Top-down and bottom-up processing of familiar and unfamiliar Mandarin dialect tone systems

Liang Zhao

Experimental Lab meeting | LLS | University of York

18 November 2021

Outline

1. **Introduction**

2. Production experiment
   - Task overview: Mandarin dialect production experiment (remote audio collection)
   - Goal: Establish the acoustic-phonetic tonal realisations in Beijing Mandarin (~Standard Mandarin) and Chengdu Mandarin

3. Perception experiment
   - Task overview: Surprisal sentence perception (accuracy and response time)
   - Goal: Determine the relative weighting of top-down and bottom-up information in the perception of lexical tone

4. Discussion

Speech perception revisited

speech signal ———————— intended utterance

- Path from acoustic signal to units of perception: *many-to-many*

- Various sources of information may jointly contribute to speech processing

Speech perception: bottom-up and/or top-down

Many early models assumed bottom-up processing as a first attempt

> e.g. The Cohort Model (Marslen-Wilson 1978, 1987), Direct Perception (Gibson 1954) & Direct Realism (Fowler 1986)

Further development of the theory did take top-down influence into consideration

- Advocate: TRACE (McClelland & Elman 1986), Acoustic landmarks & distinctive features (Stevens 2002, 2008)
- Opponent: Shortlist (Norris 1994), Merge (Norris, Cutler and McQueen 2000)

But the relative weighting and integration of bottom-up and top-down information in speech processing remains unclear

Speech perception: segmental (and suprasegmental)

Models of speech perception have been segment-oriented, and mostly on non-tonal languages

What about suprasegmental processing?

>> mechanisms for tonal speech perception (such as Mandarin dialect)

>> lexical tone processing for lexical access

Speech perception: lexical tone processing

Debate on the relative role of lexical tone and segmental information for lexical access

- Segmental information >> lexical tone in sub-lexical processing

  (Cutler & Chen 1999, Ye & Connine 2010, Li et al. 2013)

- Lexical tone >> or ≈ segmental information with top-down feedback

  (Schirmer et al. 2005, Liu & Samuel, 2007, Malins & Joanisse, 2010)

Extension of TRACE: Reverse Accessing Model (RAM, Gao et al. 2019)

- Report on a distinct advantage for segments over lexical tones, suggesting that tone information is processing only *if necessary*

Speech perception: unfamiliar lexical tone processing

Lexical tone processing

- Context seems crucial (segmental vs. tonal information)

- General consensus on using both bottom-up and top-down information

- Most often assumed for native/familiar lexical tone systems

Would **non-native/unfamiliar lexical tone** processing be different? And how?

What happens when **the tone system is unfamiliar, but the segmental system is familiar?**

First, let's get to know about the tone system of **Mandarin dialects**

Mandarin dialects:
comparable segmental inventories, but **distinct** tone inventories
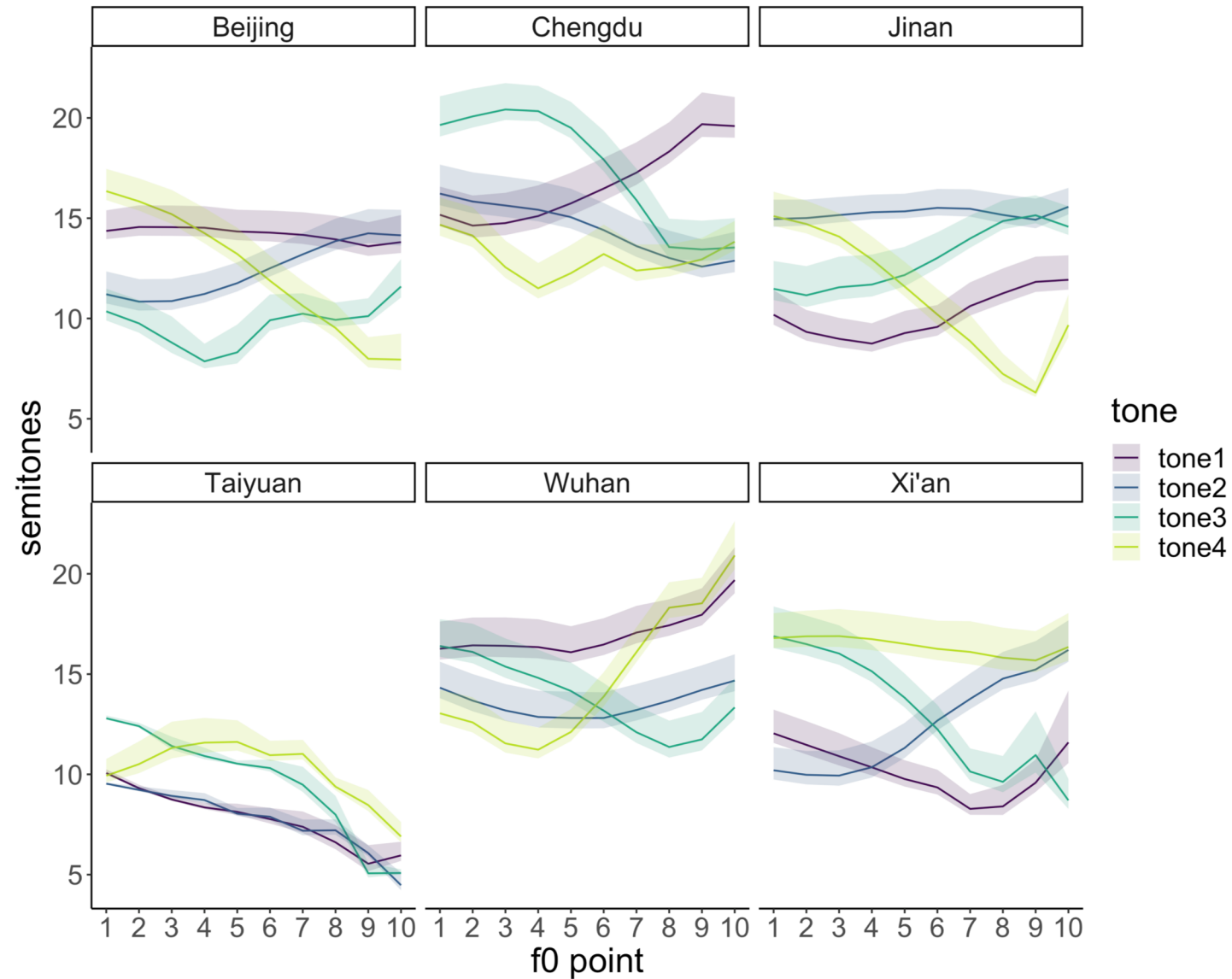
Speech perception: unfamiliar Mandarin dialects

