



# TALKER NORMALISATION OF PROSODIC CUES IN NON-NATIVE SPEAKERS

Bei Li, Bruce Xiao Wang, Si Chen, Caicai Zhang,  
Puiyin Lau & Yike Yang

The Hong Kong Polytechnic University

---

## Background

### **Speech Normalization**

- *Lack of invariance* problem (Liberman, 1967)
- Perceptual normalization provides relational invariance (e.g., Bauer & Benedict, 1997; Stilp & Theodore, 2020)

### **Theories on Speech Normalization**

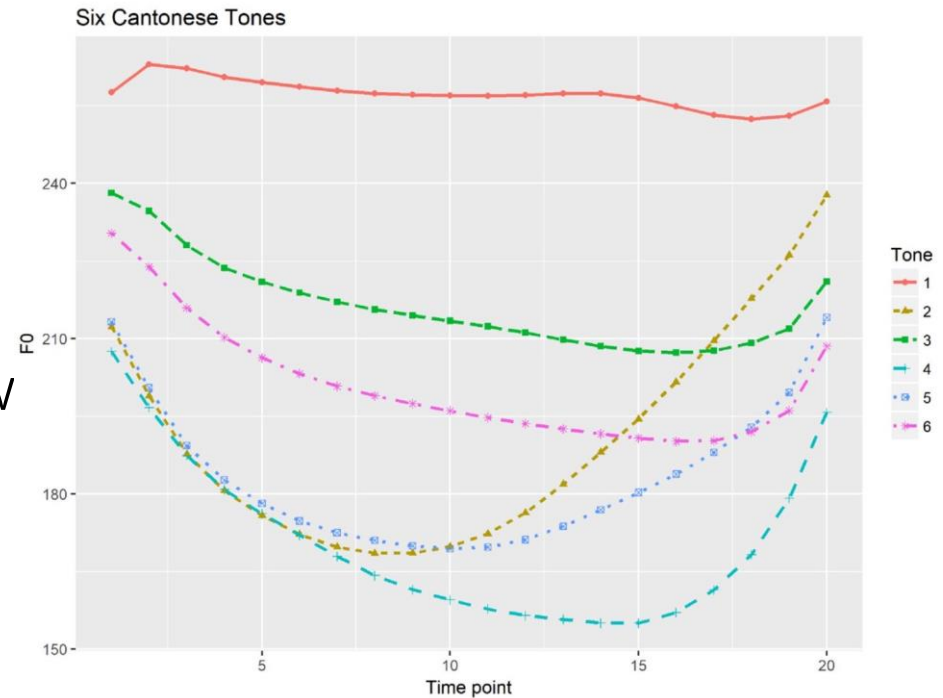
- Abstract theories of phonological representations (Hyman, 1970; Idemaru & Holt, 2011)
- Episodic theories of phonological representations (Wang, 1976; Wade & Möbius, 2010)
- Parametric Representations (Pierrehumbert, 2016)

## Cantonese Tonal System

- Three level tones: High tone T1(55/53), mid tone T3(33) and low tone T6(22)

## Context Effects

- Cantonese listeners may identify the mid tone as a low tone by shifting the  $F_0$  of the preceding context two semitones upward (Wong & Diehl, 2003; Francis et al., 2015).



Chen et al. (2022)

## **Probabilistic Parametric Representation (PPR)**

(Chen et al., 2022)

- Mental representations in isolation
- Mental representations *vs.* Contextual information

Is the phenomenon of PPR exclusive to native speakers? If non-native speakers adopt a similar approach for normalizing Cantonese level tones?

---

## Research Questions

1. Can Mandarin speakers establish and exploit the mental representations of  $F_0$  distributions to identify Cantonese level tones?
2. Can Mandarin speakers still detect such representations when the parameters are deviated from the estimated population distributions?
3. Does the contextual information affect the tonal normalization of Cantonese level tones in Mandarin speakers?

