Vocal attractiveness in Cantonese: A production study

Albert Lee[1], *Eva Ng*[2]

[1]The Education University of Hong Kong, Hong Kong
[2]Independent researcher, Hong Kong
albertlee@eduhk.hk, evang1294@gmail.com

Abstract

This paper presents a production study investigating the vocal strategies of Cantonese women when addressing an attractive vs. unattractive male. We recruited 19 young female native speakers of Hong Kong Cantonese who completed an attractiveness rating task, followed by a production task where they were presented a subset of the same faces. By comparing the rating results and corresponding acoustic data of the facial stimuli, we found that when young Cantonese women spoke to an attractive male, they were less breathy, lower in pitch, and with denser formants. Participants who were more satisfied with their own height used these vocal strategies more actively. These results are discussed in terms of the body size projection principle.

Index Terms: Cantonese, vocal attractiveness

1. Introduction

Voice and perceived physical attractiveness are closely related (e.g. [1]). It has been reported that physical attractiveness leads to advantages in situations such as dating [2] and is associated with more social support [3]. While one's face cannot be easily altered at least in the short run, adjusting voice is an immediately possible alternative. Apparently, much of what is known about vocal attractiveness may be explained by the principle of body size projection [4], which is also central to other phenomena such as animal calls. Therefore, a good understanding of vocal attractiveness is of practical, social, and theoretical importance.

The acoustic correlates of an attractive voice have been extensively studied in recent years. In languages that have been surveyed to date, vocal attractiveness seems to be neatly accounted for by the body size projection principle [4]. To male English listeners, an attractive female voice is high in pitch, breathy, and with wide formant dispersion, all of which signal a small body; to female English listeners, an attractive male voice (i) is low in pitch and (ii) has narrow formant dispersion, both signaling a large body, and (iii) is also breathy, signaling a smaller body [1]. Comparable perception studies on non-Western populations include Japanese [5] and Mandarin [6], which demonstrated cross-linguistic variations in the acoustic cues to an attractive voice. Findings of production studies (e.g. [7], [8]) agreed with those of perception studies in general. Even though creaky voice, which signals a large body size, is increasingly common among American female speakers, there is evidence that it is considered less attractive than a normal speaking voice [9], thus lending further support to the apparent universality of body size projection in vocal attractiveness. Although female English listeners' preference for a breathy male voice might appear to contradict this principle, it was suggested that breathiness may serve to neutralize otherwise aggressive-sounding male voice [1].

Although the perception of vocal attractiveness in western societies is relatively well understood, there is much less production data available, let alone from non-Western populations. To the best of our knowledge, to date there is no production study on vocal attractiveness in Cantonese. This study serves to fill this gap. While perception studies are useful for identifying the effect of individual acoustic cues, production data are essential as they show how these cues interact in everyday speech. In addition, production data can shed light on individual variability, which is increasingly important with the emergence of statistical tools capturing speakers as a random factor.

Based on the studies reviewed above (e.g. [1]), we expected that female Cantonese speakers would use vocal strategies to signal a small body size when addressing an attractive male, but they might also be creaky like their American counterparts. The prediction of creakiness is based on two reasons: (i) we tested well-educated young women in Hong Kong where the influence of western (including American) culture is prevalent, and (ii) the effect of voice quality in neutralizing one's projected body size was also observed in female English listeners' preferences in a male voice [1]. Therefore, in this study we test the following hypotheses (see Table 1):

Table 1: *Working hypotheses tested* (prediction for the attractive facial stimulus condition)

	Acoustic correlate	**Hypothesis**
H1	H1-A1	Decrease
H2	H1-A3	Decrease
H3	Formant disp.	Increase
H4	Median pitch	Increase
H5	Jitter	Increase
H6	Shimmer	Increase

H1 and H2 are related to the use of breathy voice. As the decrease in energy at higher frequencies from the first harmonic is the greatest for breathy voice and the least for creaky voice (see review in [10]), we expect to see decreased H1-A1 and H1-A3 in the attractive face condition. Formant dispersion (H3) is inversely related to vocal tract length, thus in the attractive face condition we expect to see more dispersed formants to project a shorter vocal tract. Finally, while there are different types of creaky voice [11], each with its own acoustic properties, as working hypotheses (H5 and H6) we hypothesised that Cantonese women would exhibit more cycle-to-cycle variability in the attractive face condition, thus increased jitter and shimmer.

