homosexual males (black) and heterosexual males (grey) [2].

# 3. Perceptual experiment

#### 3.1. Method

Thirty participants (15M, 15F) were recruited in the perceptual experiment. They were all graduate students at university, with an average age of 23.3 years (SD=2.1). None of them had a reported history of hearing or cognitive disorders.

It was reported that American natives could identify sexual orientation of the speaker based on English isolated words [13]. Therefore, 18 monosyllabic words composed of 7 consonants (/p/, /p<sup>h</sup>/, /m/, /s/, /g/, /tg<sup>h</sup>/, /l/), 3 vowels, and 4 tones (excluding pseudo words) were selected for a perceptual experiment. Consonants /p<sup>h</sup>/, /s/, /g/, /tg<sup>h</sup>/ were selected because they gave most significant differences in duration between the two groups, as shown in Fig. 3. The unaspirated /p/ was included to compare with its aspirated counterpart /p<sup>h</sup>/. Also, the nasal /m/ and the lateral /l/ were selected to include all manners of articulation.

The perceptual experiment was conducted in a sound-proof booth. All 18\*24 = 432 word stimuli were presented in a random order using E-Prime Professional 2.0. The stimuli were divided into four blocks (132 in one block, and 100 in each of other blocks). Each participant wore a headphone AKG K77. After a stimulus was played back, the participant scored the gayness of the speaker within a few seconds on a 5-point scale: 1 (definitely heterosexual), 2 (probably heterosexual), 3 (hard to judge), 4 (probably heterosexual), and 5 (definitely homosexual).

### 3.2. Results

The average perceptual score of gayness was 3.1 (SD=1.4) for the homosexual group, and 2.5 (SD=1.3) for the heterosexual group. A chi-square test showed that there was no significant difference between the two groups (p < 0.001).

Table 4 shows the correlations between the perceptual score and all eight acoustic parameters giving significant differences between the two groups in the acoustic experiment (i.e., mean F0, F0 max, F0 range, F1 of vowels, F2 of vowels, and durations of aspirated stops, fricatives, and aspirated affricates).

Table 4: Correlations between the perceptual score and acoustic parameters (the asterisks indicate p < 0.001).

mean F0	F0 max	F0 range	F1 vowel
-0.06*	-0.13*	-0.14*	-0.26*
Dur stop	Dur fricative	Dur affricate	F2 vowel
0.01	0.11*	$0.06^{*}$	-0.09*

#### 3.3. Discussion

The experimental result that the perceptual scores of gayness were significantly different between the speech of homosexual and heterosexual males suggested that sexual orientation of Mandarin male speakers could be identified to some extent from the speech of monosyllabic words.

Among all acoustic features, F1 of vowels shows the most significant negative correlation with the perceptual score, suggesting that lower F1 of vowels is associated with higher score of gayness. This coincides with the acoustic result that the homosexual group had lower F1 than the heterosexual group.

Next, F0 max and F0 range show significant negative correlations with the perceptual score, suggesting that lower F0

max and narrower F0 range are associated with higher score of gayness. This coincides with the acoustic result that the homosexual group has lower F0 max and narrower F0 range.

Moreover, duration of fricatives shows a significant positive correlation with the perceptual score. This is consistent with Linville's finding that identification rate of homosexual male speakers is significantly correlated with duration of /s/ [1].

From the perceptual points of view, there can be more than one acoustic cues for sexual orientation of male speakers [14]. So, even the homosexual male speakers try to hide their sexual orientations purposely by reversing the features generally associated with the stereotype of homosexual males (e.g. higher-pitch, larger mouth-opening), their speech may still be perceived as gay-sounding due to the existence of other acoustic cues that are less known. Because for homosexual male Mandarin speakers both F0 and F1 of vowels contradict with the generally expected feminine patterns as we discussed earlier, duration of fricatives is quite possibly the most salient cue for sexual orientation of male Mandarin speakers.

