VOICE QUALITY: A PRELIMINARY STUDY ON THE PHONETIC DISTINCTIONS OF TWO CANTONESE ACCENTS

Roxana Fung

The Hong Kong Polytechnic University
roxana.fung@polyu.edu.hk

ABSTRACT

Hong Kong Cantonese and Guangzhou Cantonese are two fully mutually intelligible Chinese dialects that share almost the same phonological system. However, native speakers have no difficulty in distinguishing these two accents. The baffling question of what phonetic features account for the identification of these two Cantonese variants has perplexed many minds but remains unsolved. This study is a very preliminary foray aiming at solving this puzzle from a less studied linguistic domain, voice quality. Results of acoustic analyses reveal that there are significant differences in pitch and phonation of these two accents.

Keywords: voice quality, phonation, pitch, Cantonese, sociophonetics

1. INTRODUCTION

Cantonese is the standard form of the Yue dialects, a primary branch of Chinese spoken in the southern region of China. There are two prestige varieties of Cantonese: Guangzhou Cantonese (GZC) and Hong Kong Cantonese (HKC). GZC is mainly spoken in the city of Guangzhou in Mainland China, whereas HKC is spoken in Hong Kong, which has long been under British rule. These two accents of Cantonese share almost the same phonological system and are fully mutually intelligible. The two accents differ mainly in their vocabularies due to the different socio-political systems of the two cities. The phonological distinction between them that has been frequently mentioned concerns their tones. Table 1 shows the tonal inventory of standard Cantonese. For the high level tone (T1), a high-falling contour (53) may be produced as a free variant among some older GZC speakers but are not attested in HKC. Nonetheless, most native speakers of the two accents claim that they have no difficult in distinguishing the two accents even without the hint of the high-falling tone. The question of what phonetic features help to identify these two accents has perplexed many minds.

To date, Wu [1] has been the only work devoted to the investigation of the phonetic differences of the two accents. It conducted acoustic analyses on some words and short exchanges produced by 20 undergraduate students (10 GZC + 10 HKC). However, no major phonetic differences in the realization of consonants, vowels and tones between these two accents were found.

Table 1: Cantonese Tones.

Tone Number	Contour	Pitch value
T1	High level	55
T2	High rising	35
T3	Mid level	33
T4	Extra-low level	11/21
T5	Low rising	23
T6	Low level	22

Though Wu's study has not provided an answer for the question, it revealed that the noticeable phonetic differences between these two accents might not lie in the segmental level. This conjecture is, in fact, supported by the anecdotal remarks made by the native speakers of the two accents. Some of their comments collected from the web are summarized as follows:

- The 'intonation' of HKC tends to be soft, light, crispy and uplifting; whereas that of GZC tends to be heavy, low and has a 'substantial' feeling.
- HKC speakers have a higher pitch and a flatter pitch range than the GZC speakers.
- HKC has a faster tempo than GZC.
- HKC is not accurately articulated and is full of 'lazy sounds.'

Translating these impressionistic remarks into linguistic terms, the first remark concerns the phonation of the two accents. The second remark points at the pitch range. The third remark is related to speech rhythm. The third remark may be caused by some on-going sound changes in HKC among younger speakers, such as the alevolarization of velar finals and the merger of some tones.

A pilot study has been conducted to explore the phonetic differences between the two accents in the domains of voice quality and speech prosody using the speech collected by two previous studies in tone mergers in GZC [2] and in HKC [3]. This paper reports the findings related to the first two remarks, the pitch range and the phonation of the two accents only.

2. PITCH RANGE

2.1. Background

The differences in pitch range between GZC and HKC are related to pitch level and pitch span. Pitch span involves the degree of F0 variation in a speaker's voice, whereas pitch span the shifting of the range of F0 values. Variations in pitch level and pitch span have been most readily found in emotive speech. However, such variations are also found to play a role in marking social identify for regionality and social class. For instance, many white females seemed to use a wider pitch span that African American females or males of either ethnic group in the south of the United States [4]. The English produced by Italians have overall higher pitch level and narrower pitch span than those produced by the Americans [5]. However, there is no consensus on how to quantify pitch range. Pitch level has been quantified by either the mean or the median F0 of a speaker in a particular speaking situation. Pitch span has been measured by the difference between min F0 and max F0, the standard deviation of F0 around the mean F0 or the LTD measures [6,7]. Furthermore, the gauging of pitch range is particularly challenging in tone languages.

