What is a 'substantial' voice? An LTAS study on two Cantonese accents

Eugene Wong[1], Roxana Fung[2]

[1]University of Minnesota, [2]The Hong Kong Polytechnic University
wong0703@umn.edu, roxana.fung@polyu.edu.hk

ABSTRACT

Hong Kong Cantonese (HKC) and Guangzhou Cantonese (GZC) are two major accents of Cantonese, but their phonetic differences are understudied. Laymen claim that HKC is 'crispy' whereas GZC is 'substantial'. The current study examined their differences in the domain of voice quality. We analyzed text reading and free speech samples from 35 native talkers using LTAS. Results show that: (i) HKC exhibits more energy than GZC in F1, F2 and higher frequency regions; (ii) an amplitude peak exists at around F4 in GZC but not in HKC; and (iii) the two accents differ in mean spectral energy, spectral tilt, and level difference between f0 and F1 region (L1-L0). We explored the articulatory-acoustic relationships along the dimensions of glottal tension, glottal noise, and resonance between the two accents. A 'substantial' voice may indicate a more resonant speech induced by breathy-lax voice and lowered larynx in GZC.

Keywords: Cantonese, accents, voice quality, LTAS, sociophonetics

INTRODUCTION

1.1. Hong Kong Cantonese and Guangzhou Cantonese

Cantonese is a dialect of Chinese languages spoken in the southern part of China (mainly in Guangdong province) and in Hong Kong. There are two main accents of Cantonese: Hong Kong Cantonese (HKC), which is spoken in Hong Kong, and Guangzhou Cantonese (GZC), spoken in Guangzhou city. HKC and GZC are fully mutually intelligible as they share almost the same phonological system. However, native talkers claim that they have no difficulty in discerning the two accents. Although what phonetic features account for the identification of these two Cantonese variants have perplexed many minds, studies comparing the phonetic differences between the two accents are scare. One study examined the consonants, vowels and lexical tones in university students who spoke HKC and GZC [1]. The acoustic analyses showed that there were minimal differences in segmental features between the two accents.

Unlike professional linguists, many native talkers, however, have been enthusiastically expressing their opinion on the internet. Laymen have pointed out many lexical differences between the two accents. As for pronunciation, their descriptions were very abstract. For instance, they described HKC as 'soft, light, crispy and uplifting'; whereas GZC was "blunt, heavy, low, more cadence, giving people a 'substantial' feeling." Their perceptual descriptions indeed point towards the voice quality domain. Translating into acoustic terms, the distinction between the two accents may be caused by the differences in muscular tension and excitation of glottal noise. The so-called 'substantial' feeling seems to be related to the resonating quality of the voices, induced by larynx height.

1.2. Voice quality and social information

In fact, voice quality settings are considered the quasi-permanent features of someone's voice and are a major component of one's accent (see [2] and [3]). In a broad sense, voice quality features are the habitual settings of the larynx and the vocal tract when someone speaks. These articulatory settings may be constrained by the biological makeup of one's vocal organs, yet they could still be shared by the individuals of a social group. Esling [4] remarks that voice quality characteristics reflect the accent of the entire social group, with individuals' accent differing within the boundaries of the social norms. He further points out that native talkers are very good at detecting the speech patterns that do not originate in their community.

The study by [5] was the first attempt to explore the phonetic differences between HKC and GZC in the voice quality domain. That study involved an acoustic analysis on speech samples produced by 24 talkers of the two accents across three age groups with balanced gender. A total of 864 CV syllables extracted from sentences produced in a reading task were analysed using six acoustic parameters, namely H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3* and CPP. The results showed that there were differences in the phonation of the two accents. The measurements of all the acoustic parameters suggested that HKC was tenser than GZC, except H1*-H2*, which suggested the opposite. The seemingly contradicting results may be due to the small sample size or the nature of the speech samples. Although the vowels and tones were well controlled in that study, there may be two limitations in the speech samples. First, the naturalness of voice may have been affected by a sentence reading task. Second, the short length of the syllables may not have been able to capture voice quality features that are supposed to be the longest-lasting linguistic components of an accent. Therefore, continuous speech samples are necessary to further examine the voice quality features of HKC and GZC.

