Top-down and Bottom-up Processing of Familiar and Unfamiliar Mandarin Dialect Tone Systems

Liang Zhao, Shayne Sloggett, and Eleanor Chodroff
University of York, Department of Language and Linguistic Science

UNIVERSITY of York

Introduction

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

Speech (segmental) processing frameworks have identified two high-level mechanisms:
bottom-up processing & top-down processing

- Many early models assumed bottom-up processing as a first attempt
  - The Cohort Model (Marslen-Wilson 1978, 1987), Direct Perception (Gibson 1954) & Direct Realism (Fowler 1986)

- Further development of the theories took 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)

- Current models tend to incorporate **both top-down and bottom-up processes in speech perception**, but the relative weighting and integration of these sources of information remains unclear.

What about **suprasegmental processing?**
(lexical tone processing of native tone systems)

- 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): tone information accessed only *if necessary*

- General agreement on **the use of both top-down and bottom-up information in tone processing,** but there is little consensus on the relative weighting of these sources of information and how they interact.

Non-native lexical tone processing:
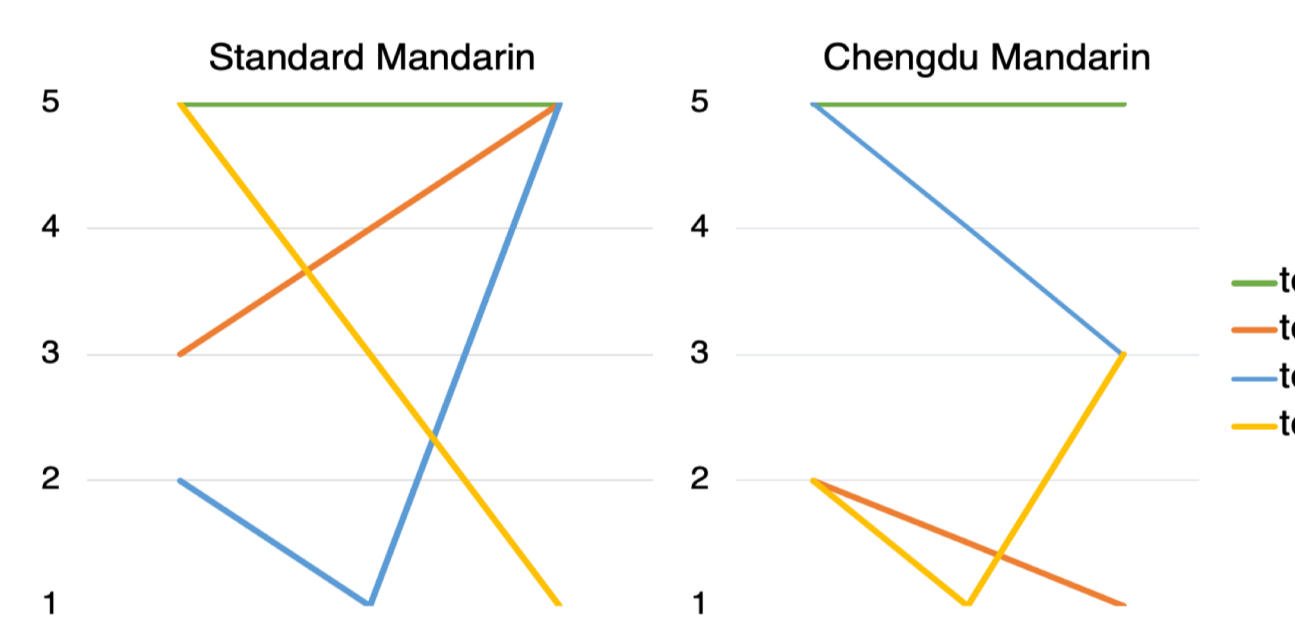unfamiliar tone system but familiar segmental inventory

Figure 1: Schematic tone contours of Standard Mandarin and Chengdu Mandarin.

