9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Perception of vocal attractiveness by Mandarin native listeners

Anqi Xu[1], *Albert Lee*[2]

[1] Department of Speech, Hearing and Phonetic Sciences, University College London, UK
[2] Department of English Language, University of Tokyo, Japan

a.xu.17@ucl.ac.uk, alee@aless.c.u-tokyo.ac.jp

Abstract

Studies on Western societies show that male voices with acoustic parameters encoding a large body size (low f_0, narrow formant dispersion and f_0 range) were considered attractive, while the opposite was true for female voices. The present work investigates whether Mandarin native listeners judge voices of the opposite sex in the same way. We replicated the design in [1] with the added parameter of creaky voice, which is prevalent in North America nowadays and hotly debated in terms of attractiveness. Thirty-two participants (16 female) rated the attractiveness of synthetic stimuli controlled for f_0 height, formant dispersion, f_0 range and voice quality. Similar to studies on Western societies, Mandarin native listeners preferred breathy and modal voices to creaky and pressed voices. Moreover, large-sounding male voices with low f_0 and narrow formant dispersion were favored. However, a narrow f_0 range significantly lowered the attractiveness ratings, regardless of the gender of the voice. These results are discussed in light of the cross-linguistic / cross-cultural divergences in vocal attractiveness.

Index Terms: attractive voice, articulatory synthesis, creaky voice, Mandarin, body size

1. Introduction

As an indicator of sex [2], age [3], personality [2], emotional state [4] and reproductive ability [5], the human voice plays a key role in daily life. An attractive voice can additionally contribute to success in dating [6], career [7] and elections [8]. A considerable amount of literature has been published on the acoustic properties that are relevant to vocal attractiveness judgment. In the male voice, low fundamental frequency (f_0) and closely spaced formants are preferred, while female voices with high f_0 and wide formant dispersion are favored [9, 10]. In addition, f_0 range is also found to be a relevant factor though its exact effect is language-dependent. A notable example is that a narrow pitch range benefits male vocal attractiveness in British English [1], whereas studies on Japanese reported that the opposite was true [11]. Regarding voice quality, breathiness is associated with intimacy [4] and renders an attractive voice for both sexes [1], whereas a pressed voice undermines attractiveness judgment [1]. Recently the creaky voice is found to be prevalent among the young generation in North America (women [12]; men [13]), but whether it is genuinely attractive remains debatable. Although some state that vocal fry is perceived as well-educated [14], another study pointed out that creaky voice actually undermined the overall judgment [15].

One account that aims to clarify the mechanism underlying the perception of attractiveness is the body-size projection theory. Morton [16] proposed motivation-structural rules to rationalize the common principles behind animal calls. Birds and mammals adopt harsh, low-pitched sounds to encode a large body size while hostile and high-frequency sounds when frightened [16]. The theory was then extended by Ohala [17], who claimed that formant frequencies were also critical in projecting body size. Just like other animals, human males descended the larynx at puberty to be prepared for mate selection [18]. Based on the aforementioned accounts, a bio-informational dimensions (BID) theory was proposed, which consists of body-size projection, dynamicity, audibility and association, based on empirical evidence on perceived attractiveness of voices with acoustic features including f_0, formant spacing and voice quality [19]. An alternative account would be the averageness theory, which states that attractive faces and voices are close to the population mean (face [20]; voice [21]). For example, medium and low fundamental frequency did not differ in enhancing male vocal attractiveness [22].

To date, these hypotheses have been tested mainly on Western societies. In contrast, there is scarce work done on non-Western populations. Therefore, whether the same assumptions hold true across languages or are language-specific needs to be investigated. In order to fill in this gap, the objective of the current study is to find out whether Mandarin native listeners perceive the attractiveness of voice in the same way as their English-speaking counterparts; and if not, identify which acoustic parameters influence their attractiveness judgment.