1.3. Long-term average spectrum

As for acoustic measurements of continuous speech, [4] suggests that long-term average spectrum (LTAS) is a productive indicator of voice quality because voice quality is indeed an average of the accumulation of traits over time. LTAS is obtained by averaging the spectra over a long duration. It reflects the distribution of sound energy across a range of frequencies. It is assumed that this procedure will eliminate the contribution of segmental features (i.e. linguistic content), and allows the examination of voice quality features that exhibit quasi-permanently in one's voice. For example, one study [6] found that an amplitude peak between 3k to 4k Hz is correlated to a more resonant 'good speaking voice' in actors. There are also some measures that help to further characterize LTAS. Mean spectral energy, i.e., the average amplitude in a spectrum, has been used to quantify glottal tension [7, 8]. Spectral tilt, the ratio of energy between 0-1kHz and 1-5kHz, has been proposed to measure glottal noise [7, 9, 10]. Similarly, spectral slope, i.e., the ratio of energy between 1-5kHz and 5-8kHz, has also been used to quantify glottal noise [11, 12]. Finally, L1-L0 is a measure of the difference between f0 and F1 regions (i.e., 0-300Hz and 3-800 Hz). This measure has been used to examine the distinction between modes of phonation, as it indicates the degree of glottal adduction [11, 13]. L1-L0 has been found to correlate with strained and breathy voices in perception [14].

1.4. The current study

The purpose of this study is to continue the exploration started by [5] on the voice quality features that discern HKC from GZC using LTAS techniques. Sets of longer and more natural speech samples produced by a larger number of talkers are adopted. This study focuses on investigating whether the perceptual descriptions made by laymen can be supported by acoustic measurements. In particular, the following three aspects of voice quality features of the two Cantonese accents will be examined: acoustic resonance, glottal tension, and glottal noise.

METHODOLOGY

2.1. Talkers

We recruited 20 HKC (F= 10) and 15 GZC (F= 8) talkers to provide speech samples. The talkers' age ranged from 23 to 35 years (HKC: mean= 27.1, s.d.= 3.4; GZC: mean= 25.3, s.d.= 2.4). We recruited talkers of this age group because younger talkers of HKC and GZC maybe too similar in their accents due to frequent contact between the cities in recent years.

The HKC talkers and their parents were all born and raised in Hong Kong. All talkers speak HKC in their daily lives and do not use any other Chinese dialects to communicate with their family and friends. They also learnt English as a second language and use English occasionally in school and/or work context.

The GZC talkers were all born and raised in Guangzhou city, China. In addition, they all grew up in the four 'old' districts of Guangzhou (i.e., Dongshan, Yuexiu, Liwan, and Haizhu). The parents of GZC talkers were also born and raised in Guangzhou city. All talkers speak GZC on a daily basis and do not use any other Chinese dialects to communicate with their family and friends. They also learnt Mandarin Chinese as a second language and used Mandarin Chinese occasionally in their education and/or work context.

All talkers reported no history of speech, language nor hearing disorder. They also reported no history of smoking, heaving drinking, laryngeal health issues nor vocal training. All talkers provided consent to participate in the study.

2.2. Production experiment

The production experiment was carried out in a sound-proof booth at the Hong Kong Polytechnic University. Speech samples were recorded through a Telefunken M80 dynamic microphone at the sampling rate of 44.1 kHz. The full production experiment included four tasks: sustained phonation, sentence reading, text reading, and free speech. The current study reported only the data from the text reading and free speech tasks.

In the text reading task, talkers were instructed to read aloud a short passage three times. The passage describes a novel product from an online shopping portal 'Taobao'. It contains 256 Chinese character, which allowed us to extract 30 to 40 seconds of voiced speech in the LTAS analysis detailed below. The speech samples of the second trial were analyzed.

In the free speech task, talkers were asked to elaborate on topics related to their daily lives. The topics included 'Yum Cha' (Cantonese dim-sum food culture), favourite pop culture, experiences on public transport, and about their family and friends. They were provided with question prompts that assisted them in narration. The speech samples of the first topic 'Yum Cha' were analysed and reported in this study. All talkers were able to talk continuously for 90 seconds, which allowed us to extract at least 40 seconds of voiced speech for subsequent LTAS analysis.