2.2. Methodology

2.2.1. Talkers and speech materials

A total of 24 talkers in three age groups with balanced gender (2 talkers x 2 gender x 3 age groups x 2 accents) were selected. The three age groups were: young (20-25), middle-aged (35-45), and senior (50-58).

All the 6 tonal contrasts of 3 CV roots: [si], [ji] and [fu], were embedded in two carrier phrases, "I am now reading the character ___ for you." and "This character is ___." for the talkers to read aloud. A total of 864 syllables (3 CV roots x 6 tones x 2 positions x 24 talkers) were analysed.

2.2.2. Measurements

The target syllables were segmented from the sentence carriers manually. The F0 of each syllable was tracked by a Praat script at 11 time points. The tracked F0 were converted to ERB rate scale according to the following formula: 16.7 x log $(0.006046 f + 1)$ where f is the frequency of sound in Hz. The pitch level was taken in this study as the mean of F0 in ERB of the 66 time points (11 time points x 6 tones) of each CV root in each sentential position. The pitch span was taken as the difference between the F0 of T1 (the highest level) tone and that of T4 (the lowest level) tone of each CV root at 50% duration. After the 50% duration, the realizations of T4 may have dramatic downdrift and creaky voice, making the measurement fairly difficult.

2.3. Results

The mean values of the pitch level and the pitch span for each accent, age group and gender are summarized in Table 2 and 3 respectively. The results were followed up by non-parametric Kruskal-Wallis statistics. For pitch level, there was no significant difference between GZC and HKC (p=0.6). As for pitch span, HKC was significantly higher than GZC, $\chi^2(1) = 17.8, p < .001$.

Table 2: Means of pitch level of the two accents (in ERB with standard deviation).

Accent	Age	Male	Female	
GZC	Young	4.1 (0.3)	5.7 (0.4)	
	Middle	4.2 (0.1)	5.2 (0.1)	
	Senior	3.7 (0.5)	4.8 (0.2)	
	Total	4.0 (0.4)	5.2 (0.5)	4.6 (0.7)
HKC	Young	3.7 (0.3)	5.3 (0.4)	
	Middle	3.6 (0.2)	5.1 (0.3)	
	Senior	4.1 (0.1)	5.5 (0.2)	
	Total	3.8 (0.3)	5.3 (0.4)	4.6 (0.8)

Table 3: Means of pitch Range of the two accents (in ERB with standard deviation).

Accent	Age	Male	Female	
GZC	Young	1.2 (0.1)	1.7 (0.4)	
	Middle	1.5 (0.3)	1.6 (0.6)	
	Old	1.1 (0.1)	1.6 (0.6)	
	Total	1.2 (0.3)	1.6 (0.5)	1.4 (0.5)
HKC	Young	1.6 (0.6)	1.8 (0.6)	
	Middle	1.4 (0.5)	2.0 (0.6)	
	Old	1.7 (0.3)	2.0 (0.3)	
	Total	1.7 (0.5)	1.9 (0.5)	1.8 (0.5)

2.4. Discussion

Our findings do not support the anecdotal remarks that HKC has higher pitch level and flatter pitch range as compared with GZC. On the contrary, the data suggest that the pitch level of HKC is comparable to that of GZC. Although there is a difference in pitch range, it is HKC, not GZC as portrayed, as a wider range. Does this mean that the intuition of native speakers is incorrect? It is possible that the measurements adopted in this study may in fact not be able to capture the differences in the pitch level and pitch span of the two accents. Due to the nature of our speech samples available, our measurements basically compute the average pitch level and the tonal space of the speakers demonstrated in reading monosyllabic words rather than natural utterances. Furthermore, some speakers have adopted a casual style in the reading, but some an exaggerated style. The style difference should critically affect the pitch range of the speech. Hence, different measures and different choice of materials may be employed in further investigation in this issue.