Mutual intelligibility (Tang & van Heuven, 2007, 2008, 2009)

- Relatively high mutual intelligibility in semantic decision task with <u>carrier sentences</u>

- Intelligibility dropped with words presented in isolation

For the unfamiliar (non-native) tone system, we hypothesize the dominance of **top-down information**

- Fully top-down?

- Hybrid process with integration of bottom-up processing?

For the familiar (native) tone system, we expect **both top-down and bottom-up processing**

Outline

1. Introduction

2. **Production experiment**
   - Task overview: Mandarin dialect production experiment
   - Goal: Establish the acoustic-phonetic tonal realisations in Beijing Mandarin (~Standard Mandarin) and Chengdu Mandarin

3. Perception experiment
   - Task overview: Surprisal sentence perception (accuracy and reaction time)
   - Goal: Determine the relative weighting of top-down and bottom-up information in the perception of lexical tone

4. Discussion

Production experiment

Participants

- 8 native Beijing speakers (represent Standard Mandarin)

- 7 native Chengdu speakers

Stimuli (reading material) & Tasks

- **80 monosyllabic words** (20 forms * 4 tone categories)

- **48 sentences** (potential perceptual stimuli)

- Production in Standard Mandarin and their regional dialect

Production experiment

Acoustic analysis

- Alignment
  - Transcripts generated from *Gorilla* and processed as text files for each speaker
  - Forced alignment using Montreal Forced Aligner
  - Manual adjustment of vowel boundaries

- Tone categories: f0 contours
  - 10 equally spaced f0 values across all vocalic intervals
 - Converted to semitones
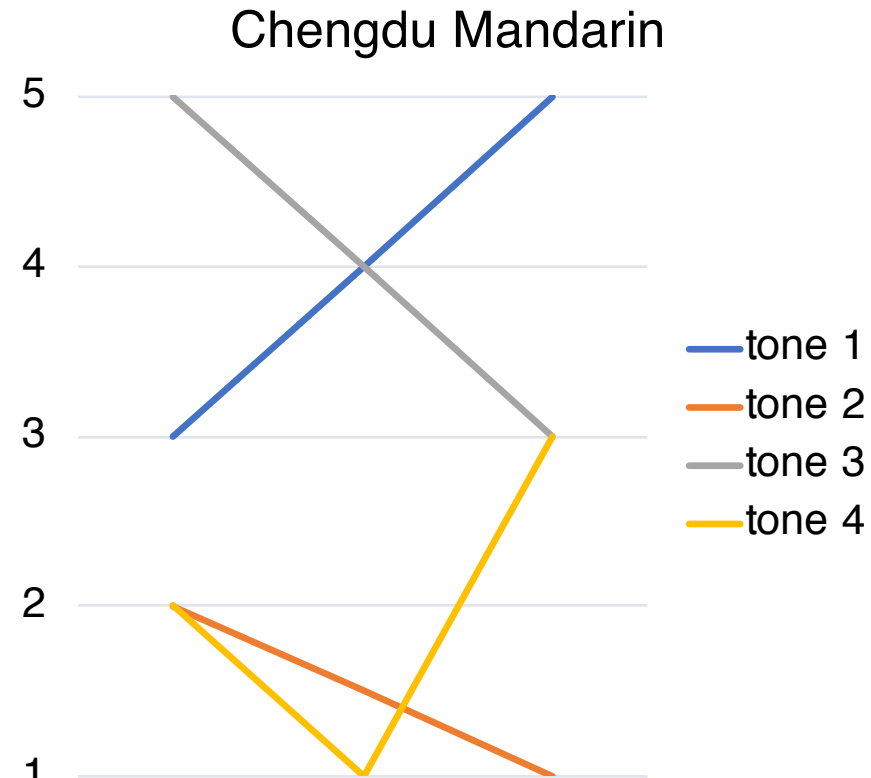
Production experiment

Results

- Highly disparate realisations of lexical tone between Standard Mandarin and Chengdu Mandarin