2. Methods

2.1. Participants

Nineteen women participated in this study. They were all recruited in Hong Kong, speaking Cantonese as their first language, and university-educated. They aged between 19 and 25, and self-identified as heterosexual. All of them also spoke English and Mandarin as second languages. Their mean height was 159.4 cm (SD ±4.4). Participation was voluntary and no one received any monetary remuneration. No one reported any (history of) speech and hearing impairment.

2.2. Warm-up task

This study comprised three tasks: warm-up, facial attractiveness rating, and production task. All tasks were completed in the same session in a quiet room on university campus. Participants were recorded using a Logitech H340 microphone at a sample rate of 44.1 kHz.

During the warm-up session, participants were asked to say the semantically neutral utterance *你好啊，你讀咩科嫁？* 'Hello. What is your major?' three times without being presented any visual stimuli. The purpose of this task was to familiarise the participants with main production task, which will be described in §2.4.

2.3. Facial attractiveness rating

Fifty different male facial stimuli were used. Forty of the faces were relatively attractive Asian male celebrities, which included masculine and feminine faces. East Asian celebrities from Hong Kong, Korea, Japan, and China were used as their features are more familiar to our participants (*cf.* [12] but note [13]). The remaining stimuli were relatively less attractive male faces. All stimuli were publicly available images obtained from the Internet.

Participants were asked to rate the attractiveness of these 50 faces on a 1 ~ 10 scale (10 = most attractive) and write down their response on an answer sheet. They were told to base their ratings purely on how much they were attracted to each face, and to ignore any past knowledge of the respective males or experience they might have with people of similar appearances. Stimuli were presented in a randomised order in Microsoft Powerpoint slides.

2.4. Production task

Based on the ratings from §2.3, for each participant the five most attractive and five least attractive faces were used as target stimuli in a subsequent production task. In the event of faces with the same rating, those that were presented later in §2.3 were chosen. Each face was presented three times on separate occasions in random order. Participants were instructed to imagine themselves in a classroom setting, and that the male face was of a classmate sitting next to them. Participants were then to ask the male classmate *你好啊，你讀咩科嫁？* 'Hello. What is your major?'. From each participant, 30 utterances were recorded. Recordings were subsequently analyzed using ProsodyPro [14] (ver. 5.7.2), which allows manual checking of vocal pulses and automatically extracts numerous acoustic measurements, as will be presented below.

2.5. Post-hoc interview

Preliminary data analysis revealed a bimodal distribution which was seemingly related to participants' height. Specifically, we seemed to observe that taller participants seemed to behave in the opposite direction from the rest. To verify this, we sent out a questionnaire to gather information on participants' height and how satisfied they were about it. There were four questions in the questionnaire: (1) 'How tall are you?', (2) 'On a scale of 1 to 10, how satisfied are you about your own height?', (3) 'If you are not satisfied, how much taller / shorter would you like to be (in centimeters)?', and (4) 'What are you doing to address your unsatisfactory height (e.g. wearing high heels)?'. All participants except one responded (i.e. N = 18). Based on their response, participants were then classified in terms of how satisfied they were about their height, namely (H)ighly satisfied, (M)oderately satisfied, and (L)east satisfied. There were six participants in each category.

3. Results

We set out to test six hypotheses (see Table 1) to examine whether Cantonese women project a small body when addressing an attractive male. Results are shown in Figure 1, where attractive (A) and unattractive (U) facial stimuli are compared (coral and turquoise respectively) for each acoustic correlate of vocal attractiveness. The X-axis of Figure 1 represents how much speakers were satisfied with their own body height (converted into the three categories H, M, L, with H being the most satisfied).