# 4. Conclusions

Both acoustic and perceptual experiments were conducted to investigate the differences between Mandarin speech produced by homosexual and heterosexual males. The results of acoustic experiment on monosyllabic words showed that the homosexual group had lower F0, smaller tVSA, and longer aspiration/frication of consonants. Especially, F0 and formants of vowels, and skewness of /s/ in homosexual males showed opposite patterns between Mandarin and American English, while longer fricative duration was commonly observed. This may be explained by social psychological differences between the two societies. American English homosexual males show a feminine pattern, whereas Mandarin homosexual males try to hide the feminine pattern intentionally because of the low acceptance of homosexual males in China. The results of the perceptual experiment suggested that the perceptual score of gayness from monosyllabic words differed significantly between the two groups. Among the acoustic features under study, longer fricative duration is commonly observed across Mandarin and American English, and may be the most important cue for sexual orientation, though it will be further tested in our future work.

### 5. Acknowledgements

This work is supported by the Major Program of the National Social Science Fund of China (13&ZD189, 12&ZD177), and the project for Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Sciences (2017STD006).

### 6. References

- S. E. Linville, "Acoustic correlates of perceived versus actual sexual orientation in men's speech," *Folia Phoniatr Logop*, vol. 50, no. 1, pp. 35-48, 1998.
- [2] B. Munson, E. C. Mcdonald, N. L. Deboe, et al., "The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech," *Journal of Phonetics*, vol. 34, no. 2, pp. 202-240, 2006.
- [3] H. Rogers, R. Smyth, "Phonetic differences between gay-and straight-sounding male speakers of North American English," *Studies*, vol. 1, no. 3, pp. 5-6, 2003.
- J. B. Pierrehumbert, T. Bent, B. Munson, A. R. Bradlow, and J. M. Bailey, "The influence of sexual orientation on vowel

production," Journal of the Acoustical Society of America, vol. 116, no. 1, pp. 1905, 2004.

- [5] R. Gaudio, "Sounding gay: Pitch properties in the speech of gay and heterosexual males," *American Speech*, vol. 69, pp. 30-57, 1994.
- [6] E. C. Tracy and N. P. Satariano, "Differentiating between gay and heterosexual male speech," *Journal of the Acoustical Society of America*, vol. 129, no. 4, pp. 2421, 2011.
- [7] E. C. Tracy, "Judgments of self-identified gay and heterosexual male speakers of American English: Certain phonemes are more salient than others in determining sexual orientation," *Journal of Phonetics*, vol. 52, pp. 13-25, 2015.
- [8] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer," Version 6.0.14, http://www.praat.org/.
- [9] D. Mirman, "Growth curve analysis and visualization using R," Boca Raton: CRC Press, 2014.
- [10] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1-48, 2015.
- [11] R. V. Lenth, "Least-squares means: The R package Ismeans," *Journal of Statistical Software*, vol. 69, no. 1, pp. 1-33, 2016.
- [12] H. M. Liu, F. M. Tsao, and P. K. Kuhl, "The effect of reduced vowel working space on speech intelligibility in Mandarinspeaking young adults with cerebral palsy," *Journal of the Acoustical Society of America*, vol. 117, no. 6, pp. 3879, 2005.
- [13] E. C. Tracy, "Vowels or consonants: Which is more effective in distinguishing between self-identified gay and heterosexual male speakers of American English?" *Journal of the Acoustical Society* of America, vol. 137, pp. 2417, 2015.
- [14] L. Zimman, "Hegemonic masculinity and the variability of gaysounding speech: The perceived sexuality of transgender men," *Journal of Language & Sexuality*, vol. 2, no. 1, pp. 1-39, 2013.
- [15] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/, 2017.
- [16] H. Chi-Kuk, "A study of acoustic characteristics of Cantonese gay speech and perceived sexual orientations," MA thesis, The Chinese University of Hong Kong, 2007.
- [17] P. Flipsen, L. Shriberg, G. Weismer, H. Karlsson, and J. Mcsweeny, "Acoustic characteristics of /s/ in adolescents," J Speech Lang Hear Res, vol. 42, no. 3, pp. 663-677, 1999.
- [18] S. Mack, "A sociophonetic analysis of perception of sexual orientation in Puerto Rican Spanish," *Laboratory phonology*, vol. 1, no. 1, pp. 41-63, 2010.
- [19] S. Fuchs, and M. Toda, "Do differences in male versus female/s/reflect biological or sociophonetic factors," *Turbulent sounds: An interdisciplinary guide*, pp: 281-302, 2010.