- Contour similarity observed for some tone categories

	Chengdu	Beijing
Tone 1	25 (55*)	55
Tone 2	21	35
Tone 3	53	214
Tone 4	213	51

- (55*) is what has been previously recorded in dialect dictionaries, but our result shows this tone is more likely a rising tone (25)
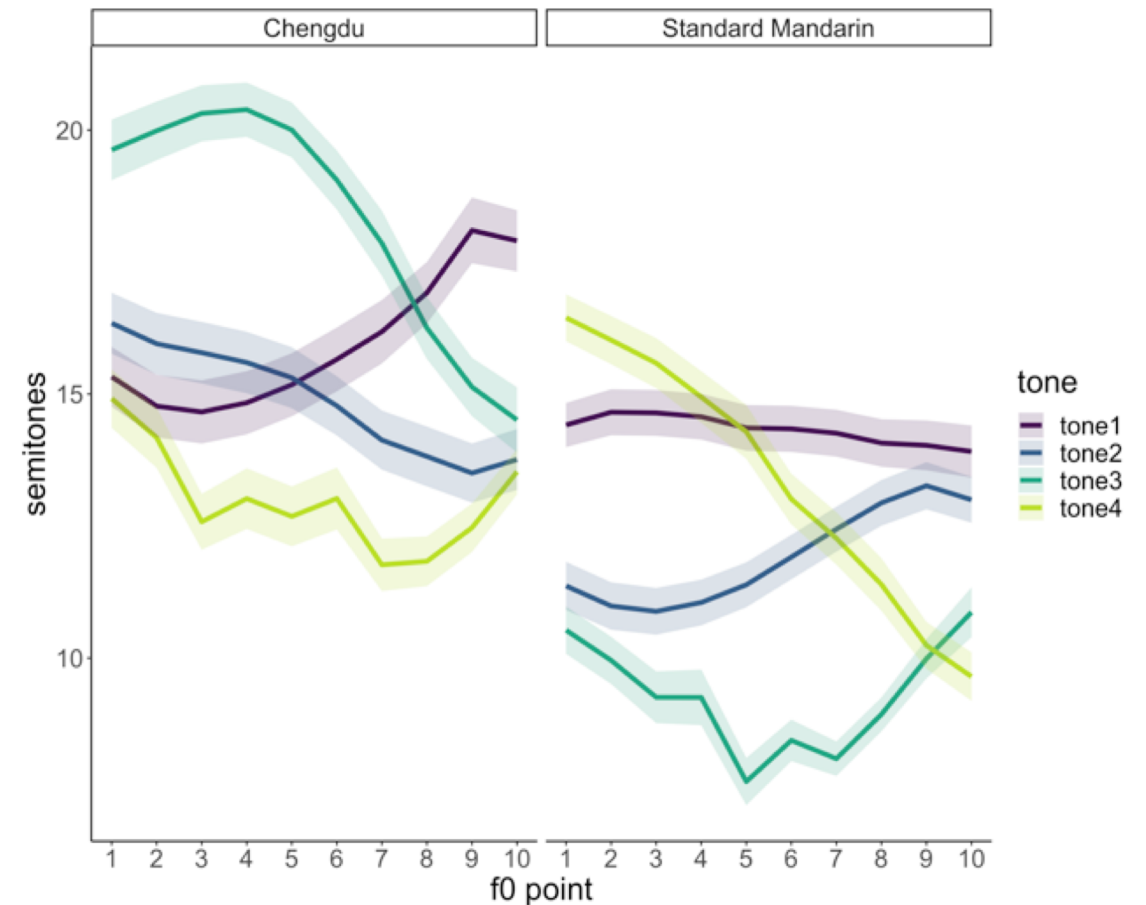
Outline

1. Introduction

2. Production experiment
   - Task overview: Mandarin dialect production experiment (remote audio collection)
   - Goal: Establish the acoustic-phonetic tonal realisations in Beijing Mandarin (~Standard Mandarin) and Chengdu Mandarin

3. Perception experiment
   - Task overview: Surprisal sentence perception (accuracy and response time)
   - Goal: Determine the relative weighting of top-down and bottom-up information in the perception of lexical tone systems

4. Discussion

Perception experiment

Participants

- 21 native Mandarin, non-Chengdu speakers

Task

- Semantic plausibility judgment task of a spoken sentence

  "***Does this sentence make sense***?" and clicked **"yes" or "no"** on the screen after hearing the whole sentence

Perception experiment

Stimuli

24 sentences were created manipulating Mandarin dialect in a between-item design (12 Standard and 12 Chengdu Mandarin sentences)

Experimental manipulation (48 trials = 24 sentences * 2 surprisal conditions)

- **Dialect conditions**: Standard Mandarin vs. Chengdu Mandarin

- **Surprisal conditions**: low surprisal vs. high surprisal

An example sentence frame manipulated by surprisal

low-surprisal sentence	high-surprisal sentence

a) 有　一只　鹰　在　天上　飞
   You3 yi4 zhi1 ying1 zai4 tian1 shang4 <u>fei1</u>
   There is an eagle in the sky <u>flying</u>
   "There is an eagle flying in the sky"

b)* 有　一只　鹰　在　天上　<u>肥*</u>
   You3 yi4 zhi1 ying1 zai4 tian1 shang4 <u>fei2</u>*
   There is an eagle in the sky <u>gaining weight</u>*
   "There is an eagle gaining weight in the sky"

Perception experiment

More on stimuli

- Standard Mandarin produced by a female speaker (age 26) and Chengdu Mandarin by a male speaker (age 29)

- All 48 sentences randomly presented with no repetitions or indication of dialect

- Listeners were made aware of the two possible dialects in the task instruction

Perception experiment

Procedure (online *Gorilla* Experiment builder)

- Familiarization phase

  - Familiarization with two pairs of high- and low-surprisal sentences In Standard Mandarin

  - "***Does this sentence make sense***?" and clicked **"yes" or "no"** on the screen after hearing the whole sentence

  - Immediate feedback on the correct answer and the sentence

- Test phase

  - Identical to the familiarization phase, except no feedback was provided

  - The presentation of trials was fully randomized

Perception experiment

Data analysis

Accuracy and response time from the test phase were measured as dependent variables across dialect and surprisal manipulations

- Accuracy
  - "Yes" responses to low surprisal (i.e., plausible) sentences
  - "No" responses to high surprisal (i.e., implausible) sentences

- Response times
  - calculated as the interval between the end of the audio file and the click registering a judgment

Perception experiment

Data analysis (models)

Accuracy: Bayesian logistic mixed-effects regression

Response time: Bayesian linear mixed-effects regression (Bürkner, 2018)

- Each model included
  - The fixed effects: surprisal, dialect, trial number, and the full set of interactions

  - The random effects:
    - For participant: an intercept for participant, slopes for surprisal, dialect, trial number, the interaction between surprisal and dialect
    - For sentence frame, an intercept and random slope for dialect

Accuracy

Expected results in the semantic plausibility task

	Low-surprisal sentence	High-surprisal sentence
Standard Mandarin	Accuracy ⬆ "yes"	Accuracy ⬆ "no"
Chengdu Mandarin	Accuracy ⬇ "yes"	Accuracy ⬇ *"yes"

Accuracy

Credible main effects of *surprisal*, *dialect* and an *interaction between surprisal and dialect*

