Voice Quality and Speech Prosody: An Exploration of the Phonological Distinctions of Two Accents of Cantonese

Roxana Fung
The Hong Kong Polytechnic University
Background

- Cantonese is the standard variety of the Yue dialects.

- There are two prestige varieties of Cantonese: Guangzhou Cantonese (GZC) and Hong Kong Cantonese (HKC).

- They share almost the same phonological system and are fully mutually intelligible.
## Background
- Cantonese Consonants

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labio-Dental</th>
<th>Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Labial-Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>[p]</td>
<td>[t]</td>
<td></td>
<td>[k]</td>
<td>[kʰ]</td>
<td></td>
<td>[kʰʷ]</td>
</tr>
<tr>
<td></td>
<td>[pʰ]</td>
<td></td>
<td>[tʰ]</td>
<td>[kʰ]</td>
<td>[kʰʷ]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td>[f]</td>
<td></td>
<td>[s][ʃ]</td>
<td></td>
<td></td>
<td></td>
<td>[h]</td>
</tr>
<tr>
<td>Affricate</td>
<td>[ʦ][ʃ]</td>
<td></td>
<td>[ʦʰ][ʃʰ]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>[m]</td>
<td></td>
<td>[n]</td>
<td></td>
<td>[n]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximate</td>
<td>[l]</td>
<td></td>
<td>[j]</td>
<td></td>
<td></td>
<td>[w]</td>
<td></td>
</tr>
</tbody>
</table>
Background
-Cantonese Vowels

[ i ] [ y ] [ u ]

[ e ] [ θ ] [ o ]

[ ε ] [ œ ] [ ø ]

[ a ]
### Cantonese Tones

<table>
<thead>
<tr>
<th>Tone number</th>
<th>Contour description</th>
<th>Pitch value</th>
<th>Example</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>High level</td>
<td>55</td>
<td>[jen]</td>
<td>‘because of’</td>
</tr>
<tr>
<td>T2</td>
<td>High rising</td>
<td>35</td>
<td>[jen]</td>
<td>‘endure’</td>
</tr>
<tr>
<td>T3</td>
<td>Mid level</td>
<td>33</td>
<td>[jen]</td>
<td>‘mark’</td>
</tr>
<tr>
<td>T4</td>
<td>Extra-low level</td>
<td>11/21</td>
<td>[jen]</td>
<td>‘people’</td>
</tr>
<tr>
<td>T5</td>
<td>Low rising</td>
<td>23</td>
<td>[jen]</td>
<td>‘attract to’</td>
</tr>
<tr>
<td>T6</td>
<td>Low level</td>
<td>22</td>
<td>[jen]</td>
<td>‘blade’</td>
</tr>
</tbody>
</table>
GZC and HKC differ mainly in their vocabularies due to the different socio-political systems of the two cities.

The only phonological distinction between them that has been frequently mentioned: For T1 (high-level tone) in Cantonese, a high-falling contour may be produced as a free variant among some older GZC speakers but not among HKC speakers.

Nonetheless, most native speakers of two accents still have no difficult in distinguishing the two accents without the hint of the high-falling tone.

The question of what phonetic features help to identify these two accents has perplexed many minds.
GZC vs HKC

- Sample 1 (‘I am now reading the character [fu]’)
- Sample 2 (‘I am now reading the character is [fu]’)
- Sample 3 (‘If who can make that guy take off his coat ……’)
- Sample 4 (‘If who can make that guy take off his coat ……’)
To date, Wu (2006) has been the only work devoted to the investigation of the phonetic differences of the two accents.

Participants: 20 undergraduate students (10 HKC speakers + 10 GZC speakers).

Methodology:
- The participants were asked to read out 88 words (from monosyllabic to tri-syllabic) and 15 sentence/short exchanges which covered all the consonants, vowels and tones in Cantonese.
- Acoustic analyses were performed on the segments collected.

Conclusion:
- More phonetic similarities than differences between GZC and HKC were found.
The Current Pilot Study

Objective:
- To explore the phonetic features that can account for the identification of GZC and HKC.

Data:
- Based on the data collected from the production task of two previous studies on tone mergers in the two accents.
- 75 GZC speakers and 120 HKC in speakers in three age groups were randomly selected.
  - Senior: 50-58
  - Middle-aged: 35-45
  - Young: 20-25
- Some CV(C) syllables were embedded in two sentence carriers for the participants to read aloud.
Impressionistic Remarks on the Differences of the Two Accents Found on the Web

- HKC is not accurately or fully articulated.

- The intonation of HKC tends to be soft, light, crispy and uplifting; whereas GZC tends to be heavy, low and has a ‘substantial’ feeling.

- HKC has a higher pitch but flatter intonation than GZC.

- HKC has a faster tempo than GZC.
Possible Acoustic Correlates of the Impressionistic Remarks

- Fully Articulated – Vowel space
- Substantial feelings – Voice quality
- Pitch – Key and Register
- Tempo – Rhythm
The claim that HKC is not fully articulated may be caused by many factors, such as co-articulation, syllable deletion or sound change.

Our exploration started with the study of the dispersion of vowels.

Vowel space refers to the area of a speaker’s acoustic vowel space formed by the frequencies of the first formant (F1) and the second formant (F2). That is the area that a speaker actually uses in producing vowel distinctions.

A number of studies have shown that there is systematic variation in vowel space size across dialects in languages such as American English and Dutch. (e.g. Clopper, Cynthia G., David B. Pisoni, and Kenneth de Jong 2005).
Vowel Space

- Methodology

- Participants: 24 (2 accents x 3 age groups x 2 gender x 2 speakers)

- Linguistic Materials:
  - The six tonal contrasts of the 2 CV roots [ji] and [fu] in sentence-internal and final positions.
  - The six tokens of T1 of the syllable [ka] in sentence-internal position.
  - A total of 1008 syllables (42 syllables x 24 participants) were collected.
Vowel Space
-Methodology

- Procedure:
  - The mid-point value of the formants F1, F2 and F3 of each vowel were measured by Praat.
  - Normalized the mean values of [i], [u] and [a] vowels with the Bark-difference technique, a vowel-intrinsic method suggested by Thomas (2011):
  - Converted F1 F2 and F3 to Bark-valued Z1, Z2 and Z3 respectively.
    - $Z_i = \frac{26.81}{(1+1960/F_i)} - 0.53$
    - $F_i$ is the value of a given formant
  - Using $Z_3-Z_1$ as vowel height, and $Z_3-Z_2$ as advancement.
  - Calculated the triangular area formed by normalized formant values of [i], [u] and [a].
Vowel Space
- Results

<table>
<thead>
<tr>
<th></th>
<th>GZC</th>
<th>HKC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>15.00 (8.17)</td>
<td>18.04 (10.10)</td>
<td>18.58 (5.22)</td>
</tr>
<tr>
<td></td>
<td>16.52 (9.74)</td>
<td>22.26 (7.14)</td>
<td></td>
</tr>
</tbody>
</table>

- The vowel space of HKC was greater than GZC.
- The difference of mean between city and gender were significant, $p \leq 0.00$. 
The vowels of HKC tended to be higher than that of GZC.

The back vowel [u] was more backward in HKC than in GZC.

Conclusion: Contradict to the impressionistic remark, the vowels in HKC were more dispersed than that of GZC.
Voice Quality

- Background

- The sense of ‘heaviness’ or ‘softness’ may be related to phonation, the process by which the vocal folds produce quasi-periodic vibration.

- ‘Voice quality is regarded as a mysterious domain by language variationists, but largely the purview of speech pathologists.’ Thomas (2011: 224)

- Nevertheless, there have been a few sociolinguistic investigations into voice quality in Scotland dialects (Esling 2000), New Zealand English (Szakay 2012), and Asian American English (Newman and Wu 2011).
Voice Quality

- Phonation types

- Glottis: most open — most closed
  phonation: breathy — modal — creaky

- Modal Voice:
  - Considered to be the ‘normal’ phonation.

- Creaky Voice:
  - Produced with constricted glottis and more overall muscular tension of the entire vocal tract.
  - More tensed voice quality.

- Breathy Voice:
  - Produced with open glottis with low muscular tension.
  - Allow more air to pass through the glottis and creates turbulence not present with modal phonation.
  - Can be characterized as a ‘bedroom voice’.
Voice Quality
-Spectral Analysis

- Creaky voice pulse:
  - A sharp quick closure of glottis.
  - Have a great deal of energy in the higher formants, such as F1(=A1).