## Method

- Identification task
- Participants: Native Mandarin speakers
- Materials: The target /ji/ with Cantonese level tones; and /si/ as fillers

### Experiment 1

Natural Tone  
Identification  
in Isolation

### Experiment 2

Synthesized  
Tone  
Identification  
in Isolation

### Experiment 3

Tone  
Identification  
in Contexts

## Participants

- 14 native Mandarin speakers (7F; 7M; Mean age  $\pm$  SD: 19.29  $\pm$  0.91 yrs)

## Stimuli

- 34 native Hong Kong Cantonese speakers
- Target syllable /ji/ & filler /si/
- Three level tones (T1, T3, T6) → “聽聽 /thiŋ thiŋ/ (Listen to) \_\_\_\_\_”
- Extracted syllables (T3) in isolation (34 speakers \* 2 syllables \* 10 repetitions)

## Experiment 1

### Natural Tone Identification in Isolation

# Method

- Procedure



Isolated syllable

- blocked by gender



Identify

醫 “to cure”  
(T1: 55)  
**Press “1”**



意 “meaning”  
(T3: 33)  
**Press “3”**



二 “two”  
(T6: 22)  
**Press “6”**





## Distribution fitting

Extracted  $F_0$   
values of the  
voiced portion  
of /ji/

Fitted **skew-normal (SN) distributions**  
to each of the 34 speakers' production and  
obtained three parameters:

**location ( $\xi$ ), scale ( $\omega$ ), and shape ( $\alpha$ )**

## Multinomial mixed effects model

**SN distributions ~ Judgments (T1, T3, T6)**

*mixcat* (Hartzel et al., 2001) package in R

# Results

## Multinomial mixed effects model

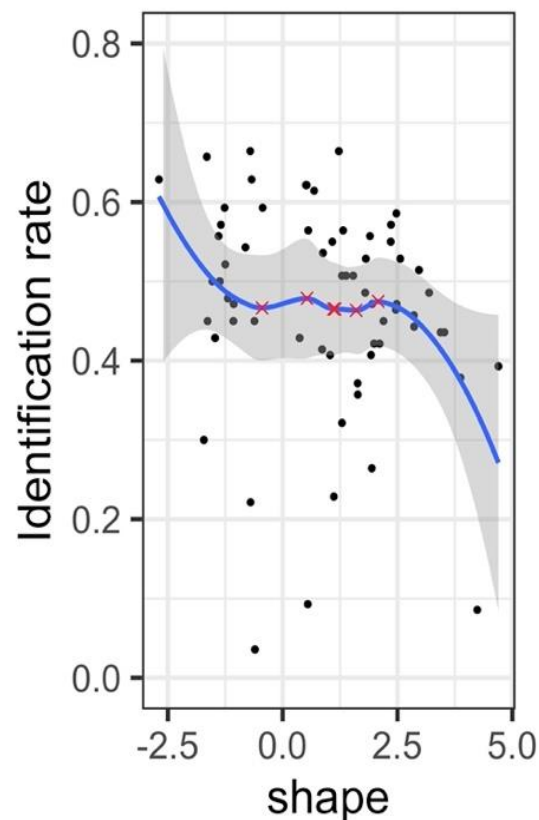
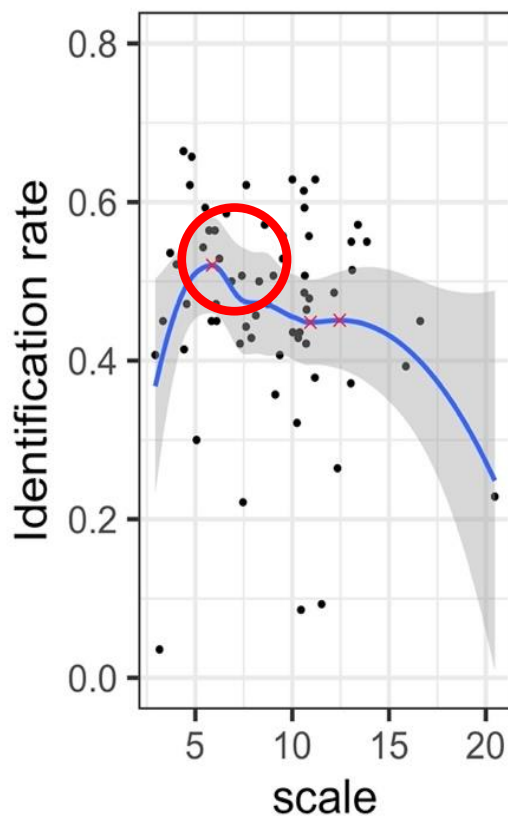
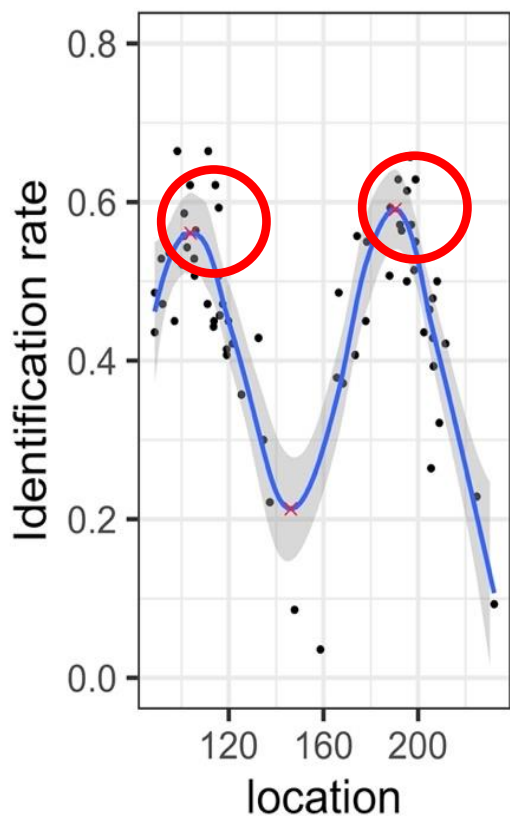
- *Location* and *shape* contribute to tone identification