- *Surprisal*: low-surprisal >> high-surprisal condition

- *Dialect*: Standard Mandarin >> Chengdu Mandarin

- *Interaction*: even less accurate in the high-surprisal Chengdu condition relative to average
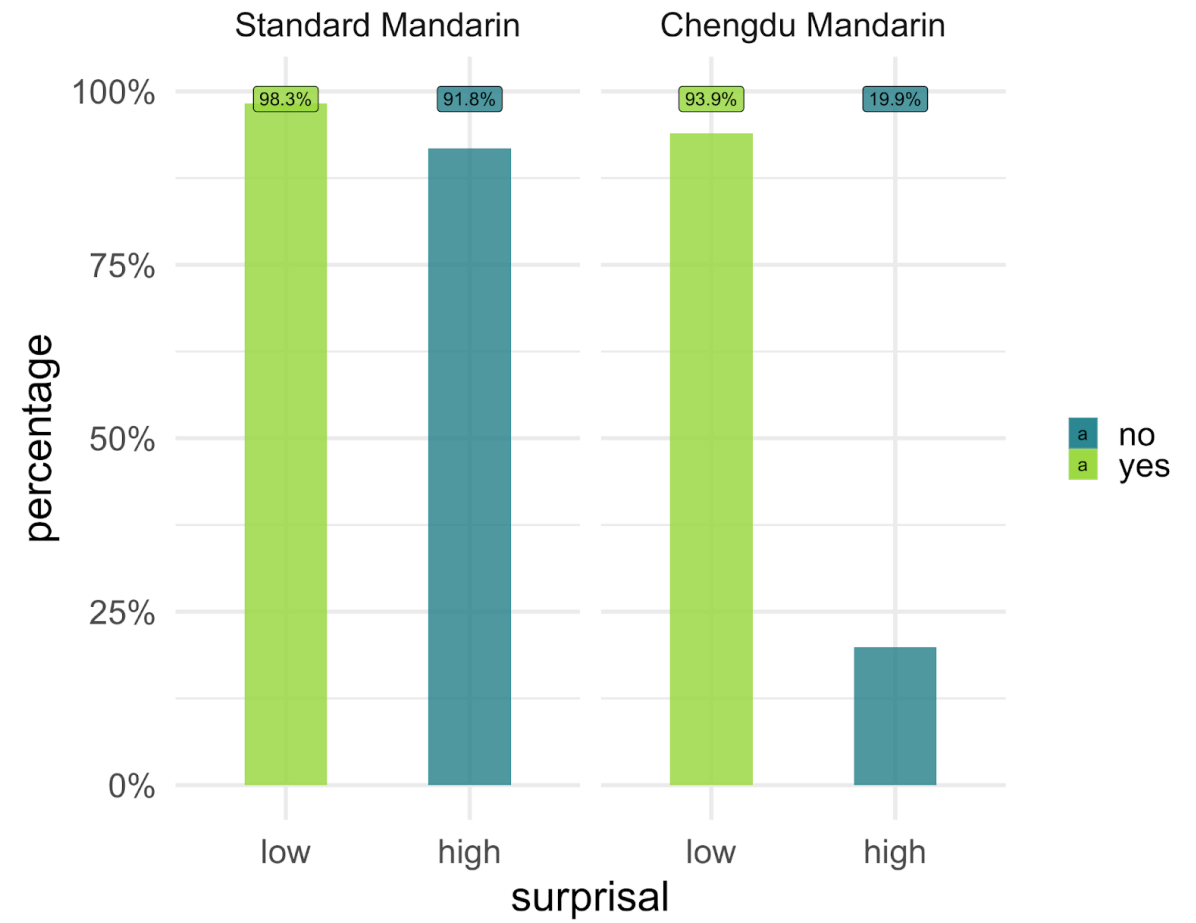
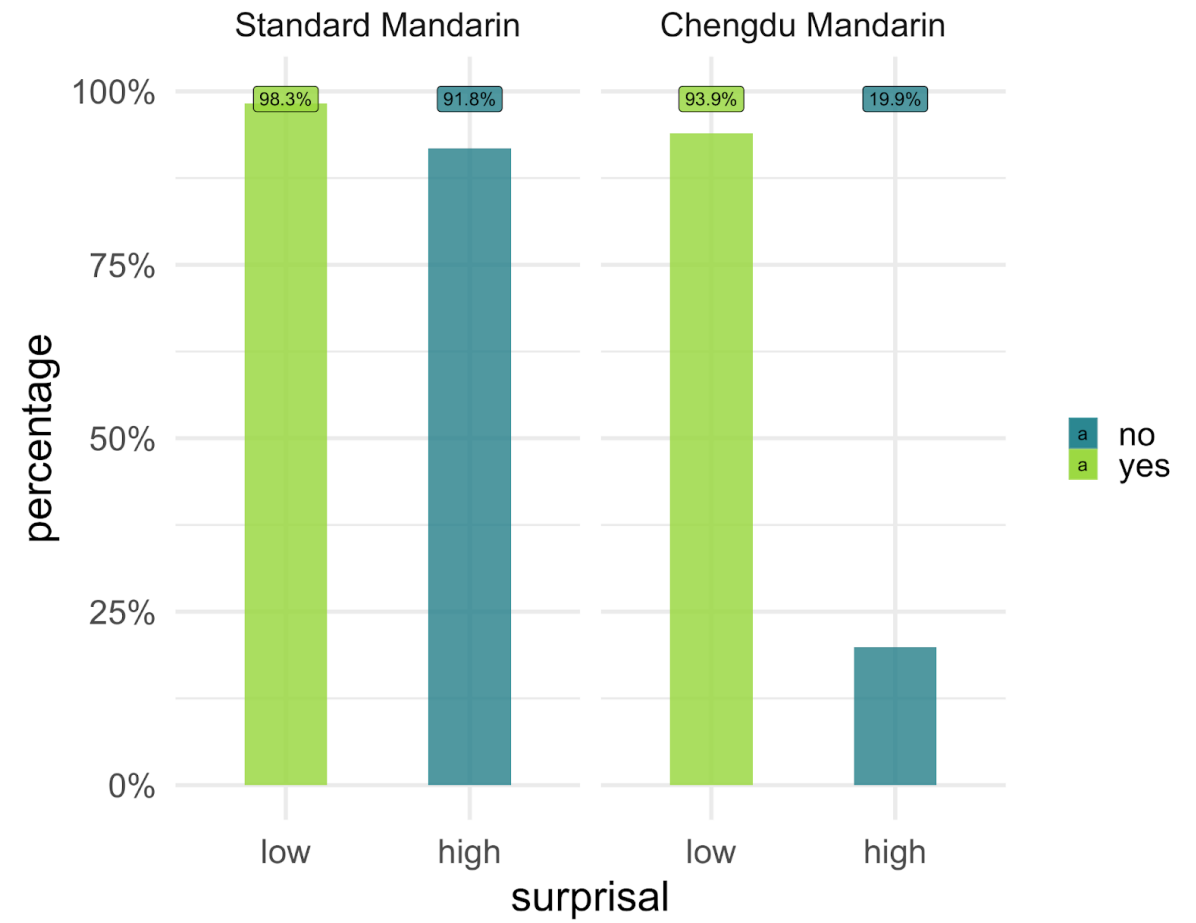# Accuracy

Familiar tone system (Standard Mandarin)

- High accuracy over 90% as expected

- The participants understood the task on general

- **Use of both top-down (sentential context) and bottom-up information (word-tone knowledge)**
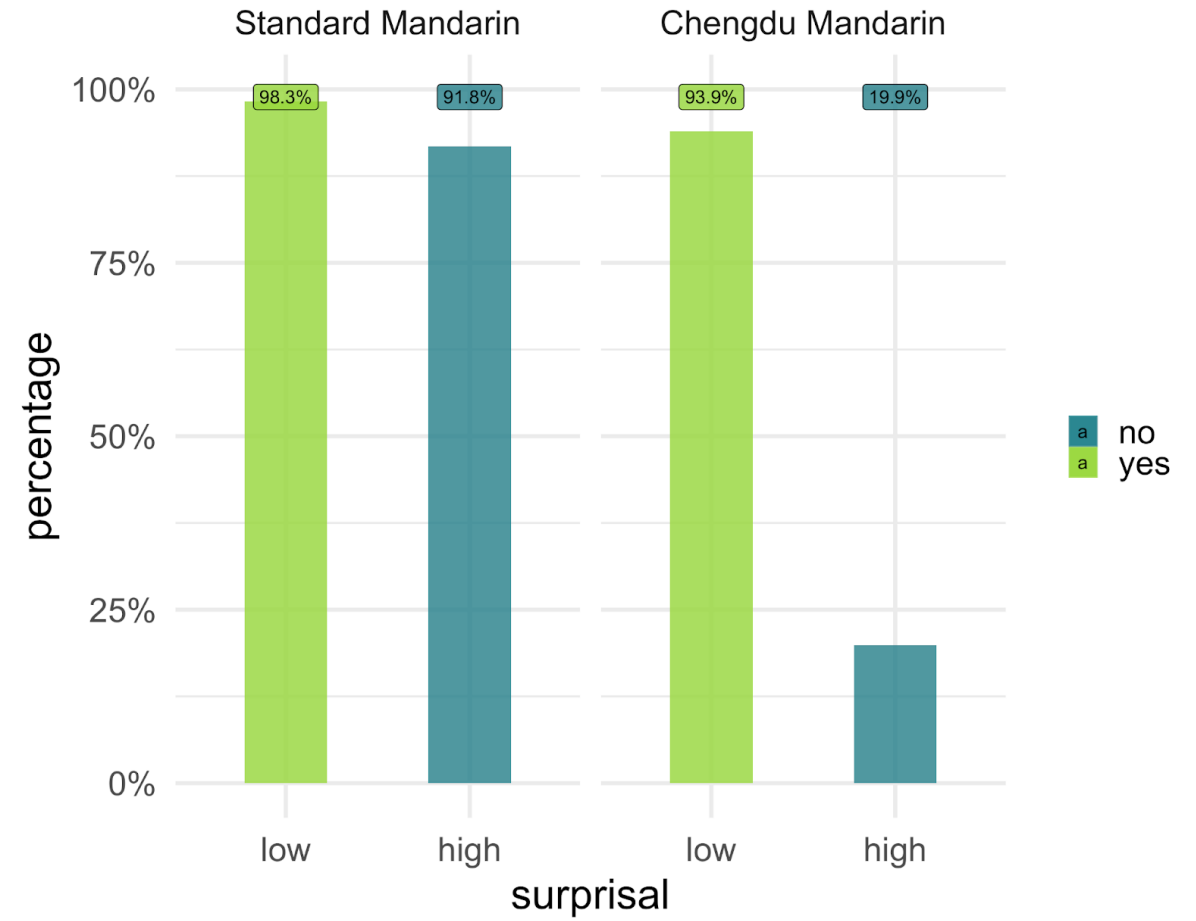
Accuracy

## Unfamiliar tone system (Chengdu Mandarin)

- Bias to respond "semantically plausible"/ **"yes"** Suggests

  - Major bottom-up *failure* in identifying tone mismatch in high-surprisal condition

  - relatively high intelligibility

- **Top-down influence overriding bottom-up tone acoustics**

Accuracy

No reliable effects of *trial number* and *its interactions with surprisal and dialect*