For voice quality, H1-A1 was higher for unattractive stimuli, indicating more use of breathiness when participants spoke to an unattractive face; the same was true for H1-A3. Cantonese women also appeared to lengthen their vocal tract with denser formants in the H condition, thus projecting a larger body. Similarly, participants' median pitch was lower in the H condition. In terms of creakiness, participants showed higher jitter but lower shimmer in the H condition.

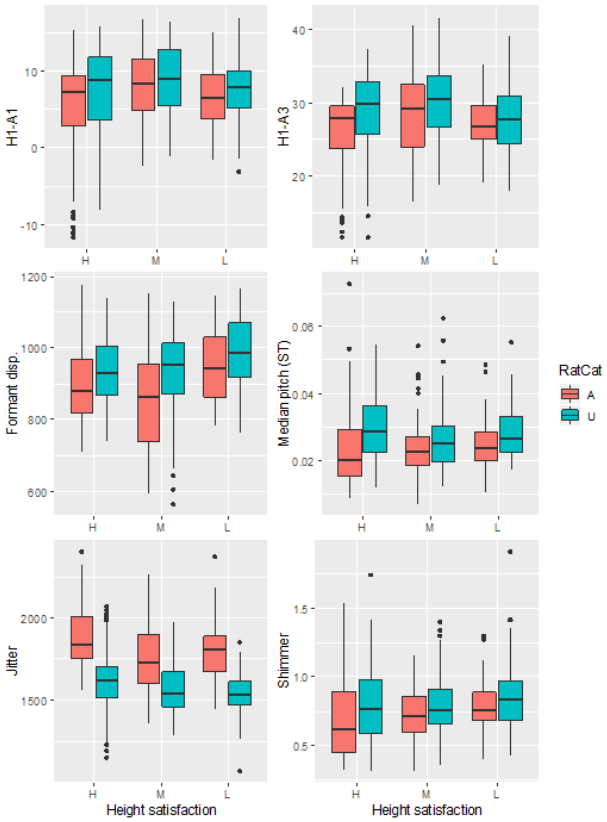

Figure 1: *Boxplots comparing acoustic correlates of projected voices in attractive vs. unattractive facial stimulus conditions.*

Initial analysis based on Pearson's *r* revealed substantial individual variability in vocal strategies. Therefore, for each acoustical parameter in Figure 1, we fitted a linear mixed effects model using the *lme4* package in R [15] (ver. 1.1-18-1) to model by-speaker variations. Model summaries are shown in Table 2. All models contained the continuous predictor of Rating (of male facial stimuli). In some models, we also included the interaction between Rating and Desire (i.e. desired change in height, see Question 3 in §2.5), which appeared to be a good heuristic of the individual variation in the initial analysis. No other manipulation of the data was performed. All models included by-speaker and by-item random intercepts; most also included by-speaker random slope for Rating (except for shimmer, in which model including the random slope for Rating would lead to non-convergence). A fixed effect is considered significant if the absolute value of the t-statistic is greater than or equal to 2.0 [16].

Table 2 shows that there was a significant main effect of Rating on all acoustical correlates of vocal attractiveness analysed ($t > 2.0$). This indicates that, after taking into account by-speaker variation, in general an attractive male face elicited significantly less breathiness (lower H1-A1 and H1-A3), longer vocal tract (denser formant dispersion), lower f_0 (median pitch), less regular cycle-to-cycle variation in f_0 (higher jitter) but more regular cycle-to-cycle variation in amplitude (lower shimmer). In addition, there was a significant interaction between Rating and Desire in all voice quality-related measurements. It can be seen in Figure 1 and Table 3 that the contrast between H and L faces in terms of acoustic correlates were bigger for speakers who are satisfied with their own height (in bold in Table 3).