		Location <i>P</i> value	Scale <i>P</i> value	Shape <i>P</i> value
Baseline T6	T1	<0.001*	0.94	<0.001*
	T3	<0.001*	0.17	<0.001*
Baseline T1	T3	<0.001*	0.196	<0.001*
	T6	<0.001*	0.608	<0.001*

# Results

## Locally Weighted Scatterplot Smoothing (LOESS) Curves

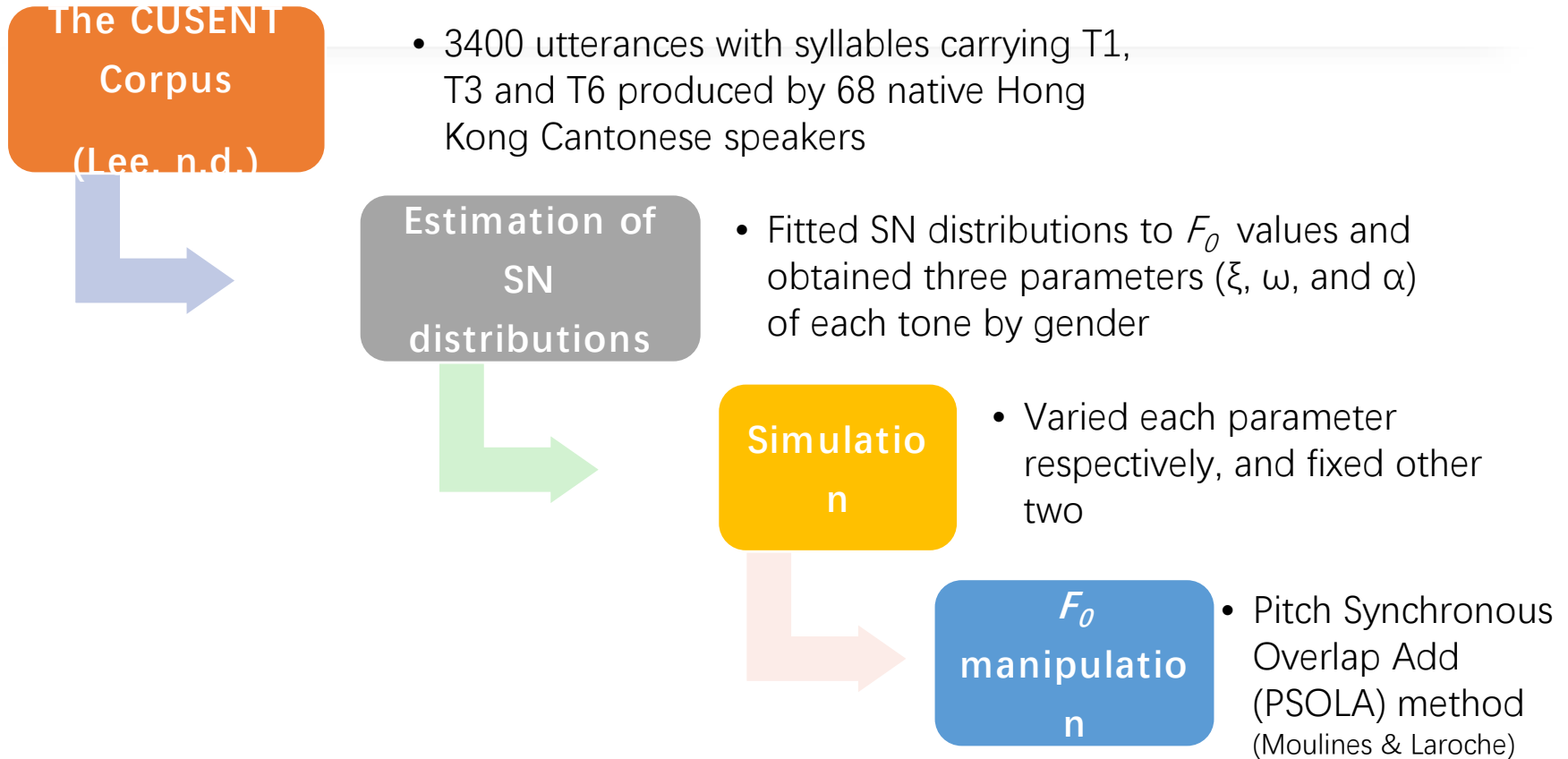
- **Two peaks of Location:** Identify T3 spoken by male and female speakers
- **One peak of Scale:** Identify T3 production with variability



# Method

## Experiment 2: Synthesized Tone Identification in Isolation

- Stimuli



Semitone distance	T1 vs. T3	T1 vs. T6	T3 vs. T6
<b>Male</b>	2.64	3.56	0.93
<b>Female</b>	3.01	3.90	0.88

## Participants:

- The 14 native Mandarin speakers in Experiment 1

## Stimuli:

Two carrier sentences:

- High tones T1(55):
  - 聽聽 /thiŋ thiŋ/\_\_\_\_ (Listen to \_\_\_\_ )
- Low tones T6(22):
  - 就係 /thsəu hɛi/\_\_\_\_ (This is just \_\_\_\_ )