- *Accuracy did **not reliably improve** in any condition over the course of experiment*
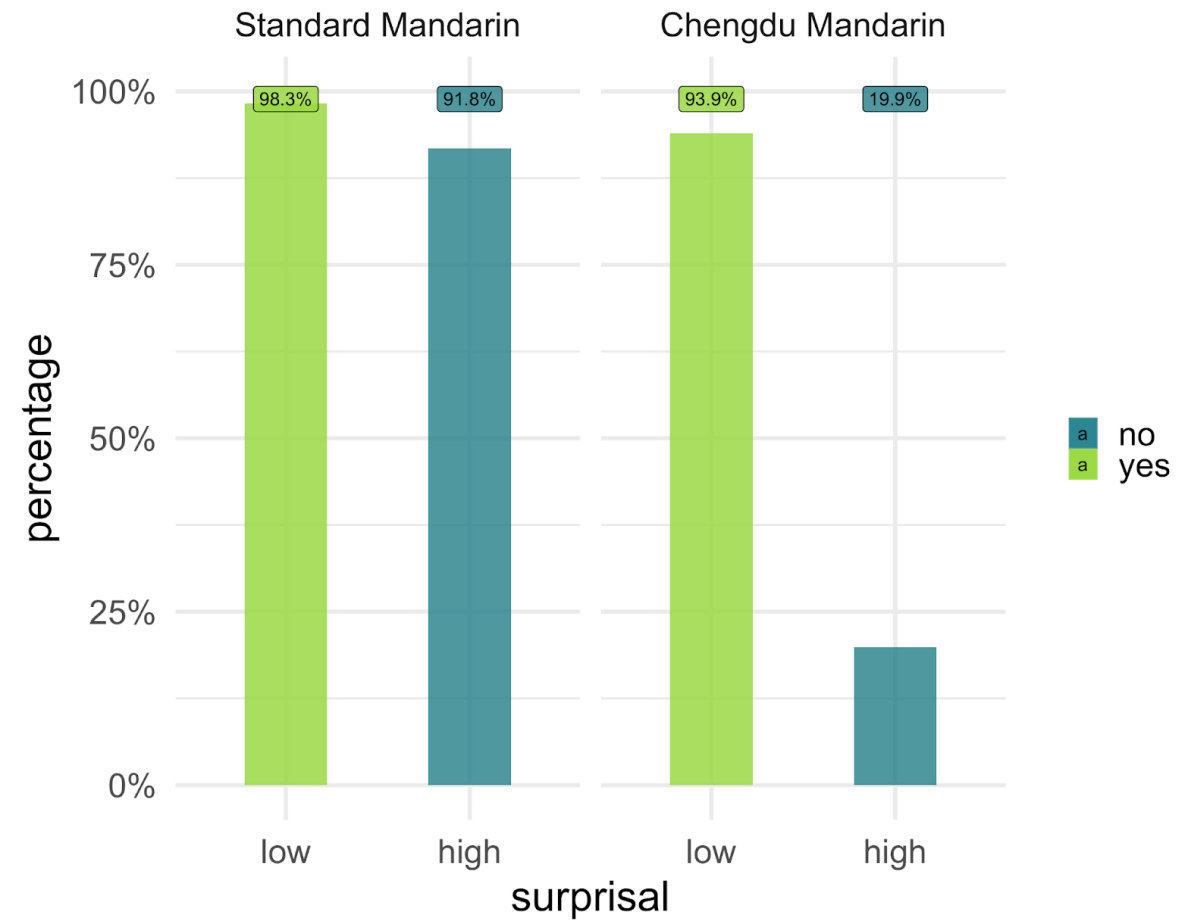
Response time

Expected results in the semantic plausibility task

|	Low-surprisal sentence	High-surprisal sentence
Standard Mandarin	baseline	slowdown ⬇
Chengdu Mandarin	==	==

Response time

Reliable main effects for *surprisal, dialect,* and *the interaction between surprisal and dialect*

Reliable slowdown:

- high-surprisal >> low-surprisal condition
- Chengdu Mandarin >> Standard Mandarin

Difference between high- and low-surprisal:

- Standard Mandarin (estimated mean difference = 800 ms) >> Chengdu Mandarin (300 ms)
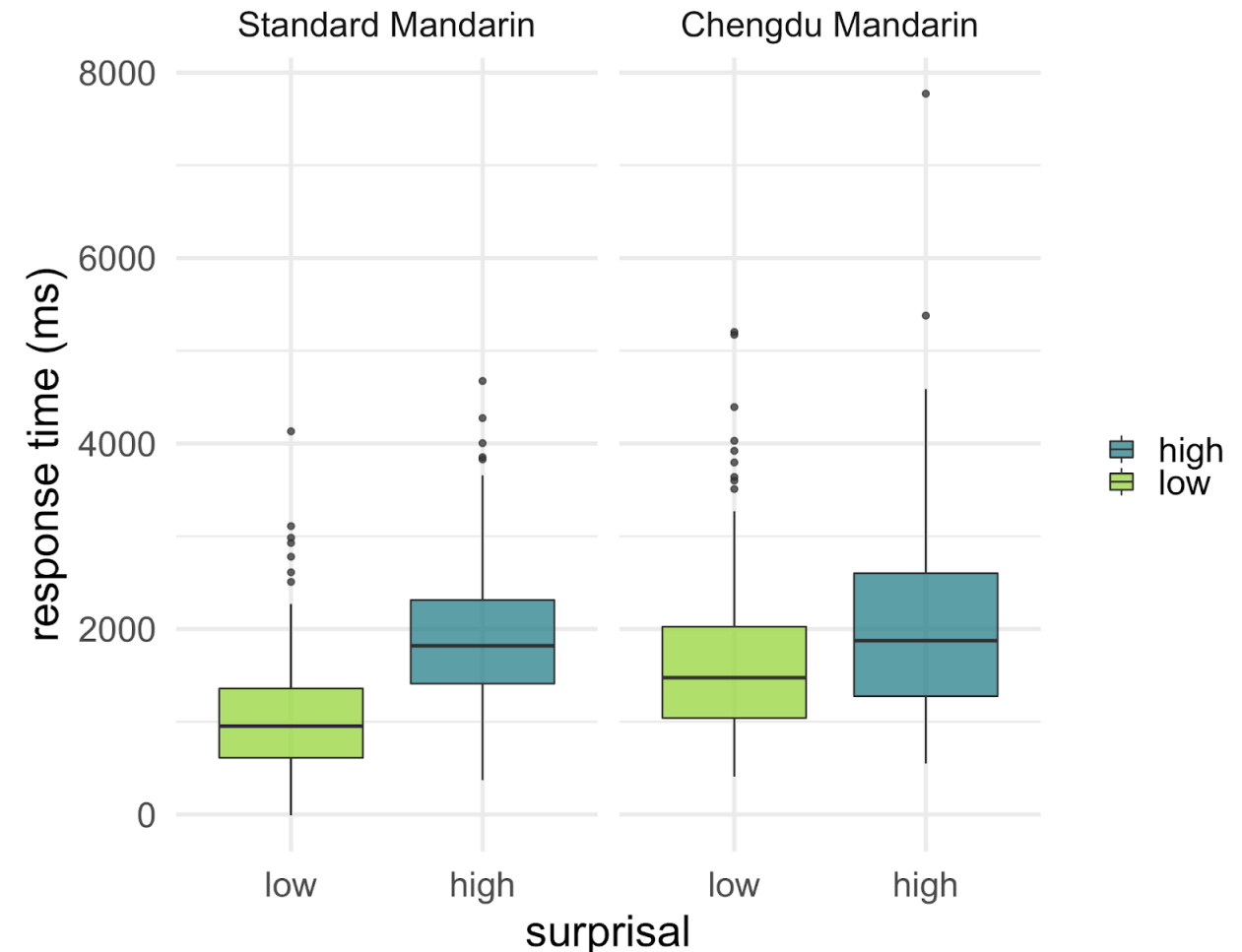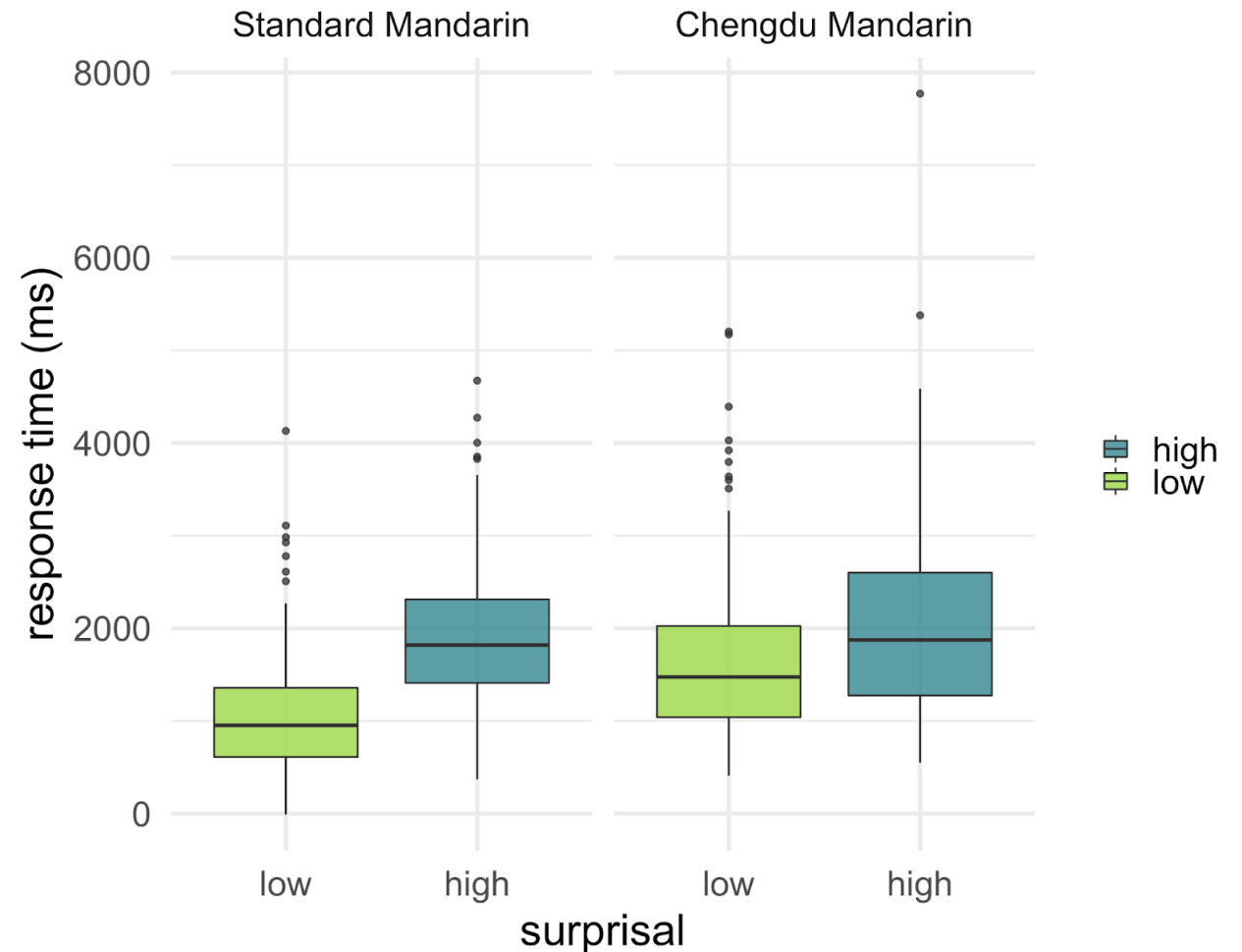
Response time

Reliable slowdown in high-surprisal condition consistent for **both** familiar and unfamiliar tone systems

- o   For native speech: expected
- o   For non-native speech: unexpected!