2.4. Data Analysis

Data analyses were conducted through Praat [15]. The voiced portions (including all Cantonese vowels and the approximants /j w l/) were first annotated and extracted. The pitch-corrected LTAS were then obtained at intervals of 100 Hz, up to 8 kHz. The LTAS were then normalized with all amplitudes expressed as negative values relative to the peak.

A total of five parameters were used to measure the LTAS, which include mean spectral energy (MSE), spectral tilt (ST), spectral slope (SS), speaker's formant (SPF), and L1-L0. MSE was measured as the average frequency across the entire LTAS (0-8 kHz). A higher (less negative) MSE is associated with more laryngeal tension. ST was measured as the ratio of energy between 0-1 kHz and between 1-5 kHz. SS was measured as the ratio of energy between 1-5 kHz and between 5-8 kHz. Higher ST and SS are associated with less aspiration noise, hence perceived less breathy. SPF was measured as the energy difference between 0-1 kHz and 3-4 kHz. A lower SPF suggests more resonant speech. L1-L0 is the level differences between the F1 region (300-800Hz) and F0 region (0-300 Hz). A high L1-L0 indicates high glottal adduction, whereas low L1-L0 indicates low glottal adduction.

3. RESULTS

3.1. Overall LTAS

To examine whether HKC and GZC differ in energy distribution in LTAS, we fitted mixed effect regression models to the amplitude values at each frequency bin, controlling for the effects of gender and task. All models included accent, gender, and task as the fixed effects and a random intercept for talker. Before examining the differences between accents, we first report the task effects as the continuous speech samples were collected from two different tasks. Overall, there were task effects only in the 500 and 600 Hz frequency bins across the entire LTAS ($p<0.05$). This suggests that there were minimal differences in voice quality features across the two continuous speech tasks.

As for differences between HKC and GZC, we found significant effects of accent on the amplitude values in the frequency bins of 6-700 Hz, 16-1900 Hz, and 55-6600Hz (all $p<0.05$). For all these frequency bins, HKC had higher amplitudes than GZC. Figure 1 illustrates the mean LTAS for HKC and GZC.

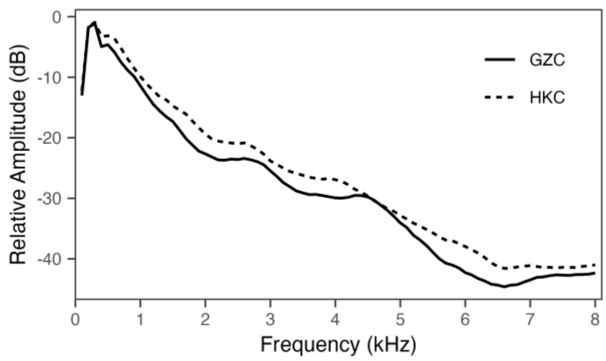

Fig 1. LTAS display of HKC and GZC.

In addition to the energy differences in the various frequency bins reported above, we also observed that GZC had a peak of energy around 4500 Hz, compared to the neighbouring frequencies. In HKC, the 4500 Hz region was rather flat. Although in terms of amplitude values the two accents were quite close, this peak at the F4 region may contribute to the differences in perceptual quality of resonant speech in the two accents of Cantonese. We will return to this in the discussion section below.

3.2. LTAS parameters

We also fitted mixed effect models to the LTAS parameters to determine the effect of accent, after gender and task effects were being controlled for. The models included a random intercept for talker. Due to the exploratory nature of our study, the significant and trending main effects ($p<0.10$) are also reported. Overall, we found a trending main effect of accent on MSE ($p=0.076$), ST ($p=0.086$) and L1-L0 ($p=0.099$). Means and standard deviation for each parameter are provided in Table 1. Overall, HKC has higher MSE, lower ST, and higher L1-L0 than GZC.

Table 1. Means and standard deviations (in parentheses) of the LTAS parameters.