2. Methodology

2.1. Stimuli

The Chinese sentence wǒmen yǎng yīngwǔ 我们养鹦鹉 'We keep a parrot' was used. The base sentences with four voice qualities were generated with the articulatory speech synthesizer VocalTractLab 2.1 [23]. Breathy, modal and pressed voices were generated using the default settings in the triangular glottis model. As for creaky voice, it is characterized by highly irregular periodicity and a sensation of repeating impulses [24]. In order to generate the irregularity, small uneven time constants were added under the 'f_0 gestures' function with modal glottal shape.

To confirm that the voice qualities of the base sentences were successfully synthesized, we conducted acoustic analysis (Table 1) and energy-band analysis (Figure 1) on the utterances using ProsodyPro 5.6.0 [25]. As shown in Figure 1, when voice quality went from pressed to breathy, signal energy decreased. By contrast, harmonic structure H1-H2* and measurements of spectral tilt H1-A1* and H1-A3* increased as indicated in Table 1. Creakiness can take many forms [26], and the version used here has a high deviation in periodicity (measured by jitter)

10.21437/SpeechProsody.2018-70

but also a spectral tilt that is more similar to that of modal voice, so as to tease apart the effects of creaky and pressed voices. The female base sentences were converted from the male ones using the 'ChangeGender' function in Praat [27] with formant shift by 1.2 and pitch medium shift by 12 semitones. The 8 base sentences (4 voice qualities × 2 gender) were preliminarily tested by five trained linguists. They reported no difficulty in distinguishing the voice qualities. Subsequently, the base sentences were manipulated independently of all other parameters (Script 3 in [1]), resulting in 108 stimuli for each gender (4 voice qualities × 3 fundamental frequency shifts × 3 formant dispersion shift ratios × 3 pitch range shift ratios), as shown in Table 2. In order to confirm that the synthetic stimuli sounded natural to native listeners, we ran a pilot experiment with 4 female native Mandarin listeners. They reported that most of the voices sounded extremely tall. Considering the height difference between the British [28] and Chinese populations [29], the formant dispersion of base sentences was further adjusted (increased by 1.1 compared to [1]).

Table 1: *Acoustic measurements of Chinese base sentences.*

Intended voice quality	H1-H2*	H1-A1*	H1-A3*	Center of Spectral Gravity	Jitter (%)	Shimmer (%)	HNR
Breathy	-0.5	-0.5	23.7	402.8	0.8	5.2	17.8
Modal	-1.3	-2.3	12.0	540.8	0.6	6.6	15.2
Creaky	-1.4	-2.2	12.2	536.4	1.2	6.1	15.5
Pressed	-3.5	-4.6	6.9	652.0	1.0	9.4	14.0

Table 2: *Manipulation of acoustic parameters*

Body size projection	Voice Quality	Pitch shift	Formant shift ratio	Pitch range shift ratio
Small	Breathy	+2 semitone	×1.1	×2
Medium	Modal	0	×1	×1
Large	Creaky/ Pressed	−2 semitone	×0.9	×0.25

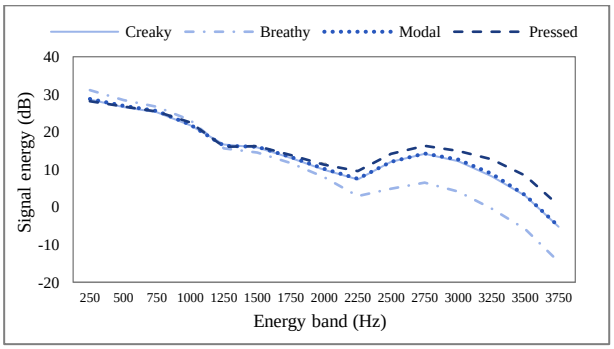

Figure 1: *Band energy profiles of Chinese base sentences.*

2.2. Listeners

Sixteen Mandarin female native listeners (mean ± SD age: 24 ± 1.15) and sixteen Mandarin male native listeners (mean ± SD age: 24.07 ± 2.11) were recruited to rate the stimuli. The participants were all born and raised in the northern part of China, speaking different varieties of Mandarin or Putonghua as their mother tongues. They were all university students with no self-reported history of speech or hearing impairment and reported to be heterosexual. All participants were paid a small remuneration for their time. Ethics approval for the collection of the data was obtained from the Faculty of Arts Research Committee at the University of Hong Kong (EA-MA00071).

