Normalization of talker variation in the perception of Cantonese level tones: Impact of speech and nonspeech contexts

Caicai Zhang, Gang Peng, and William S-Y. Wang
Language Engineering Laboratory, The Chinese University of Hong Kong
TAL 2012, Nanjing, 2012 May 27
1.1 Background: Contrastive context effect
1.1 Background: Contrastive context effect

The contrastive context effect refers to the phenomenon where a particular tone or sound is perceived differently based on its context within a sentence or phrase. In the diagram, the mid level tone is indicated by the red line, with the specific tones represented as `ʦʰiŋ25 lei23 tuk22 ji33 lei21 tʰiŋ55 ha23`. The frequency (Hz) and pitch (Hz) are also shown on the graph, highlighting the linguistic and acoustic properties of the sounds.
1.1 Background: Contrastive context effect

tʰiŋ25 lei23 tuk22 ji33 lei21 tʰiŋ55 ha23

High level tone
1.1 Background: Contrastive context effect

\[ ts'^h_iŋ25 \, lei23 \, tuk22 \quad ji33 \quad le^i21 \, t'^h_iŋ55 \, ha23 \]

Low level tone
1.1 Background: Talker normalization

Adapted from Ryalls and Pisoni (1997)
1.1 Background: Talker normalization

- Pitch range assessment (Wong and Diehl, 2003)
1.1 Background: Talker normalization

- Pitch range assessment (Wong and Diehl, 2003)

Mid level tone
1.2 Research question

- Is context-dependent talker normalization a *speech-specific* process or *general auditory* process?
  - Speech vs. Nonspeech context

- Conflicting findings in the literature:
  - Qualitatively *similar effects* of speech and nonspeech contexts on tone perception in *Mandarin* (Huang and Holt, 2009; 2011);
  - *[ə]* sound context generated with the neutral vocal tract had *no effect* on tone perception in *Cantonese* (Francis et al., 2006).
2. Experimental design

- Talker variability:
  - **Female High** talker: 180-350 Hz
  - **Female Low** talker: 180-280 Hz
  - **Male High** talker: 110-190 Hz
  - **Male Low** talker: 80-130 Hz

- 請你讀意嚟聽下。
  
  [tsʰiŋ25 lei23 tuk22 ji33 lei21 tʰiŋ55 ha23]
  
  ‘Please read /ji33/ for me.’
2. Experimental design

- Factorial design: Context x F0 shift
  - Speech and Nonspeech (triangle wave)
  - Raised (2 semitones), unshifted and lowered (3 semitones).
2. Experimental design

- Factorial design: Context x F0 shift
  - Speech and Nonspeech (triangle wave)
  - Raised (2 semitones), unshifted and lowered (3 semitones).

Low level tone  Mid level tone  High level tone

Low level tone  Mid level tone  High level tone
2. Experimental design

• Blocked design
  • (1) Isolated target words (/ji33/ produced by 4 talkers);
  • (2) Nonspeech context condition;
  • (3) Speech context condition.

• Task: Three-alternative forced-choice identification
  • 医 /ji55/ ‘a doctor’ (high level tone);
  • 意 /ji33/ ‘meaning’ (mid level tone);
  • 二 /ji22/ ‘the second’ (low level tone);

• Subjects: 16 native speakers of HK Cantonese (8F, 8M)
3. Results

- Unshifted F0 condition: *Mid level tone expected*

**Results**

**•** p<0.01

**••** 0<0.001
3. Results

- Lowered F0 condition: *High level tone expected*

![Graph showing identification rate of T1 for different talkers and context conditions](image)

**Notes:**
- ****: p<0.01
- *******: 0<0.001
3. Results

- Raised F0 condition: *Low level tone* expected

**Results**

**p > 0.05**: $p < 0.01$

**p > 0.05**: $p < 0.001$

**$p < 0.01$**

**$p < 0.001$**
4. Discussion

• (1) Unequal effects of speech and nonspeech contexts:
  • Only speech contexts effectively facilitate the normalization of talker variability in tone perception, whereas nonspeech contexts show no obvious effect (Francis et al., 2006).

• Our findings are congruent with the speech-specific mechanism, but not with the general perceptual mechanism.

• Why?
  • Speech contexts are human vocalizations, which allow listeners to map a particular talker’s phonetic space.
4. Discussion

- (2) Talker-specific normalization patterns:
  - Some talker (such as FL) can be normalized well without contextual cues.

- Listeners’ expectation of population-average pitch ranges (Peng et al. 2012).
  - FH talker: 180-350 Hz
  - FL talker: 180-280 Hz
  - MH talker: 110-190 Hz
  - ML talker: 80-130 Hz

- **Contextual cues** and the expected pitch ranges co-contribute to talker normalization in an interactive way:
  - Contextual cues enable listeners to tune to a particular talker’s pitch range, reducing perceptual bias when a talker’s pitch range is far away from the population-average.
4. Discussion

• (3) Why do our findings differ from that of Huang and Holt (2009, 2011)?

  • We studied **level tones**, which are ambiguous without contextual cues;
  • Huang and Holt examined **contour tones**, which are less ambiguous; therefore, contextual cues may have a smaller effect.

• It is likely that contour tones constrained the effect of **speech contexts** more than it did to the **nonspeech contexts**, thereby equalizing the effects of speech and nonspeech contexts in Huang and Holt (2009, 2011).
Acknowledgements

• Research Grant Council of Hong Kong (GRF: 455911)
• National Science Foundation of China 11074267 and 61135003
• 973 Grant from the National Basic Research Program of the Ministry of Science and Technology of China (2012CB720700)

• We thank all the LELers:
  Dr. James W. MINETT
  Dr. Hong-Ying ZHENG
  Mr. Manson FONG
  Mr. Guangting MAI
  Ms. Ruijing WANG
  Ms. Ruoxiao YANG
  Ms. Lin ZHOU
  Mr. Yifan ZOU

Thank you!
References


• Picture of train station: