Temporal Location of Perceptual Cues for Cantonese Tone Identification

Zoe Wai-Man Lam, Kathleen Currie Hall and Douglas Pulleyblank

Department of Linguistics
University of British Columbia
Outline of the presentation

1. Background
2. Research question
3. Methodology
4. Results
5. Discussion
6. Conclusion
F0 variation in language

Fundamental frequency (F0): the frequency of vocal fold vibration (in Hz)

F0 variation takes place within a **temporal window**.

- Lexical tone: temporal window = duration of a syllable
- Intonation: temporal window = duration of an utterance
Acoustic vs. perceptual

- F0 is an *acoustic* term: the physical properties of the signal
- Pitch is a *perceptual* term: the hearer’s perception of the signal
- Tone is a *linguistic* term: phonological categories to distinguish words (Yip 2002)

- Example: Yoruba has 3 level tones (H M L)
- In a disyllabic noun $\sigma_1\sigma_2$:

<table>
<thead>
<tr>
<th>F0 of $\sigma_2$</th>
<th>Perceived tone of $\sigma_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Falling</td>
<td>Low tone</td>
</tr>
<tr>
<td>Flat</td>
<td>Mid tone</td>
</tr>
</tbody>
</table>

(Hombert 1976)

- Acoustic cues $\neq$ perceptual cues
Phonemic inventory of Cantonese tones

6 lexical tones

- **Tone 1**: High level
- **Tone 2**: High rising
- **Tone 3**: Mid level
- **Tone 4**: Low falling*
- **Tone 5**: Low rising
- **Tone 6**: Low level

3 **level** tones (1, 3, 6)
3 **contour** tones (2, 4, 5)

*Slope of Tone 4 is similar to that of a level tone phonetically*
Previous studies on tonal production

Khouw & Ciocca (2007)

- Focus on syllables produced in isolation
- Goal: investigate which parts of the vocalic segment provide F0 information that is most strongly related to the identification of Cantonese tones
- F0 measurements are made for eight consecutive sections of the whole vocalic segment for each syllable.
- Discriminant analysis shows that the most important correlates of tone identity are in the later part of the vocalic segment, specifically the 6th and 7th sections.
Previous studies on tonal production

Wong (2007)
- Focuses on tonal coarticulatory effects in disyllabic sequences in a carrier sentence
- *lau lau* as target syllables
- Carryover effects are so strong that F0 transition from the tone of $\sigma_1$ to $\sigma_2$ takes up 50% of $\sigma_2$, resulting in a great magnitude of F0 variation across different tonal contexts
Research question

Do native Cantonese speakers rely particularly on the perceptual cues from the latter portion of the tone for identification?

Our hypothesis
Stimulus preparation

- 4 naïve native speakers: 2 male, 2 female; age range 19-28
- Carrier sentence
  \[ \text{jau5 go3 giu3 } \sigma_1 \sigma_2 \text{ ge3 je5} \]
  \[ \text{haveCL call } \underline{\text{__}} \underline{\text{__}} \text{ of thing} \]
  ‘There is something called \underline{\text{__}} \underline{\text{__}}.’
- \( \sigma_1 \sigma_2 \) is any disyllabic combination of the following:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>si</th>
<th>se</th>
<th>fu</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>high level</td>
<td>55</td>
<td>詩 Poetry</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>high rising</td>
<td>25</td>
<td>史 History</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>mid level</td>
<td>33</td>
<td>試 To try*</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>low falling</td>
<td>21</td>
<td>時 Time</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>low rising</td>
<td>23</td>
<td>市 Market*</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>low level</td>
<td>22</td>
<td>是 Right</td>
</tr>
</tbody>
</table>

- 1152 sentence tokens, divided into 8 groups of stimuli
Four stimulus types

Control stimuli

**Type A:** Not manipulated

<table>
<thead>
<tr>
<th>jau5</th>
<th>go3</th>
<th>giu3</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>ge3</th>
<th>je5</th>
</tr>
</thead>
</table>

**Type B:** $\sigma_2$ not affected, but there is some disruption elsewhere

<table>
<thead>
<tr>
<th>jau5</th>
<th>go3</th>
<th>giu3</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>ge3</th>
<th>je5</th>
</tr>
</thead>
</table>
Four stimulus types

**Type C:** 1st half of the tone of $\sigma_2$ zeroed out

```
<table>
<thead>
<tr>
<th>jau5</th>
<th>go3</th>
<th>giu3</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>ge3</th>
<th>je5</th>
</tr>
</thead>
</table>
```

**Type D:** 2nd half of the tone of $\sigma_2$ zeroed out

```
<table>
<thead>
<tr>
<th>jau5</th>
<th>go3</th>
<th>giu3</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>ge3</th>
<th>je5</th>
</tr>
</thead>
</table>
```

Zeroing out done by a Praat script

Boundaries of the vocalic portion of the syllable were marked by the first author manually.
Four predictions

Based on the hypothesis that the second half of the tone contains the crucial perceptual cues for tone identification, we predict that:

1. Type A = Type B
2. Type A = Type C
3. Type A > Type D
4. Type C > Type D
Perception study: Procedures

- Participants were presented with one sentence on the screen in each trial.
- They listened to an audio stimulus at the same time.
- **They were always asked what \( \sigma^2 \) was.**
- They were instructed to respond by pushing 1,2,3,4,5 or 6 on the keyboard.
Perception study: Procedures

Practice set

• 12 trials, using the syllable *jau*, produced by the first author
• All stimuli in the practice set were not manipulated (i.e. belong to Type A).
• Feedback was provided to the participants during the practice period.
• Data of participants who failed to get **at least 9 correct responses** were excluded from analysis.

Actual experiment

• 144 trials per participant, using *si*, *se* or *fu* syllables
• Randomized, not blocked according to stimulus type
• Participants were told that it’s normal if some parts of the recording are not clear. They just needed to try their best to choose an answer.
• Feedback was NOT provided.
Perception study: The participants

- 63 participants from the UBC community participated in the study, who were either paid or given course credit.

- The data of 39 participants were excluded from the analysis.
  - Not a speaker of the Hong Kong variety of Cantonese (e.g. Malaysia)
  - Got fewer than 9 correct answers out of 12 Qs in the practice set

- The data of 24 participants were included in the analysis.
  - Got at least 9 correct answers out of 12 Qs in the practice set
  - Self-rated 6 or 7 out of 7 for their Cantonese proficiency
  - Able to read Traditional Chinese characters, had spent at least 14 years in Hong Kong
  - Subject 107 and 501 have never lived in Hong Kong, but their parents are from Hong Kong; results similar to other participants
  - Age: 18-36
Results (Version 1)

Are these predictions correct?

1. Type A = Type B  Yes
2. Type A = Type C  No
3. Type A > Type D  Yes
4. Type C > Type D  No
Results (Version 2)

Are these predictions correct?

1. Type A = Type B
2. Type A = Type C
3. Type A > Type D
4. Type C > Type D

Yes only for contour tones but not level tones
Type A  Prediction = Type C

Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half
Level tones: Type A (control) vs Type C (1st half removed)

Prediction: Type A = Type C

✗ Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half
Contour tones: Type A (control) vs Type C (1st half removed)

Prediction: Type A = Type C
Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Prediction

Type A > Type D
Level tones: Type A (control) vs Type D (2nd half removed)

Prediction: Type A > Type D  ✗
Contour tones: Type A (control) vs Type D (2nd half removed)

Prediction: Type A > Type D
Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half

Prediction

Type C > Type D
Level tones: Type C (1st half removed) vs Type D (2nd half removed)

Prediction: Type C > Type D
Contour tones: Type C (1\textsuperscript{st} half removed) vs Type D (2\textsuperscript{nd} half removed)

Prediction: Type C > Type D
Data analysis

- A logistic mixed effects model with percentage correct as the dependent variable was fit with fixed effects for Stimulus Type (A, B, C, D), Tone Type (Level, Contour) and their interactions.

- The random effect structure was as maximally specified as possible with random effects for Subject and Talker, and by-Subject random slopes for the interaction of Stimulus Type and Tone Type.
Type C = cannot hear the 1st half
Type D = cannot hear the 2nd half
Discussion

Listener’s bias

When hearing Type C → tend to choose a contour tone
When hearing Type D → tend to choose a level tone

<table>
<thead>
<tr>
<th>Type</th>
<th>% of listeners’ response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>level</td>
</tr>
<tr>
<td>A</td>
<td>50.3</td>
</tr>
<tr>
<td>B</td>
<td>48.3</td>
</tr>
<tr>
<td>C</td>
<td>37.9</td>
</tr>
<tr>
<td>D</td>
<td>53.6</td>
</tr>
</tbody>
</table>

$\sigma^2$

Type A

Type B

Type C

Type D

$p < 0.001$

$p = 0.03$

Type C = cannot hear the $1^{st}$ half
Type D = cannot hear the $2^{nd}$ half
Discussion

Tone 4 patterns with other contour tones despite being phonetically similar to level tones.
Conclusion

Listeners rely on perceptual cues from different parts of the syllable, depending on the tone type.

This half is needed for the identification of level tones.

This half is needed for the identification of contour tones.
Acknowledgements

This project was supported by a research grant from Social Sciences and Humanities Research Council of Canada (F10-04735).

We would like to thank Molly Babel and participants of the research seminar “Attention and Salience” at the University of British Columbia in 2014-2015, who offered invaluable comments throughout the process. All errors are ours.