2.3. Procedure

The stimuli were presented using the PRAAT object ExperimentMFC [27]. Participants listened to 108 stimuli of the opposite sex twice resulting in 216 trials in total. The stimuli were played in a randomized order with a 0.5 s silence in between. Participants listened to the stimuli through headphones in a quiet room and judged the attractiveness of the voice on a 1-5 scale with 1 being the most unattractive and 5 being the most attractive. No further instruction was given to the participants on how to rate the stimuli. They could replay any given stimulus up to three times. Before the experiment, every participant had three practice trials. To reduce fatigue, the participants were forced to take a 5-minute break after 108 trials. The entire experiment took approximately 25 minutes.

2.4. Analysis

We fitted mixed-effect models using the *lmer()* function of the *lme4* package [30] in R [31] for female and male voice ratings separately. The manipulated acoustic parameters 'voice quality', 'fundamental frequency', 'formant dispersion', 'pitch range' and their interactions were treated as potential fixed effects. We started with the simplest (null) model, which included only the random intercept for listener. By-listener random slopes for the fixed effects were introduced maximally if it achieved convergence and then the fixed effects were added incrementally. Likelihood ratio tests were used to check whether the inclusion of additional predictors contributed significantly to a better model. The post-hoc Tukey's comparisons were conducted by the *multicomp* package [32] in R [31].

3. Results

3.1. Male vocal attractiveness perceived by female listeners

Figure 2 displays the kernel probability density of each ratings of male voices by female listeners (1-5 scale) across the four different acoustic parameters. For any given violin plot, the width of the shaded area represents the proportion of the data located there. The boxplot inside displays the median of the distribution, interquartile range, maximum, minimum and outliers. Random intercept for listener and by-listener random slopes for voice quality and pitch were included in the mixed model. There were significant main effects of voice quality ($\chi2 = 17.306$, df = 3, $p < .001$), fundamental frequency ($\chi2 = 8.33$, df = 2, $p = .015$), formant dispersion ($\chi2 = 530.79$, df = 2, $p < .001$) and pitch range ($\chi2 = 496.55$, df = 2, $p < .001$). As verified by post-hoc Tukey's tests, breathy voice ($p = .013$) and modal voice ($p = .010$) were rated as more attractive than pressed voice. Additionally, large-sounding male voices with closely distributed formants increased the judgment compared to both narrow formant spacing and the baseline ($p < .001$). Furthermore, low fundamental frequency that encodes a large body size improved the ratings as well ($p < .001$). However, low and medium f_0 did not differ in attractiveness assessment ($p = .489$). As to f_0 range, unlike formant dispersion and fundamental frequency, a narrow pitch range was significantly inferior to both a normal and a wide one ($p < .001$).