Goals

- To examine whether tone processing differs between familiar (Standard Mandarin) and unfamiliar (Chengdu Mandarin) tone systems
- To investigate how top-down and bottom-up processing might interact and their relative weighting in tone processing—fully top-down or hybrid processing

Expected results

- Familiar tone system:
  - Both top-down & bottom-up processing of tonal information
- Unfamiliar tone system:
  - Dominance of top-down information from sentential context
  - Little or no use of lexical tone due to the unfamiliarity of the tone system

References

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive psychology, 10(1), 29-63.
Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. Cognition, 25(1-2), 71-102.
Gibson, J. J. (1954). A theory of pictorial perception. Audiovisual communication review, 2(1), 3-23.
Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. Journal of phonetics, 14(1), 3-28.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive psychology, 18(1), 1-86.
Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. The Journal of the Acoustical Society of America, 111(4), 1872-1891.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. Cognition, 52(3), 189-234.
Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences, 23(3), 299-325.
Taft, M., & Chen, H. C. (1992). Judging homophony in Chinese: The influence of tones. In Advances in psychology (Vol. 90, pp. 151-172). North-Holland.
Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. Perception & Psychophysics, 59(2), 165-179.
Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. Language and cognitive processes, 14(5-6), 609-630.
Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. Language and speech, 59(1), 59-82.
Liu, S., and Samuel, A. G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. Lang. Cogn. Process. 22, 566–594.
Gao, X., Yan, T. T., Tang, D. L., Huang, T., Shu, H., Nan, Y., & Zhang, Y. X. (2019). What makes lexical tone special: A Reverse Accessing Model for tonal speech perception. Frontiers in psychology, 10, 2830.
李荣 (Li, Rong.). (2002). 现代汉语方言大词典 (The Modern Dictionary of Chinese Dialects). 江苏教育出版社 (Jiangsu Education Press).
Anwyl-Irvine, A.L., Massonié J., Flitton, A., Kirkham, N.Z., Evershed, J.K. (2020). Gorilla in our midst: an online behavioural experiment builder.
Zhao, L., & Chodroff, E. (2022). ManDi: Mandarin Chinese Dialect Corpus. Retrieved from https://osf.io/69fx5/.
Bürkner, P. C. (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms." The R Journal, 10(1), 395–411. doi:10.32614/RJ-2018-017.

Method

Participants
21 native speakers of Standard Mandarin (little/no knowledge of Chengdu Mandarin)

Stimuli
24 pairs of low/high-surprisal spoken sentences manipulating Mandarin dialect (Standard vs. Chengdu Mandarin) in a between-item design

Experimental manipulation (24×2 =48 trials)
Surprisal: high surprisal (+1) vs. low surprisal (-1)
Dialect: Chengdu Mandarin (+1) vs. Jinan Mandarin (-1)

Parentheses indicate sum coding for the statistical models

Table 1: An example sentence item across surprisal conditions

low-surprisal sentence	a) 有 一只 鹰 在 天上 飞 You3 yi4 zhi1 ying1 zai4 tian1 shang4 fei1 There is an eagle in the sky flying "There is an eagle flying in the sky"
high-surprisal sentence	b)* 有 一只 鹰 在 天上 肥* You3 yi4 zhi1 ying1 zai4 tian1 shang4 fei2* There is an eagle in the sky gaining weight* "There is an eagle gaining weight in the sky"

Procedure
(online *Gorilla* Experiment builder, Anwyl-Irvine et al., 2018)

Familiarization phase
- "***Does this sentence make sense***?" and clicked "**yes**" or "**no**" on the screen after hearing the whole sentence (stimuli: two pairs of sentences in Standard Mandarin)
- 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

Data Analysis
Accuracy: expected judgment on sentence plausibility counted as correct
  - "Yes" responses to low surprisal (i.e., plausible) sentences
  - "No" responses to high surprisal (i.e., implausible) ones

Response times: the interval between the end of the audio file and the click registering a judgment

The analysis included a total of 920 trials (range of 43–44 per participant)

Results

Statistical models
Accuracy: Bayesian logistic mixed-effects regression
Response time: Bayesian log-normal mixed-effects regression
both with weakly informative priors (Bürkner, 2018)

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

Credible main effects of *surprisal, dialect* and *the 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