## Experiment 3

## Tone Identification in Contexts

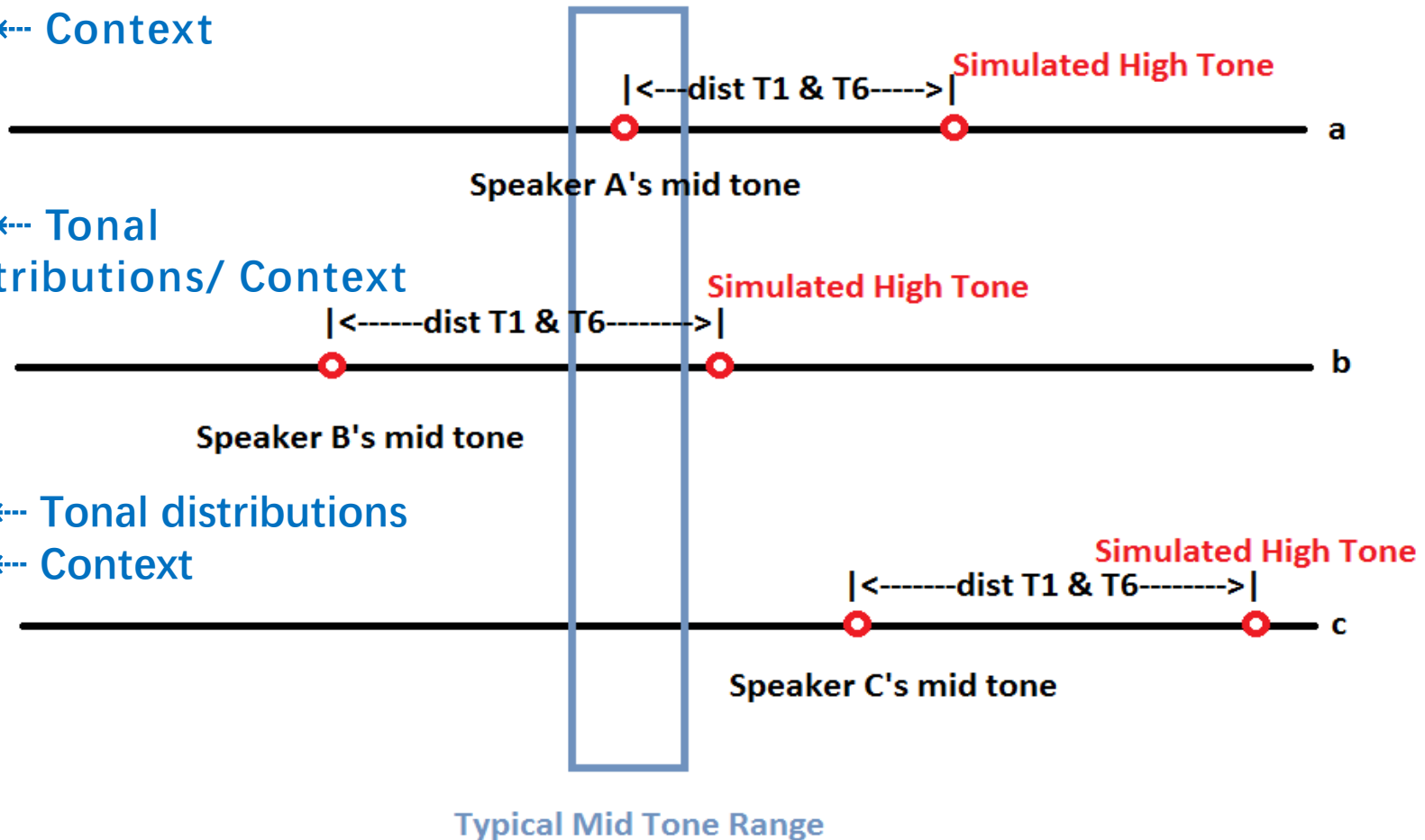
T3 ← Tonal distributions

T6 ← Context

T6 ← Tonal distributions/  
Context

T1 ← Tonal distributions

T6 ← Context



Simulated high tones (T1) treating mid tones (T3) as low tones (T6)

T1 as the baseline	Simulated low tone			Simulated high tone	
	Estimate	p-value		Estimate	p-value
<b>Indicator 3</b>	0.53	< 0.001	<b>Indicator 3</b>	0.49	< 0.001
<b>Indicator 6</b>	-0.61	< 0.001	<b>Indicator 6</b>	-0.69	< 0.001
<b>tt 3</b>	0.46	< 0.001	<b>zh 3</b>	0.10	0.413
<b>tt 6</b>	1.38	< 0.001	<b>zh 6</b>	0.90	< 0.001