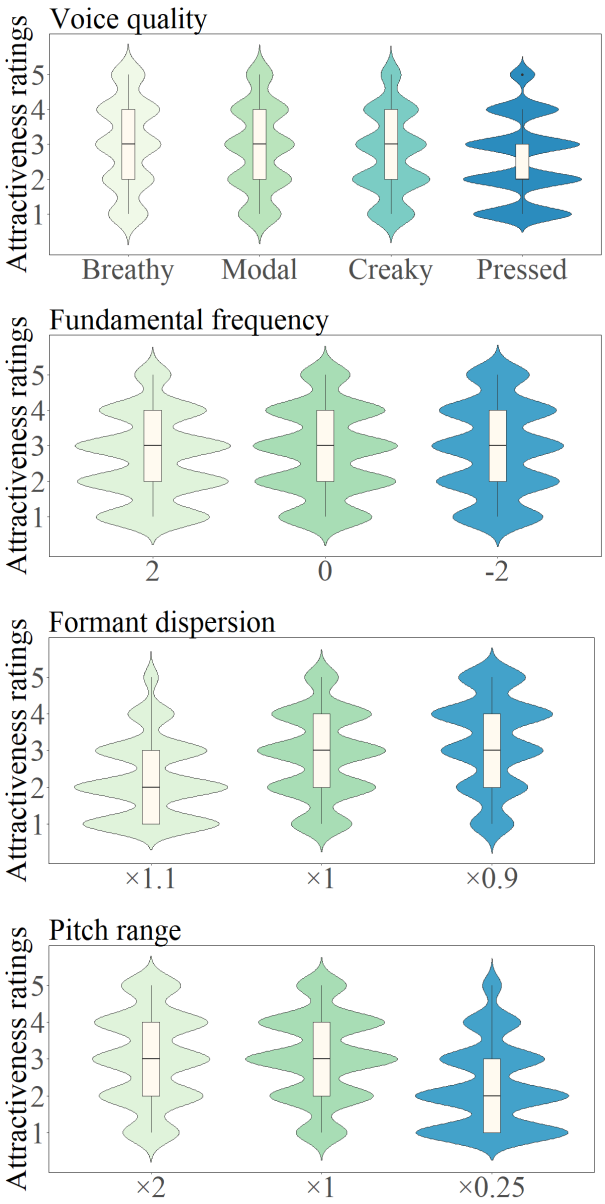

Figure 2: *Attractiveness ratings of male voice manipulated acoustic parameters by female listeners.*

3.2. Female vocal attractiveness perceived by male listeners

Figure 3 illustrates the effect of acoustic feature on female voice ratings by male listeners. Random intercept for listeners and by-listeners random slopes of for fundamental frequency, formant dispersion and pitch range were included in the mixed model. Similar to male voices, there were significant main effects of voice quality ($\chi2$ = 69.519, df = 3, p < .001), fundamental frequency ($\chi2$ = 6.928, df = 2, p = .031), formant dispersion ($\chi2$ = 20.366, df = 2, p < .001) and pitch range ($\chi2$ = 10.58, df = 2, p = .005) on female vocal attractiveness. As revealed in post-hoc comparisons, breathy voice was perceived to be significantly more attractive than creaky voice (p < .001) and pressed voice (p < .001). Modal voice led to higher ratings compared with creaky voice (p < .001). For fundamental frequency, manipulated conditions were not significantly different from their baseline conditions. Regarding formant frequency, widely distributed formants did not benefit the attractiveness judgment compared with unmodified female voices (p<.001) and narrowly spaced formants (p=.016). Similar to male voices, a narrow pitch range lowered attractiveness judgment compared with a normal pitch range (p<.001).