No credible main effects of trial number and its interactions with surprisal and dialect: accuracy did not reliably improve in any condition across the course of the experiment
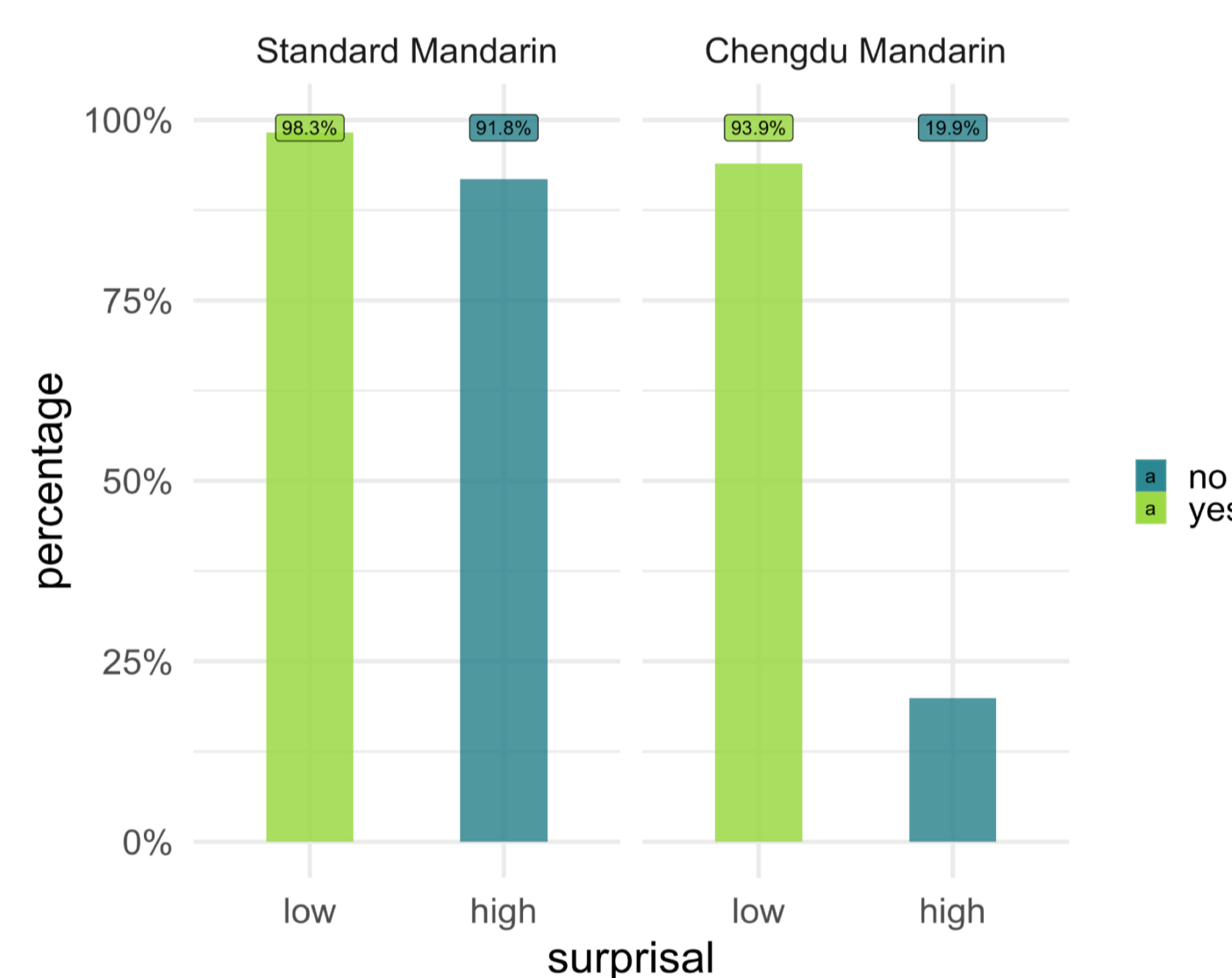

Figure 2: Percentage of "correct" responses across dialect and surprisal conditions

Response time

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

Surprisal: high-surprisal >> low-surprisal condition
Dialect: Chengdu Mandarin >> Standard Mandarin
Difference between high- and low-surprisal:
Standard Mandarin (estimated mean difference = 800 ms) >> Chengdu Mandarin (300 ms)