3. PHONATION

3.1. Background

Voice quality has long been regarded as a mysterious domain by language variationists [4]. Nevertheless, there have been a few sociolinguistic investigations into voice quality in recent years. It was found that particular phonation types characterize local dialects in Scotland [8, 9]. In the American English, breathier voice was identified as a factor separating the speech from Asian Americans from the non-Asian Americans [10]. In two varieties of New Zealand English, phonation was also proved to be a marker of ethnicity [11].

In tone languages, tone and phonation co-vary to various extents. In some Chinese dialects, a particular non-modal phonation type may be consistently associated with a tonal category as either a contrastive feature (e.g. Jingning, a Wu dialect) or non-contrastive feature (e.g. Longquan, a Wu dialect). In other dialects, a non-modal phonation type is inconsistently associated with a tonal category (e.g. Mandarin and Cantonese). In Cantonese, it is found that some speakers tend to display creakiness in the production of the low level tone (T4) and at the turning point of the high rising tone (T2) [e.g. 12]. However, creakiness is a personal style rather than a phonemic feature in the tonal system. However, Yu and Lam [12] claimed that Cantonese listeners were sensitive to details of creak production and may use the phonation information as a cue for tone perception. As previous studies mainly focuses on the investigation of the correlation of phonation and tone in Chinese dialects, the exploration of the role of phonation in accent identification has never been conducted.

3.2. Methodology

3.2.1. Phonation Talkers and speech materials

A total of 1584 syllables (66 syllables x 24 talkers) were analysed. The 66 syllables consisted of four vowels, [i], [ɔ], [a], [i], with all six contrastive tones. These syllables were chosen since we tried to use as many commonly occurred syllables as possible from two different previous studies to form a set of data with controlled vowels and tones.

3.2.2. Measurements

Acoustic measurements were made semi-automatically over the mid-portion of the entire vowel duration of the selected syllables with VoiceSauce [13]. The formant values of some vowels were corrected manually. The following six parameters: H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, CPP were obtained. Asterisks indicate that the harmonic amplitudes were corrected to for the effects of formants.

H1*-H2* is the difference in amplitude between the first harmonic (H1) and the second harmonic (H2). It is the most popular measure for phonation type. H2*-H4* was the difference in amplitude between H2 and H4. It is not so commonly used in phonation study. However, it was found to be associated with a less still vocal fold-layer and/or breathy voice based on a physical model of vocal folds [14]. H1*-An* was the difference in amplitude between H1 and the strongest harmonic in the nth formant (An). CPP was the difference in amplitude between the cepstral peak and the value of regression line at the cepstral peak. It was an indicator of harmonics-to-noise ration. Noise excitation was an important component of breathy voice.

The values for the first five parameters were expected to be large and positive for breathy voice and small and/or negative for creaky voice. But the value of CPP was expected to be small for breathy voice and large for creaky voice, just the opposite of all the other measures.

3.3. Results

The mean values of the acoustic parameters for each accent, age group and gender are summarized in Table 4. The mean values of the acoustic measures between the two accents were followed up with non-parametric Kruskal-Wallis statistics. For H1*-H2* and CPP, the mean values of HKC was significantly higher than those of GZC ($p < .001$ for both). As for the remaining 4 parameters, the mean values of GZC were significantly higher than those of HKC ($p < .001$for all): H1*-H2* ($\chi^2(1) = 38$); H2*-H4* ($\chi^2(1) = 82$); H1*-A1* ($\chi^2(1) = 137$), H1*-A2* ($\chi^2(1) = 94$), H1*-A3* ($\chi^2(1) = 62$), CPP ($\chi^2(1) = 25$).

Table 4: Mean values of the acoustic parameters (in dB with standard deviation).