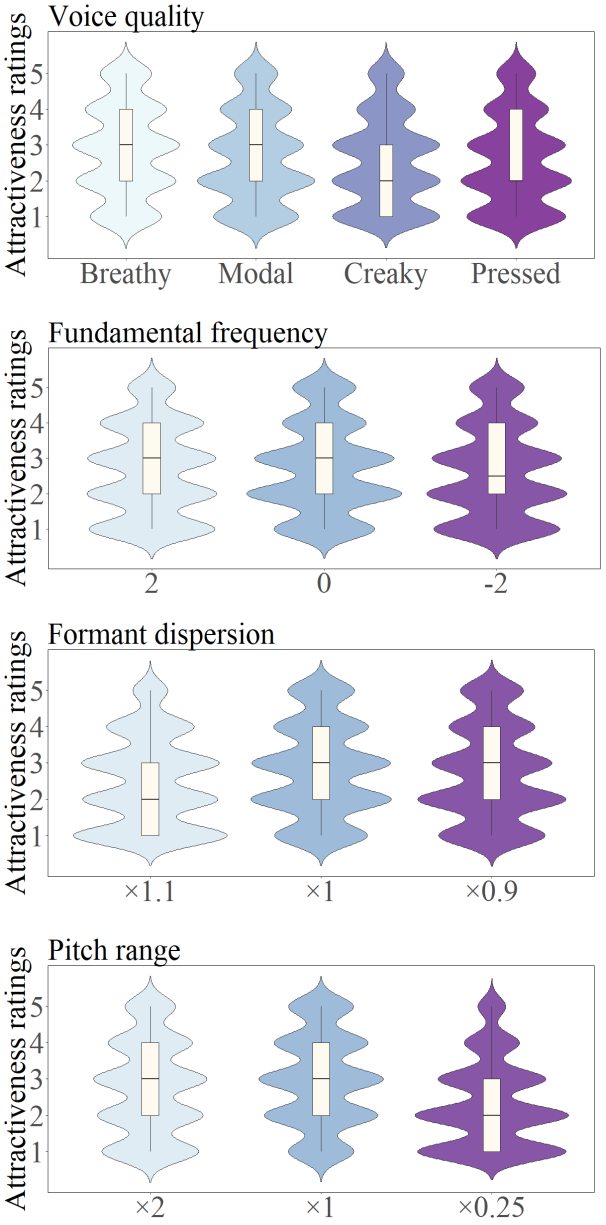

Figure 3: *Attractiveness ratings of female voice with manipulated acoustic parameters by male listeners.*

4. Discussion

The present work examined how acoustic properties biased Mandarin native listeners' assessment of attractive voices by replicating the methods in [1] with creakiness as a new parameter. A breathy low-pitched male voice with narrowly distributed formants and normal pitch range was perceived as the most attractive. With respect to a female voice, a breathy voice with moderately spaced formants and normal pitch range

sounded the most attractive. Below the similarities and differences between our data and those on English listeners are discussed.

In line with previous studies on Western societies, breathy voice raised the attractiveness perception of both male and female voices whereas pressed voice lowered the attractiveness judgment [1] in Mandarin. While creaky voice is reportedly prevalent in the USA [12, 13], we found that it did not benefit attractiveness ratings by Mandarin native listeners, which is consistent with [15]. Moreover, an attractive Mandarin male voice is low-pitched with narrow formant spacing, which projects a large body size just as in Western countries [1, 9].

Previous studies on Westerners indicated that a small-sounding female voice with high f_0 and wide formant dispersion was appealing [1]. Further, it has been demonstrated that exaggerated feminine voices, that is, even a relatively high-pitched female voice was preferred than averagely-pitched female voice [10, 33]. For Mandarin, however, neither high-pitched nor low-pitched female voice was perceived to be attractive, seemingly supporting the averageness theory. Listeners' preference for formant dispersion also showed a similar pattern. Interestingly, research on cross-cultural facial attractiveness likewise reported that Chinese male was indifferent to feminine faces in contrast with Western male [34]. These results would suggest that the preference for the small size code in a female voice is not universal but culturally specific. Up to now, research has tended to focus on cross-cultural variation in facial rather than on perceived vocal attractiveness. A possible area of future research would be to investigate what factors are associated with the cross-cultural divergence in the preference for sexually-dimorphic voice.

That a narrow pitch range had a negative effect on perceived vocal attractiveness regardless of gender contradicts with studies on British English native listeners [1]. Yet, it is somehow similar to what has been found in another tone language, i.e. Japanese [11, 35].The perception of pitch range may be swayed by the fact that in both Mandarin and Japanese f_0 movements are essential for distinguishing word meanings. Remarkably, within tonal languages, the perception of pitch range may be affected by speakers' utilization of pitch range as well. Mandarin native speakers were observed to have a wider pitch range than Min speakers in Taiwan [36]. Whether listeners from other parts of China also perceive pitch range in the same way needs further investigation.