- Breathy voice pulse:
  - A less sharply closure.
  - Have much of the energy in the F0(=H1), and is composed semi-random noise produced by the turbulent glottal airflow.

- Modal voice pulse:
  - Closing phase is shorter than the opening, but two are nearly the same.
  - H2 is slightly lower than H1.
Voice Quality
-Acoustic Measures

- **H1-H2**
  - The most popular measure.
  - The difference in amplitude between the first harmonic (H1) and the second harmonic (H2).
  - Expected to be large and positive for breathy voices and small and/or negative for creaky voice.

- **H1-An (H1-A1, H1-A2, H1-A3)**
  - The difference in amplitude between H1 and the strongest harmonic in the nth formant (An).
  - Expected to be large and positive for breathy voices and small and/or negative for creaky voices.
Voice Quality
-Acoustic Measures

- CPP
  - The difference in amplitude between the cepstral peak and the value of regression line at the cepstral peak.
  - An indicator of harmonics-to-noise ratio. Noise excitation is an important component of breathy voice.
  - Expected to be small for breathy voices.

- H2-H4
  - Not a commonly used measure.
  - The difference in amplitude between H2 and H4.
  - High H2-H4 seems to be associated with a less stiff vocal fold body-layer and/or breathy voice.
  - Low H2-H4 with a stiffer body-layer and/or falsetto phonation.
Voice Quality - Acoustic Measures

- Many linguists just focus on H1-H2.

- There are different ways to produce breathy vs creaky voice. Speakers may differ in the acoustic details of their contrast.

- Different measures reflect different aspects of production, and may or may not distinguish phonation category.

- Men and women may produce the phonation contrast differently. Hence, different gender may need different measures.

- Different languages of known phonation type are distinguished by different measures.
Voice Quality
-Methodology

- **Materials:**
  - A total of 1584 syllables (66 syllables x 24 participants) were adopted for analysis.
  - The six contrastive tones of [ji], [si], [fu] in two sentential positions = 3 CV roots x 6 tones x 2 positions = 36 syllables

- **Procedures:**
  - Acoustic measures were made semi-automatically over the mid-portion of the entire vowel duration with VoiceSauce (Keating and Vicenik 2009).
  - The six measures (H1*-H2*, H2*-H4*, H1*-A1*, H1*-A2*, H1*-A3*, CPP) were calculated.
  - Asterisks indicate that the harmonic amplitude were corrected to recover the source spectrum of the vocal fold pulses by reducing the influence of formant resonances.
Voice quality - Results

<table>
<thead>
<tr>
<th></th>
<th>H1*-H2*</th>
<th>H2*-H4*</th>
<th>H1*-A1*</th>
<th>H1*-A2*</th>
<th>H1*-A3*</th>
<th>CPP</th>
</tr>
</thead>
<tbody>
<tr>
<td>GZC</td>
<td>4.47 (5.27)</td>
<td>5.82 (5.59)</td>
<td>17.41 (9.46)</td>
<td>21.87 (9.30)</td>
<td>14.47 (10.78)</td>
<td>20.92 (2.82)</td>
</tr>
<tr>
<td>HKC</td>
<td>5.30 (4.27)</td>
<td>3.32 (5.34)</td>
<td>14.43 (8.07)</td>
<td>18.17 (8.17)</td>
<td>10.98 (9.70)</td>
<td>22.40 (3.42)</td>
</tr>
</tbody>
</table>

- The differences between city and age were significant.
- All the measures, except H1*-H2*, indicated that GZC was breathier than HKC.
- Keating and Esposito (2006) on Mandarin: H1*-H2* was a measure most clearly related to F0: the low pitch creakier and the pitches breathier.
Voice Quality - Results

- Phonation co-varies with some tones in Cantonese. The extra-low level tone (T4), low rising tone (T5) may be with creaky voice among some speakers.

- In general, our findings explained why GZC sounded more ‘heavy’ than HKC.
**Key and Register**

**-Background**

- The difference in pitch range between HKC and GZC should be related to the key and register, another laryngeal factor of voice quality.

- Key involves the degree of F0 variation in a speaker’s voice, whereas register involves the shifting of the range of F0 values.

- Thomas (2011:226): Many white females seem to use a wider key that African American females or males of either ethnic group in the south of the United States.

- Busa and Urbani (2011) demonstrated that the English produced by Italians have overall higher register and narrower key than those produced by the Americans.