Parameter	Accent	Age	Male	Female
H1*-H2*	GZC	Young	6.7 (3.4)	4.5 (6.3)
		Middle	2.6 (4.4)	7.0 (5.4)
		Senior	0.1 (3.7)	5.9 (4.2)
		Total	4.5 (5.3)	
	HKC	Young	3.4 (2.9)	6.2 (5.4)
		Middle	2.9 (3.2)	5.8 (4.9)
		Senior	7.4 (2.8)	6.0 (4.0)
		Total	5.3 (4.3)	
H2*-H4*	GZC	Young	6.2 (5.5)	2.9 (5.1)
		Middle	5.5 (5.9)	5.3 (4.8)
		Senior	9.4 (5.2)	5.6 (5.1)
		Total	5.8 (5.9)	
	HKC	Young	5.4 (5.2)	1.0 (6.7)
		Middle	4.4 (4.8)	1.5 (4.4)
		Senior	5.2 (4.9)	2.5 (3.9)
		Total	3.3 (5.3)	
H1*-A1*	GZC	Young	19.7 (7.9)	12.9 (8.3)
		Middle	15.0 (10.9)	18.0 (10.0)
		Senior	21.2 (9.8)	17.5 (6.9)
		Total	17.4 (9.5)	
	HKC	Young	14.8 (9.0)	13.5 (7.3)
		Middle	13.5 (8.4)	12.6 (7.5)
		Senior	18.6 (7.6)	13.6 (7.2)
		Total	14.4 (8.1)	
H1*-A2*	GZC	Young	24.4 (8.2)	16.5 (8.4)
		Middle	15.1 (10.6)	23.9 (8.2)
		Senior	26.1 (7.0)	22.5 (7.5)
		Total	21.4 (9.3)	
	HKC	Young	19.6 (8.0)	16.7 (6.7)
		Middle	18.6 (6.9)	14.6 (8.2)
		Senior	24.0 (7.8)	15.4 (7.4)
		Total	18.2 (8.2)	
H1*-A3*	GZC	Young	12.7 (8.8)	9.6 (10.1)
		Middle	6.7 (13.6)	19.5 (8.4)
		Senior	21.6 (7.6)	16.6 (6.4)
		Total	14.5 (10.8)	
	HKC	Young	11.2 (9.8)	9.8 (7.8)
		Middle	9.6 (10.3)	8.8 (9.5)
		Senior	17.9 (9.5)	8.7 (7.9)
		Total	11.0 (9.7)	
CPP	GZC	Young	21.6 (2.2)	20.3 (3.4)
		Middle	20.8 (2.4)	20.2 (2.6)
		Senior	22.4 (2.0)	20.3 (3.3)
		Total	20.9 (2.8)	
	HKC	Young	21.6 (3.9)	22.7 (4.0)
		Middle	23.1 (3.3)	21.4 (2.5)
		Senior	21.3 (2.9)	24.4 (2.6)
		Total	22.4 (3.4)	

3.4. Discussion

The results of the six acoustic parameters demonstrated that there were differences in phonation production between GZC and HKC. The measurements of all the acoustic parameters suggested that GZC was breathier than HKC, except H1*-H2*, which suggested the opposite. Several considerations can be drawn from these seemingly contradictory findings in the current study. There are different ways to produce breathy vs creaky voice. Speakers may differ in the acoustic details of their contrast. Furthermore, different measures reflect different aspects of production. They may or may not distinguish phonation category. Previous studies have shown that different languages of known phonation type are distinguished by different measures [15]. Although H1*-H2* has been identified as one of the most successful measures for phonation and which has been applied to many languages, it may fail to distinguish phonation types in some languages. It has been reported that H1*-H2* distinguished breathy vs modal phonation in eight out of 10 language samples. H1*-An* and CPP distinguish the modal and breathy categories of the two languages that H1*-H2* did not work [16]. Cantonese may be similar to these two languages in that H1*-H2* is not an effective measure of their phonatory production. We may conclude in this study that our findings confirm the impressionistic remark that GZC is breathier than HKC.