No credible main effects of trial number and its interactions with surprisal and dialect; credible interaction of trial, surprisal and dialect possibly driven by a reliable slowdown in Chengdu high-surprisal condition
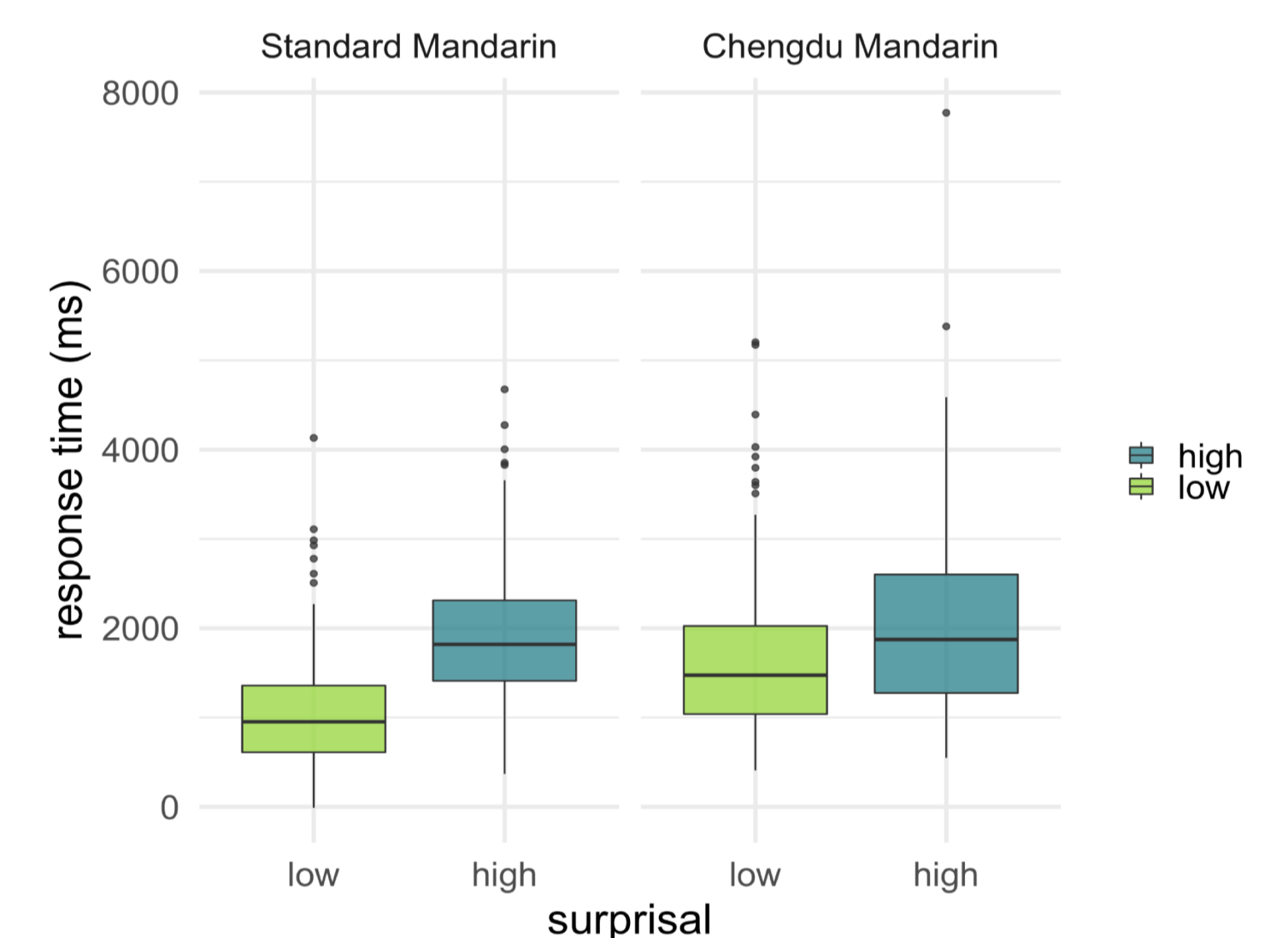

Figure 3: Response times across dialect and surprisal conditions.

Discussion

Accuracy results

Familiar tone system (Standard Mandarin)
- High accuracy over 90%, as expected
- The participants understood the task in general (sensitive to the sentence surprisal)

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

Unfamiliar tone system (Chengdu Mandarin)
- Bias to respond "semantically plausible"/ "yes":
  - Major bottom-up *failure* in identifying tone mismatch in high-surprisal condition
  - Relatively high intelligibility of this non-native dialect

Top-down influence overriding bottom-up tone acoustics

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 results

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

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

Both bottom-up and top-down processing are present
  - Processing tone information in the unfamiliar tone system (Chengdu Mandarin)!

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

Conclusion

Findings:

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

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

Familiar tone system perception (Standard Mandarin):
Strong representations of tones and segments (bottom-up)
Relatively high sensitivity to surprisal tone mismatch (top-down)

Unfamiliar tone system perception (Chengdu Mandarin):
Biased plausible perception overlooking tone mismatch (overriding top-down)
Certain sensitivity to surprisal tone mismatch (bottom-up)

Rapid learning of the unfamiliar tone system: listeners seem to be 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.