Table 2: *Model summaries for different acoustic correlates (significant fixed effects are in bold).*

		Fixed effects			**Random effects SD**	
		β	***SE***	***t***	**Item**	**Part't**
H1-A1	(Intercept)	8.205	.940	8.733	.430	3.927
	Rating	-.411	.119	**-3.462**		.310
	Rating:Desire	.056	0.021	**2.642**		
H1-A3	(Intercept)	29.076	1.217	23.897	.430	5.163
	Rating	-.435	.135	**-3.223**		.393
	Rating:Desire	.055	.022	**2.489**		
F1-F3	(Intercept)	968.015	21.400	45.235	.000	89.727
	Rating	-8.909	2.179	**-4.088**		8.377
f_0	(Intercept)	.029	.001	21.915	.000	.005
	Rating	-.001	.000	**-4.218**		.001
Jitter	(Intercept)	1538.590	37.417	41.120	23.200	154.810
	Rating	42.467	5.411	**7.849**		15.460
	Rating:Desire	-1.942	.900	**-2.157**		
Shimm.	(Intercept)	.823	.036	23.012	.000	.144
	Rating	-.015	.003	**-4.692**		
	Rating:Desire	.001	.001	**2.011**		

Table 3: *Summary of differences (attractive* less *unattractive facial stimuli) in each acoustic correlate by speakers' satisfaction in their own height (H being the most satisfied).*

	Attractive *less* unattractive facial stimuli			
	H1-A1	H1-A3	Jitter	Shimmer
H	**-2.308**	**-2.627**	**27.447**	**-0.025**
M	-0.781	-1.773	18.349	-0.020
L	-0.787	-0.253	24.565	-0.016

4. Discussion

This study explored how Cantonese women projected their voice when speaking to an attractive vs. unattractive face. Results showed that, in the attractive face condition, most acoustic cues pointed to a larger body size. Participants were significantly less breathy in the attractive condition, supporting H1 and H2. In terms of vocal tract length, participants showed narrower formant dispersion in the attractive condition, signaling a larger body, rejecting H3. Their median f_0 was also significantly lower when addressing an attractive face, thus rejecting H4. For creaky phonation, in the attractive condition there was greater jitter but smaller shimmer, thus supporting H5 but not H6. For all voice quality-related measurements, there was a significant interaction between facial attractiveness rating and desired change in body height.

As mentioned in §1, male English listeners judged small-sounding acoustic cues to be more attractive [1], so even with cross-linguistic variation one would have expected Cantonese women to at least use some small-sounding cues in their production. Unexpectedly, in our data participants seemed to be always trying to project a large-sounding voice when speaking to an attractive face, unlike what the body-size projection account would have predicted. This is reminiscent of the prevalent use of creaky voice by female American speakers, despite that creakiness is considered unattractive [9]. The case of creaky voice in American female speech shows that speakers do not necessarily use vocal strategies that listeners consider attractive – knowingly or otherwise; as for Cantonese, as there is no comparable perception data available, it is unclear whether the large-sounding cues used by our female speakers are attractive to the local population. Another conceivable speculation is that speakers were taking into account social factors (classroom setting with friends nearby, interlocutor being a classmate), such that they deliberately avoided sounding ambitious in front of an attractive potential mate. This speculation, needless to say, needs to be carefully verified.

In our initial analysis, we had the impression that speakers' height might affect their vocal strategies – this was confirmed in Table 3. For all voice quality-related acoustic cues, participants who were satisfied with their own height manifested a larger contrast between the attractive and the unattractive stimulus conditions. Our data thus seem to suggest that although female Cantonese speakers have the same set of vocal strategies for attractive vs. unattractive mates, it is those who are confident in their own height that are using them more actively.

Participants in this study were well-educated young women who had been exposed to western culture since a very young age. They also spoke fluent English and Mandarin as second languages. This group of speakers thus represents only a subset of the local population. Future studies should look at other groups of speakers in the community, such as older monolingual speakers. Another potentially interesting factor to investigate would be the effect of menstrual cycle on speech production. To the best of our knowledge, to date there is only preliminary data on how the menstrual cycle affects voice quality in Cantonese women [17]. Understanding how physiological factors interact with vocal attractiveness would shed new light on this issue. Finally, it would also be useful to verify the present findings with articulatory data, such as electroglottography.