	GZC	HKC
Mean Spectral Energy (MSE, in dB)	-29.3 (3.55)	-27 (3.78)
Spectral Tilt (ST)	0.0663 (0.0139)	0.0587 (0.0155)
Spectral Slope (SS)	0.808 (0.108)	0.783 (0.133)
Speaker's Formant (SPF, in dB)	25.7 (5.21)	22.9 (5.43)
L1-L0 (in dB)	-3.67 (2.81)	-2.09 (2.76)

4. DISCUSSION

In this study, we analyzed the continuous speech samples of HKC and GZC talkers using the LTAS procedure. We will discuss our findings with respect to the three aspects of voice quality: acoustic resonance, glottal tension, and glottal noise.

First, we found higher energy in HKC at the 6-700 Hz (F1 region), 16-1900 Hz (F2 region), and 55-6600Hz regions. In addition, we found that there was a peak at around 4500 Hz (F4 region) in GZC, but not in HKC. This feature, however, was not able to be captured by our statistical analysis. As Wu found no differences in the vowel quality of HKC and GZC, it is assumed that these acoustic features are attributed to articulatory settings in the lower vocal tract, such as larynx height and vocal folds vibratory patterns.

For glottal tension, we found that HKC had higher MSE than GZC. This replicates those in [5] where she found HKC producing tenser phonation than GZC. A higher MSE suggests that HKC has more tension in their voice quality settings than GZC. Such tension may come from adductive tension in the vocal folds, or an overall increase in tightness in the lower vocal tract. However, an acoustic analysis would not allow use to point out which set of muscles give rise to this tension. On the other hand, studies have also reported the interactions between resonance and tension: [7] suggested that a tenser phonation may be related to the use of 'posterior resonance' in bilingual Cantonese-English talkers when speaking HKC. The study by [16] also observed that increased tension is negatively correlated to the perception of resonance. Taken together, HKC talkers may be raising their larynx, which lead to constriction in the lower vocal tract [17]. This constriction increases the overall laryngeal tension and give rise to increased spectral energy in LTAS. Such articulatory settings may cause HKC to be perceived as less resonant than GZC.

As for glottal noise, we found that HKC had higher L1-L0. A lower L1-L0 value suggests more energy in f0 region compared to F1 region. In previous studies [13, 14], low L1-L0 was found to associate with low glottal adduction and the perception of breathy voice. Therefore, the lower L1-L0 suggests that the vocal fold vibratory pattern of GZC is similar to that of lax-breathy voice.

Furthermore, we also found that HKC had lower ST than GZC. As a lower ST is associated with increased aspiration noise and the perception of breathy voice [9, 10], this finding may contradict to the other findings in the current study, which point towards GZC being breathier. Interestingly, this seemingly contradictory result is also attested in [5] and [7]. In the study by [7], LTAS was used to examine the voice quality differences when HKC talkers spoke Cantonese versus English. The findings also showed that HKC talkers exhibited more glottal tension (as indicated by MSE) and more glottal noise (as indicated by ST) when they spoke Cantonese, compared to English. In the study by [5], the various acoustic parameters suggested that HKC was tenser than GZC, except H1*-H2*, which suggest the reverse. In the current study, the increase in glottal tension and noise in HKC maybe a result of the raised larynx posture proposed earlier. According to the Laryngeal Articulator Model [17], raising the larynx would induce constriction in the larynx, narrowing and shortening the epilaryngeal tube, giving rise to turbulent noise. It should be noted that phonation is a complex mechanism, and the interactions of glottal tension and glottal noise are not straightforward. Further studies are required to examine the articulatory mechanism of HKC and GZC. Perceptual effects of these acoustic features are also warranted, particularly for the perception of resonance, as it is traditionally associated with vocal health but not the perception of accent characteristics.

In the last section, we explore how the acoustic findings can relate to the laymen's perceptual descriptions. Overall, our results suggest that HKC exhibits higher glottal tension, and more glottal noise than GZC. Recall that HKC was described as 'light, soft and crispy'. 'Crispy' may refer to the tenser muscular settings, whereas 'light, soft' may be related to increased noise and the perception of less resonant speech, presumably due to raised larynx. Normally, a lowered larynx is considered more relaxed and is the optimal larynx height for speech [18].