In conclusion, although the body size projection theory and averageness theory likewise apply to Mandarin listeners' voice perception, there exist language-specific features, namely, the preferences for medium f_0 and formant spacing in female voices and the effect of pitch range on attractiveness judgment.

5. Acknowledgements

We are indebted to Dr.-Ing. Peter Birkholz for his advice on synthesizing creaky voice using VocalTractLab.

6. References

[1] Y. Xu, A. Lee, W. Wu, X. Liu and P. Birkholz, "Human vocal attractiveness as signaled by body size projection," *PloS ONE,* vol. 8, no. 4, p. e62397, 2013.

[2] C. D. Aronovitch, "The voice of personality: stereotyped judgments and their relationship to voice quality and sex of speaker," *Journal of Social Psychology,* vol. 99, no. 2, p. 207–220, 1976.

[3] P. H. Ptacek and E. K. Sander, "Age recognition from voice," *Journal of Speech,* vol. 9, no. 2, pp. 273-277, 1966.

[4] C. Gobl and A. Ní Chasaide, "The role of voice quality in communicating emotion, mood and attitude," *Speech and Communication,* vol. 40, pp. 189-212, 2003.

[5] J. R. Wheatley, C. A. Apicella, R. P. Burriss, R. A. Cárdenas, D. H. Bailey, L. L. M. Welling and D. A. Puts, "Women's faces and voices are cues to reproductive potential in industrial and forager societies," *Evolution and Human Behavior,* vol. 35, p. 264–271, 2014.

[6] C. Hodges-Simeon, S. Gaulin and D. A. Puts, "Voice correlates of mating success in men: examining "contests" versus "mate choice" modes of sexual selection," *Archives of Sexual Behavior,* vol. 40, no. 3, pp. 551-557, 2011.

[7] W. J. Mayew, C. A. Parsons and M. Venkatachalam, "Voice pitch and the labor market success of male chief executive officers," *Evolution and Human,* vol. 34, no. 4, pp. 243-248, 2013.

[8] C. C. Tigue, D. J. Borak, J. J. M. O'Connor, C. Schandl and D. R. Feinberg, "Voice pitch influences voting behavior," *Evolution and Human Behavior,* vol. 33, no. 3, pp. 210-216, 2012.

[9] S. A. Collins, "Men's voices and women's choices," *Animal Behaviour,* vol. 60, no. 6, pp. 773-780, 2000.

[10] D. R. Feinberg, L. M. Debruine, B. C. Jones and D. I. Perrett, "The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices," *Perception,* vol. 37, no. 4, pp. 615-623, 2008.

[11] M. Camp, "Japanese Lesbian Speech: Sexuality, Gender Identity, and Language," The University of Arizona Thesis and Dissertation Database, 2009. [Online]. Available: http://hdl.handle.net/10150/195371. [Accessed 25 February 2017].

[12] L. Wolk, N. B. Abdelli-Beruh and D. Slavin, "Habitual use of vocal fry in young adult female speakers," *Journal of Voice,* vol. 26, no. 3, pp. e111-e116, 2012.

[13] N. B. Abdelli-Beruh, L. Wolk and D. Slavin, "Prevalence of vocal fry in young adult male American English speakers," *Journal of Voice,* vol. 28, no. 2, pp. 185-190, 2014.

[14] I. P. Yuasa, "Creaky voice: a new feminine voice quality for young urbanoriented upwardly mobile American women," *American Speech,* vol. 85, no. 3, pp. 315-337, 2010.

[15] R. C. Anderson, C. A. Klofstad,, W. J. Mayew and M. Venkatachalam, "Vocal fry may undermine the success of young women in the labor market," *PLoS ONE,* vol. 9, no. 5, p. e97506, 2014.

[16] E. W. Morton, "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," *American Naturalist,* vol. 111, no. 981, pp. 855-869, 1977.