4. CONCLUSION

This is the first acoustic analysis of voice quality differences in Cantonese, and in Chinese dialects as well. The findings confirm that there are acoustic differences both in pitch range and phonation between GZC and HKC. This pilot study lends a strong support for future large-scale investigations on the phonetic distinction between the accents in the voice quality domain.

The future study should collect articulatory data, acoustic data and perceptual data. It is important to note that phonation is a complex process, and the acoustic measures are unlikely to inform us all the details of the phonatory production in languages. Recently, a better understanding on how languages use articulators other than the vocal folds to produce different phonation has been put forward through direct observation of the laryngopharynx (e.g. [17]) The findings of this study prompt future work in determining which articulatory manoeuvres are responsible for the phonation differences in the two accents. After identifying the effective acoustic measures through acoustic analysis, perception experiment can then be conducted to confirm if native speakers are able to perceive the acoustic differences identified, and find out the acoustic measure that best distinguishes the two accents.

5. ACKNOWLEDGEMENTS

The author wishes to thank Dr Ling Zhang for her assistance in conducting the analysis and acknowledge support from the Internal Competitive Research Grant provided by The Hong Kong Polytechnic University (4-ZZDD).

6. REFERENCES

[1] Wu, W.-L. 2006. *A Comparative Analysis of the Phonetics of Hong Kong Cantonese and Guangzhou Cantonese*. MPhil thesis, The University of Hong Kong.

[2] Ou, Jinghua. 2012. *Tone Merger in Guangzhou Cantonese*. MPhil thesis, The Hong Kong Polytechnic University.

[3] Fung, R., Wong, C. 2015. The asymmetry of the asymmetry: An interplay of production and perception in the multiple tone mergers in Hong Kong Cantonese. Manuscript in preparation.

[4] Thomas, E. 2011. *Sociophonetics: An Introduction*. Basingstoke: Palgrave Macmillan.

[5] Busa, M., Urbani M. 2011. A cross linguistic analysis of pitch range in English L1and L2. *Proc. 17th ICPhS Hong Kong*, 380-383

[6] Mennen, I., Schaeffler, F., Docherty, G. 2008. A methodological study into the linguistic dimensions of pitch range differences between German and English. *Proc. 4th Speech Prosody University of Campinas, 5527-530.*

[7] Patterson, D. 2000. *A Linguistic Approach to Pitch Range Modeling*. PhD. dissertation, University of Edinburgh.

[8] Esling, J. H. 2000. Crosslinguistic aspects of voice quality. In: Kent R. D., Ball, M. J. (eds). *Voice Quality Measurement*. San Deigo: Thomson Learning Singular, 25-35.

[9] Stuart-Smith, J. 1999. Glasgow: Accent and voice quality. In: Foulkes P., Dochetty G. (eds), *Urban Voices*. London: Arnold, 203-222.

[10] Newman, M., Wu, A. 2011. 'Do you sound Asian when you speak English?' Racial identification and voice in Chinese and Korean Americans' English. *American Speech* 86.2: 152-178.

[11] Szakay, A. 2012. Voice quality as a marker of ethnicity in New Zealand: From acoustics to perception. *Journal of Sociolinguistics* 16.3: 382-397.

[12] Yu, K. M., Lam, H. W. 2011. The role of creaky voice in Cantonese tonal perception. *Proc. 17th ICPhS Hong Kong*, 2240-2243.

[13] Shue, Y.L., Keating, P., Vicenik, C. 2009. VoiceSauce: A program for Voice Analysis. Presented at Fall 2009 meeting of ASA San Antonio.

[14] Zhang, Z., Kreimna, J., Gerratt, B.R. 2011. Perceptual sensitivity to changes in vocal fold geometry and stiffness. *J. Accoust. Soc. Am*. 129, 2529.

[15] Keating, P. A., Esposito, C.M. 2006. Linguistic Voice Quality. *Proc. 11th Australasian International Conference on Speech Science and Technology.*

[16] Esposito, C.M. 2006. *The Effects of Linguistic Experience on the Perception of Phonation*. Ph.D. dissertation, UCLA.

[17] Edmondson, J. A., Esling, J. H. 2006. The values of throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology* 23: 157-191.