T6 as the baseline	Simulated low tones			Simulated high tones	
	Estimate	p-value		Estimate	p-value
<b>Indicator 1</b>	0.61	< 0.001	<b>Indicator 1</b>	0.69	< 0.001
<b>Indicator 3</b>	1.14	< 0.001	<b>Indicator 3</b>	1.18	< 0.001
<b>tt 1</b>	-1.38	< 0.001	<b>zh 1</b>	-0.90	< 0.001
<b>tt 3</b>	-0.92	< 0.001	<b>zh 3</b>	-0.80	< 0.001

# Results

## Non-native natural/synthesized tone identification in isolation:

- Distribution parameters significantly contribute to the normalization of Cantonese level tones in Mandarin speakers.
- Akin to the findings among native Cantonese speakers (Chen et al., 2022).
- However, Mandarin speakers display a reduced sensitivity compared to Cantonese speakers reported in Chen et al. (2022).
- Mandarin speakers may not have established mental representations of non-native tones as parametrically as their native counterparts.



## Non-native tone identification in contexts:

- Both mentally stored representations and contextual information contribute to the identification of non-native lexical tones.
- However, the effect of contextual information is weaker than mentally stored parameters under certain conditions.
- Mentally stored distributional parameters remain accessible even in the presence of contextual information.

1. Bauer, R. S. & Benedict, P. K. Modern Cantonese phonology. *De Gruyter* (1997).
2. Si, C., Zhang, C., Lau, P. *et al.* Modelling representations in speech normalization of prosodic cues. *Sci Rep* **12**, 14635. <https://doi.org/10.1038/s41598-022-18838-w> (2022).
3. Francis, A. L., Ciocca, V. & Ng, B. K. C. On the (non) categorical perception of lexical tones. *Percept. Psychophys.* **65**(7), 1029–1044. <https://doi.org/10.3758/BF03194832> (2015).
4. Hyman, L. How concrete is phonology? *Language*. **46**(1), 58–76. <https://doi.org/10.2307/412407>(1970).
5. Idemaru, K. & Holt, L. L. Word recognition reflects dimension-based statistical learning. *J. Exp. Psychol. Hum. Percept. Perform.* **37**(6), 1939–1956. <https://doi.org/10.1037/a0025641> (2011).
6. Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. Perception of the speech code. *Psychol. Rev.* **74**(6), 431–461. <https://doi.org/10.1037/h0020279> (1967).
7. McMurray, B. & Jongman, A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychol. Rev.* **118**(2), 219–246. <https://doi.org/10.1037/a0022325> (2011).
8. Moulines, E. & Laroche, J. Nonparametric techniques for pitch-scale and time scale modification of speech. *Speech Commun.* **16**, 175–205 (1995).
9. Pierrehumbert, J. B. Phonological representation: Beyond abstract versus episodic. *Ann. Rev. Linguist.* **2**(1), 33–52. <https://doi.org/10.1146/annurev-linguistics-030514-125050> (2016).
10. Stilp, C. E. & Theodore, R. M. Talker normalization is mediated by structured indexical information. *Atten. Percept. Psychophys.* **82**(5), 2237–2243. <https://doi.org/10.3758/s13414-020-01971-x> (2020).
11. Wade, T. & Möbius, B. Detailed phonetic memory for multi-word and part-word sequences. *Lab. Phonol.* **1**(2), 283–294. <https://doi.org/10.1515/labphon.2010.014> (2010).
12. Wang, W. S. Y. Language change. *Ann. N. Y. Acad. Sci.* **280**(1), 61–72. <https://doi.org/10.1111/j.1749-6632.1976.tb25472.x>(1976).
13. Wong, P. & Diehl, R. L. Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *J. Speech Lang. Hearing Res. JSLHR* **46**(2), 413–421. [https://doi.org/10.1044/1092-4388\(2003/034\)](https://doi.org/10.1044/1092-4388(2003/034)) (2003).

# Thank you!