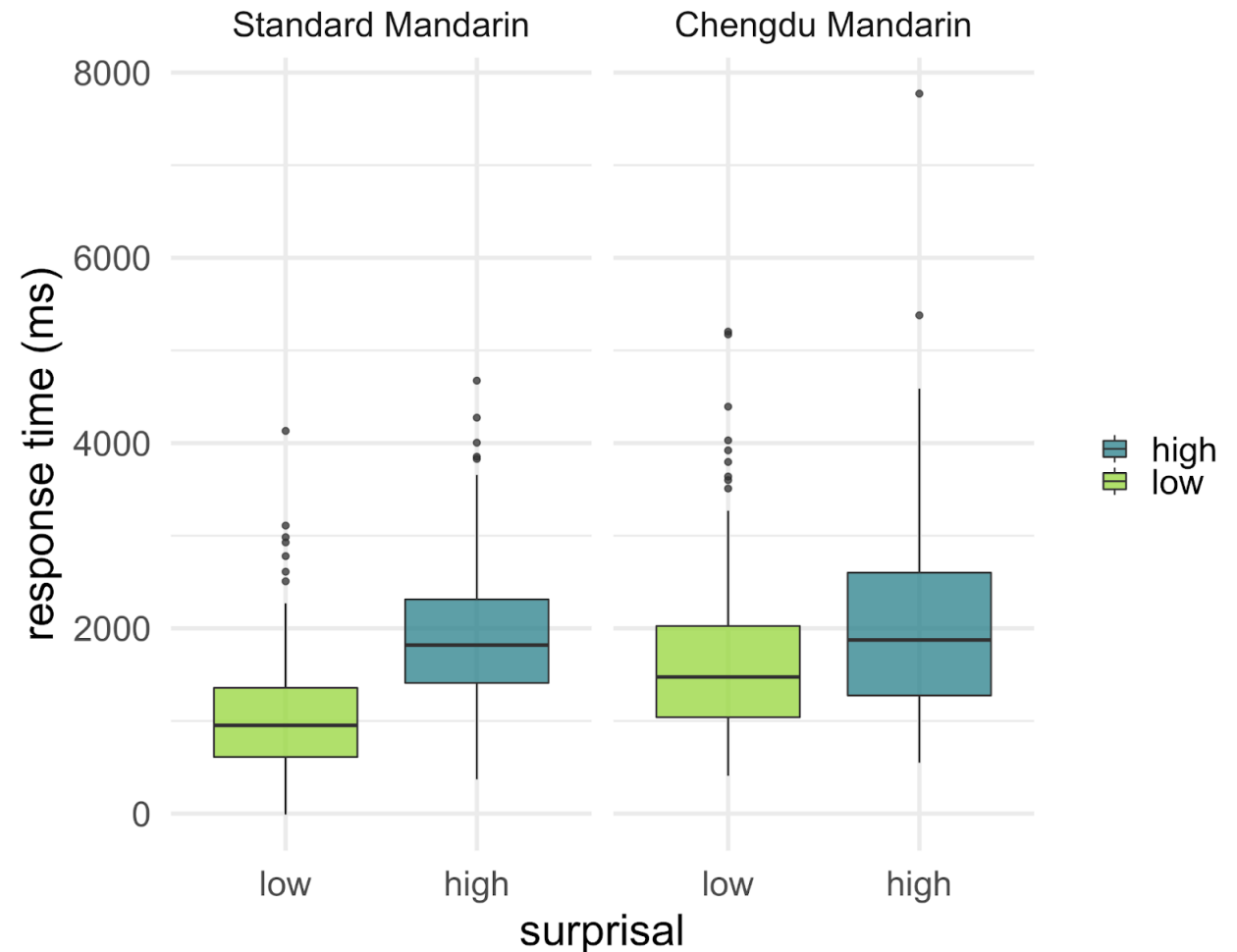**Both bottom-up and top-down processing are present**

- o   Processing tone information in the unfamiliar tone system (Chengdu Mandarin)!

Response time

No reliable effects of *trial number* and *its interaction with surprisal and dialect*

- *Response times did **not reliably differ** in any condition over the course of experiment*

- **Rapid learning** of the unfamiliar tone system

   ----Attention to the tone surprisal as early as the experiment commenced

Summary

Findings

- Highly **disparate tone systems** between Chengdu and Standard Mandarin

- **High intelligibility** of Chengdu dialect by native speakers of Standard Mandarin

- Accuracy results suggested **dominance of higher-level top-down information** in perception of unfamiliar tone system (Chengdu Mandarin)

- Response time results suggested listeners' sensitivity to high-surprisal tones, indicating **bottom-up processing** of unfamiliar tone system

Discussion

Integration of bottom-up and top-down processing of familiar and unfamiliar Mandarin dialect tone systems

- Opposite to RAM (Gao et al. 2019) which proposed that tone information is "processed *if necessary*", we argue that tone information is *always* processed

Integration of bottom-up and top-down processing of familiar and unfamiliar Mandarin dialect tone systems

Familiar Standard Mandarin speech

- Accuracy results suggested that listeners have strong representations of tone and segments (bottom-up)

- Response times suggested they are sensitive to the sentence surprisal (top-down)

Integration of bottom-up and top-down processing of familiar and unfamiliar Mandarin dialect tone systems

Unfamiliar Chengdu Mandarin speech

- Accuracy results suggested an overriding effect of top-down information in determining sentence meaning (ultimate bias towards plausible judgment)

- Differences in response times between high- and low-surprisal indicated unexpected integration of bottom-up information
  - One explanation: listeners are constructing *impoverished tone representations* during online processing (bottom-up)

  - But, the updated tone systems are not reliable enough and thus top-down information overrides the output of tone-level processing

  - Rapid learning of the unfamiliar tone system: trial order and relevant interactions not reliable in either accuracy and response time models

Discussion

Follow-up experiment with Chengdu Mandarin exposure

- Accuracy improves after two minutes of uninterrupted exposure

- Listeners may indeed be learning tone representations or gaining confidence of them

Future directions

Current study could be extended:

- To revise the between-item design to a true Latin Square design

- To experiment on a broad range of Mandarin dialects to test the hypotheses

- To investigate whether some tone categories are easier to learn in the unfamiliar tone systems

Thank you!

Dialect code	City	Province
BEI	**Standard (Beijing)**	
TYN	Taiyuan	Shanxi
JNN	Jinan	Shandong
XIA	Xi'an	Shaanxi
CHD	**Chengdu**	Sichuan
WHN	Wuhan	Hubei