Building Speech Databases of Mandarin Chinese

Hongwei Ding

Speech-Language-Hearing Center, School of Foreign Languages
Shanghai Jiao Tong University, China

LDC Virtual Workshop for Penn China Research
November 9th & 10th

1. **Introduction**

❖ Spoken version of Chinese Treebank

  o Collaborated with LDC

  o Finished

  o Preliminary research work

❖ Speech Database for mental disorders

  o One part of Major Program of National Social Science Foundation of China

  o Under construction

  o Seek advice for improvements

❑ Background

- Proposed and guided by Mark Liberman
- Text materials selected by Jiahong Yuan from Chinese Treebank
- Recorded by SJTU team by Hongwei Ding
  - Recruited more than 20 speakers (Class 2 Level 1 or better on *Putonghua Shuiping Ceshi* (national standard Mandarin proficiency test))
  - Only 16 speakers finished (100 hours)
- Automatically aligned by assistants of Mark Liberman
- Ready for research work

❑ Potential Investigations

- To learn correlation between syntactic analysis and prosodic features
- To improve parsing by adding speech-derived features
- To predict prosody from text for TTS
- To study phonetics and phonology of Mandarin Chinese

❏ **Investigation of Tones in Text**
- o Word segmentation
  - • Automatically conducted with tools
  - • Manually corrected
  - • Compared with results in Treebank (will be done)
- o Text-to-pinyin conversion
  - ▪ Automatically conducted with tools
  - ▪ Manually corrected
- o Measurement of occurrences of tones in text
  - o Each tone in isolation → Tone 4 most frequent

    (Hou & Huang, 2020)
  - o Disyllabic tones → 2nd syllable is usually Tone 4 with 4+4 as most frequent

❑ **Questions**

❖ Why does Tone 4 appear so frequently in Mandarin Chinese?

➢ Possible explanations

- Tone 4 is distinctive from other Tones
- Tone 4 is easy to segment when heard in sequence

❑ **Evidence from other Investigations**

❖ Tone 4 (high-falling) → most salient for Mandarin Chinese

a) Tone 4 alone can produce observed language differences between English and Mandarin Chinese (Keating & Kuo, 2012)

b) **Tone 4 is also favored by Mandarin Chinese speakers to produce alphabets in alphabetical words (Ding et al. 2017, 2020)**

c) **Tone 4 is also responsible for Chinese-accented German and English (Ding et al. 2012a)**

d) Tone 4 is specially difficult for German learners of Mandarin Chinese (Ding et al. 2012b)

❑ **Tonal Adaptation of Disyllabic Letter-Character Pattern in Mandarin Alphabetical Words (MAW)** (Ding et al. 2020)
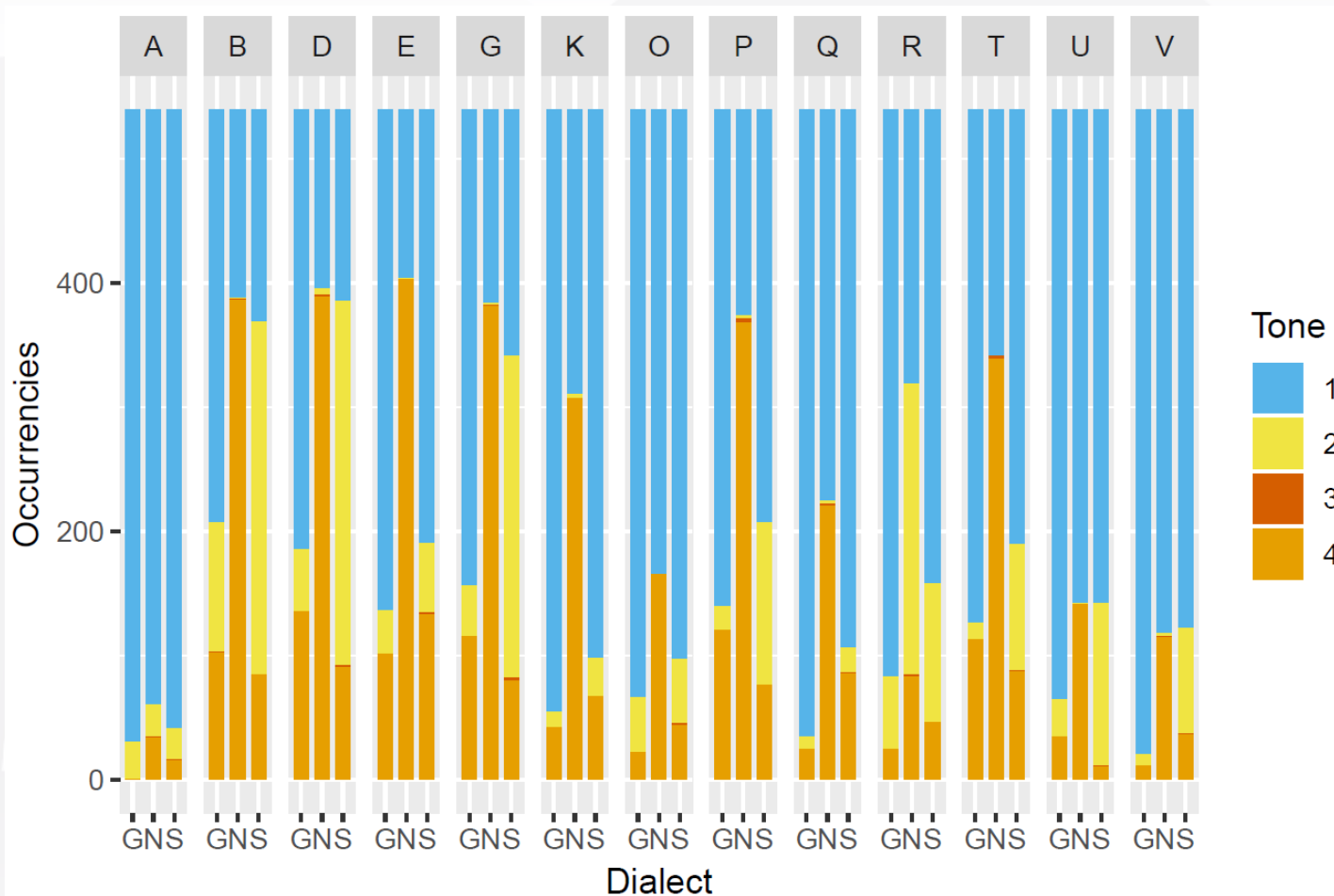