[17] J. J. Ohala, "An ethological perspective on common cross-language utilization of F0 of voice," *Phonetica*, vol. 41, pp. 1-16, 1984.

[18] W. T. Fitch and D. Reby, "The descended larynx is not uniquely human," *Proceedings: Biological Sciences*, vol. 268, no. 1477, pp. 1669-1675, 2001.

[19] Y. Xu, A. Kelly and C. Smillie, "Emotional expressions as communicative signals.," in *Prosody and Iconicity*, Philadelphia, John Benjamins Publishing Co., 2013, pp. 33-60.

[20] J. H. Langlois and L. A. Roggman, "Attractive faces are only average," *Psychological Science*, vol. 1, no. 2, pp. 115-121, 1990.

[21] L. Bruckert, P. Bestelmeyer, M. Latinus, J. Rouger, I. Charest, G. A. Rousselet, H. Kawahara and P. Belin, "Vocal attractiveness increases by averaging," *Current Biology*, vol. 20, no. 2, pp. 116-120, 2010.

[22] D. Riding, D. Lonsdale and B. Brown, "The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women," *Journal of Nonverbal Behavior*, vol. 30, no. 2, pp. 55-61, 2006.

[23] P. Birkholz, "VocalTractLab: Towards high-quality articulatory speech synthesis," [Online]. Available: http://www.vocaltractlab.de/. [Accessed 25 February 2017].

[24] J. Laver, The Phonetic Description of Voice Quality, Cambridge: Cambridge University Press, 1980.

[25] Y. Xu, "ProsodyPro — A tool for large-scale systematic prosody analysis," in *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 2013.

[26] P. Keating and M. Garellek, "Acoustic analysis of creaky voice," in *2015 Annual Meeting of the Linguistic Society of America*, Portland, 2015.

[27] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," [Online]. Available: http://www.praat.org/. [Accessed 25 February 2017].

[28] Center, Health & Social Care Information, "Health survey for England, 2014: Trend tables [NS]," 2015. [Online]. Available: http://www.hscic.gov.uk/catalogue/PUB19297. [Accessed 25 February 2017].

[29] General Administration of Sports of China, "2014 National Physical condition Survey," 2015. [Online]. Available: http://www.sport.gov.cn/n315/n329/c216784/content.html. [Accessed 25 Februay 2017].

[30] B. Douglas, M. Mächler, B. Bolker and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1-48, 2015.

[31] R Core Team, "R: A language and environment for statistical computing R Foundation for Statistical Computing," [Online]. Available: http://www.R-project.org/. [Accessed 25 February 2017].

[32] F. Bretz, T. Hothorn and W. Peter, Multiple Comparisons Using R, Boca Raton: CRC Press, 2010.

[33] D. E. Re, J. J. M. O'Connor, P. J. Bennett and D. R. Feinberg, "Preferences for Very Low and Very High Voice Pitch in Humans," *PLoS ONE*, vol. 7, no. 3, p. e32719, 2012.

[34] U. M. Marcinkowska, M. V. Kozlov, H. Cai, J. Contreras-Gardun˜o, B. J. Dixson, G. A. Oana, G. Kaminski, N. P. Li, M. T. Lyons, I. E. Onyishi, K. Prasai, F. Pazhoohi, P. Prokop, S. L. R. Cardozo, N. Sydney, J. C. Yong and M. J. Rantala, "Cross-cultural variation in men's preference for sexual dimorphism in women's faces," *Biology Letters*, vol. 10, p. 20130850, 2016.

[35] A. Xu, S. Leung and A. Lee, "Universal vs. language-specific aspects in human vocal attractiveness: An investigation towards Japanese native listeners' perceptual pattern," in *Proceedings of Meetings on Acoustics*, Honolulu, Hawaii, 2016.

[36] S. Chen, "The effects of tones on speaking fundamental frequency and intensity," *The Journal of the Acoustical Society of America*, vol. 117, no. 5, pp. 3225-3230, 2005.