5. Conclusions

This paper has found that young Cantonese women projected a large-sounding voice when speaking to an attractive male face. This seems to disagree with the widely held body size projection principle which states that an attractive female voice is small-sounding. We also found that women who were confident in their own height adjust their voice more actively depending on the attractiveness of their mates. Further investigation is needed to understand the relationship between the present findings and those observed in other languages.

6. References

[1] Y. Xu, A. Lee, W.-L. Wu, X. Liu, and P. Birkholz, "Human vocal attractiveness as signaled by body size projection," *PLoS One*, vol. 8, no. e62397, pp. 1–9, 2013.

[2] E. Berscheid, K. Dion, E. Walster, and G. W. Walster, "Physical attractiveness and dating choice: A test of the Matching Hypothesis," *J. Exp. Soc. Psychol.*, vol. 7, pp. 173–189, 1971.

[3] B. R. Sarason, I. G. Sarason, T. A. Hacker, and R. B. Basham, "Concomitants of Social Support : Social Skills , Physical Attractiveness , and Gender," *J. Pers. Soc. Psychol.*, vol. 49, no. 2, pp. 469–480, 1985.

[4] E. S. Morton, "On the occurrence and significance of motivation-structural rules in some bird and mammal sounds," *Am. Nat.*, vol. 111, pp. 855–869, 1977.

[5] A. Xu, S.-S. Leung, and A. Lee, "Universal vs. language-specific aspects in human vocal attractiveness: An investigation towards Japanese native listeners' perceptual pattern," *Proc. Meet. Acoust.*, vol. 29, no. 60001, pp. 1–8, 2017.

[6] A. Xu and A. Lee, "Perception of vocal attractiveness by native Mandarin listeners," in *Proceedings of the 9th International Conference on Speech Prosody (SP2018)*, 2018, pp. 344–348.

[7] C. R. Hodges-Simeon, S. J. C. Gaulin, and D. A. Puts, "Different vocal parameters predict perceptions of dominance and tttractiveness," *Hum. Nat.*, vol. 21, no. 4, pp. 406–427, 2010.

[8] C. R. Hodges-Simeon, S. J. C. Gaulin, and D. A. Puts, "Voice correlates of mating success in men: Examining 'contests' versus 'mate choice' modes of sexual selection," *Arch. Sex. Behav.*, vol. 40, no. 3, pp. 551–557, 2011.

[9] 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. e97506, pp. 1–8, 2014.

[10] M. K. Gordon and P. N. Ladefoged, "Phonation types: A cross-linguistic overview," *J. Phon.*, vol. 29, pp. 383–406, 2001.

[11] L. Redi and S. R. Shattuck-Hufnagel, "Variation in the realization of glottalization in normal speakers," *J. Phon.*, vol. 29, pp. 407–429, 2001.

[12] V. Coetzee, J. M. Greeff, I. D. Stephen, and D. I. Perrett, "Cross-cultural agreement in facial attractiveness preferences: The role of ethnicity and gender," *PLoS One*, vol. 9, no. e99629, pp. 1–8, 2014.

[13] D. Burke, C. Nolan, W. G. Hayward, R. Russell, and D. Sulikowski, "Is there an own-race preference in attractiveness?," *Evol. Psychol.*, vol. 11, no. 4, pp. 855–872, 2013.

[14] 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)*, 2013, pp. 7–10.

[15] D. M. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, "Fitting Linear Mixed-Effects Models using {lme4}," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015.

[16] A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press, 2007.

[17] C. H. Li, "Voice onset time in Cantonese women across the menstrual cycle," *J. Acoust. Soc. Am.*, vol. 140, no. 4, p. 3112, 2016.