- However, the perceptual salience of key and register is not yet known in tone languages, which employ F0 change to contrast lexical meaning.
Key and Register
- Methodology

- Linguistic materials:
  - All the tonal contrasts of [si], [ji], [fu] were embedded in two sentential positions.
  - A total of 864 syllables (3 cv x 6 tones x 2 positions x 24 participants) were analyzed.

- Procedures:
  - The F0 of each syllable was tracked by a Praat script at 11 time points.
  - The F0 were converted to ERB rate scale according to the formula:
    $$16.7 \times \log_{10} (0.006046 f + 1)$$ where $f$ is the frequency of sound in Hz.
  - The register for each participant was taken as the median of F0 in ERB of 66 time points.
  - The key for each participant was the difference between the F0 of T1 (the highest level) tone and that of T4 (the lowest level) tone at 50% duration.
Our findings support the impressionistic remarks that the register of HKC was higher than that of GZC, especially among the female.
The register of HKC was higher than that of GZC, especially among the female.

The difference was significant for city and gender.

Our findings contradict the impressionistic remark: GZC has a flatter pitch range than HKC.
Speech Rhythm - Background

- In social-interactional research, tempo or speech rate is generally quantified as number of syllables or words per time unit.

- It is well-known that this acoustic measure is significant in language typology:
  - According to speech rhythm, English is a syllable-timed language, Chinese a stressed-time language, and Japanese a *mora*-timed language.
There is a growing literature showing that there can be considerable variation in metrics between dialects of the same language.

Giordani and D’Anna (2010) on 15 regional varieties of Italian: certain varieties resemble mora-timed and stress-timed languages, even though Italian is generally considered to be syllable-timed.

Arabic (Ghazali et al. 2002); British English (Ferragne and Pellegrino 2004); Greek (Tropintzi et al. 2010), Portuguese (Frota and Vigario 2001); Acadian French (Wladyslaw Chichocki et al 2013).
Rhythm
-Methodology

- Participants
  - 1 GZC middle age female + 1 HKC female old age female.

- Linguistic materials
  - The narration of the North Wind and the Sun story.

- Procedure
  - The GZC speaker recorded the story in a spontaneously manner. The content was then transcribed and read by the HKC speaker in a semi-spontaneously manner.
  - The speech was segmented into consonantal and vocalic elements manually.
  - The durations of the segments were measured by a Praat script.
Rhythm - Methodology

- Rhythm metrics (following Grabe and Low 2002, and Ramus et al. 1999) were calculated:
  - %V (percentage of total utterance duration composed of vocalic intervals)
  - ΔV (standard deviation of vocalic interval duration)
  - Varco ΔV (coefficient of variation of vocalic interval duration)
  - ΔC (standard deviation of consonantal interval duration)
  - Varco ΔC (coefficient of variation of consonantal interval duration)
  - nPVI (normalized pairwise variability indices for vocalic intervals - measures the durational contrast between successive vowels in a spoken sentence.)

- Characteristics of stress-timed languages:
  - Higher ΔC
  - Higher Varco ΔC
  - Lower %V
  - Higher Varco ΔV
  - Higher nPVI
More data was required to explore this linguistic domain.
<table>
<thead>
<tr>
<th></th>
<th>Our findings</th>
<th>Impressionistic remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel Space</td>
<td>GZC was smaller than HKC</td>
<td>X</td>
</tr>
<tr>
<td>Voice Quality</td>
<td>GZC was breathier than HKC</td>
<td>V</td>
</tr>
<tr>
<td></td>
<td>HKC used more muscular force in production</td>
<td></td>
</tr>
<tr>
<td>Register</td>
<td>GZC has lower pitch than HKC</td>
<td>X</td>
</tr>
<tr>
<td>Key</td>
<td>GZC has narrower pitch range than HKC</td>
<td>X</td>
</tr>
<tr>
<td>Rhythm</td>
<td>GZC has a lower speech rate than HKC</td>
<td>V</td>
</tr>
</tbody>
</table>
Conclusion

- Voice quality and speech prosody are two promising linguistic domains to explore the phonetic distinction between GZC and HKC.

- After identifying the acoustic correlates of the two accents, perception experiments should be performed to prove if native speakers used the correlates to distinguish the two accents.

Thank you!