On the other hand, GZC was described as "blunt, heavy, low, more cadence, giving people a 'substantial' feeling." The impression of 'heavy and low' may come from a rather lax muscular settings and breathier voice in GZC, compared to HKC. The 'substantial' feeling may be indicating that GZC is more resonant than HKC, presumably due to its lower larynx height.

To conclude, our acoustic analysis provided some evidence to support the voice quality differences between the two accents and the perceptual descriptions by native talkers. Our findings are promising amid the scarcity of studies in the relationships between perception, articulation, and acoustics of voice quality in different accents. They should be followed up by perception experiments to investigate whether native talkers are able to discern the two accents using the acoustic cues of resonance, glottal tension, and glottal noise.

5. ACKNOWLEDGEMENT

This study was supported by an internal competitive research grant from the Hong Kong Polytechnic University offered to the second author.

6. REFERENCES

[1] Wu, W. L. 2006. A comparative analysis of the phonetics of Hong Kong Cantonese and Guangzhou Cantonese. M.Phil. thesis, University of Hong Kong, Hong Kong SAR.

[2] Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh University Press.

[3] Laver, J. D. M. 1968. Voice quality and indexical information. *British Journal of Disorders of Communication*, *3*(1), 43–54.

[4] Esling, J. H., Edmonson, J. A. 2011. Acoustical analysis of voice quality for sociophonetic purposes. In: di Paolo, M., Yaeger-Dror, M. (eds), *Sociophonetics: A student's guide*. Routledge, 131-148.

[5] Fung, R. S. 2015. Voice quality: A preliminary study on the phonetic distinctions of two Cantonese accents. *Proc. 18th ICPhS* Glasgow.

[6] Leino, T., Laukkanen, A. M, Radolf, V. 2011. Formation of the actor's/speaker's formant: A study applying spectrum analysis and computer modeling. *J Voice* 25(2), 150-158.

[7] Ng, M. L., Chen, Y., Chan, E. Y. 2012. Differences in vocal characteristics between Cantonese and English produced by proficient Cantonese-English bilingual speakers- a long-term average spectral analysis. *J Voice* 26(4), 171-176.

[8] Fuller, B. F., Horii, Y. 1988. Spectral energy distribution in four types of infant vocalizations. *Journal of Communication Disorders*, *21*(3), 251-261.

[9] Mendoza, E., Valencia, N., Muñoz, J. and Trujillo, H., 1996. Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *J Voice* 10(1), 59-66.

[10] Hillenbrand, J., Houde, R. A. 1996. Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *J Speech Lang Hear Res* 39(2), 311-321.

[11] Guzman, M., Correa, S., Munoz, D.,Mayerhoff, R., 2013. Influence on spectral energy distribution of emotional expression. *J Voice* 27(1).

[12] Peters, J. 2018. Regional Variation of Voice Quality: A pilot study of High and Low German in the Bersenbrücker Land. *Zeitschrift für Dialektologie und Linguistik*. 1-34.

[13] Kitzing, P. 1986. LTAS criteria pertinent to the measurement of voice quality. *J Phon* 14(3-4), 477-482.

[14] Antonetti, A. E. D. S., Siqueira, L.T.D., Gobbo, M.P.D.A., Brasolotto, A.G., Silverio, K.C.A. 2020. Relationship of cepstral peak prominence-smoothed and long-term average spectrum with auditory–perceptual analysis. *Applied Sciences*, 10(23).

[15] Boersma, P., Weenink, D. Praat: doing phonetics by computer [Computer program]. Version 6.2.14. Accessed: May 2022. [Online]. Available: https://www.praat.org

[16] Master, S., De Biase, N., Chiari, B. M., Laukkanen, A.M. 2008. Acoustic and perceptual analyses of Brazilian male actors' and nonactors' voices: long-term average spectrum and the "actor's formant". *J Voice 22*(2), 146-154.

[17] Esling, J. H., Moisik, S. R., Benner, A., Crevier-Buchman, L. 2019. *Voice quality: The laryngeal articulator model.* Cambridge University Press.

[18] Elliot, N., Sundberg, J., Gramming, P. 1997. Physiological aspects of a vocal exercise. *J Voice* 11(2), 171-177.