Figure 1: Number of preferred tonal choices for each letter across different dialects in all tasks.

MAW: L+C (Letter + Character) Adaption to Chinese phonological system is most probable

G: Guangdong
N: Northern China
S: Shanghai

For northern speakers, Tone 4 exceeded 60% for 6 (B, D, E, G, P and T) out of 13 letters, which end with front vowel /i/

❑ **A Phonetic Investigation of Intonational Foreign Accent in Mandarin Chinese Learners of German** (Ding et al. 2012)

❖ German truncates falling accents and falls do not become steeper but simply end earlier → High falling tones are rare in German speech. (Grabe, 1998)

❖ Mandarin L2 learners tend to use high-falling pitch accent to stress syllables, which is not popular in German.

❖ German sentences with high-falling tones (similar to Tone 4) produced by Mandarin learners are regarded as having Chinese-accent.

❑ **Extend Study from Text to Speech**

❖ Acoustics in phonetics and phonology

○ Syllable level (esp. Tone 4)

○ Word level
- Disyllabic words (e.g. Sandhi rules)
- Trisyllabic words (e.g. 2+1, 1+2)
- Multisyllabic words
- Neutral tones
- ...

○ Phrase level

○ Utterance level

1. **Recordings from mental disorders**
   - ❖ Patients include
     - o schizophrenia
     - o depression
     - o bipolar disorder
     - o anxiety disorder
     - o Alzheimer
   - ❖ Materials include
     - o Clinical interviews
     - o Reading
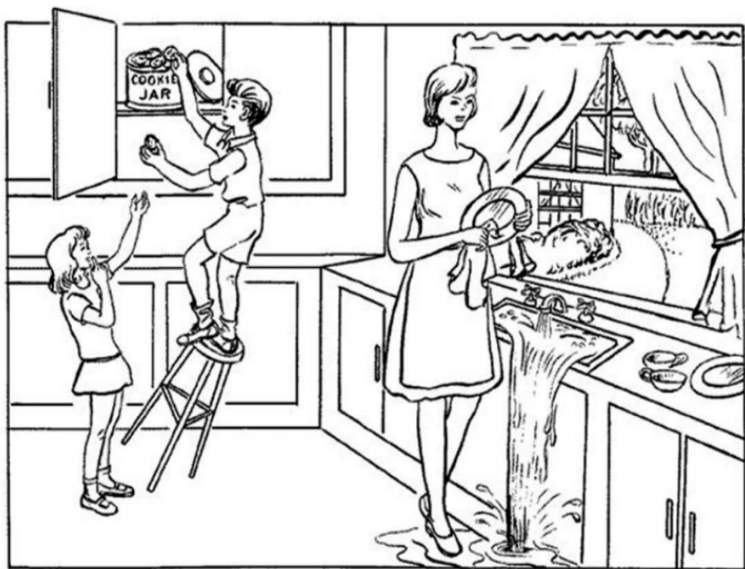   - o Speaking

2. **Recordings from controlled groups**

❑ **Recordings from controlled groups (our plan)**

❖ Materials include

o Questions similar to clinical interviews

o Reading of prosodic passages

o Speaking (picture description)

❖ Time

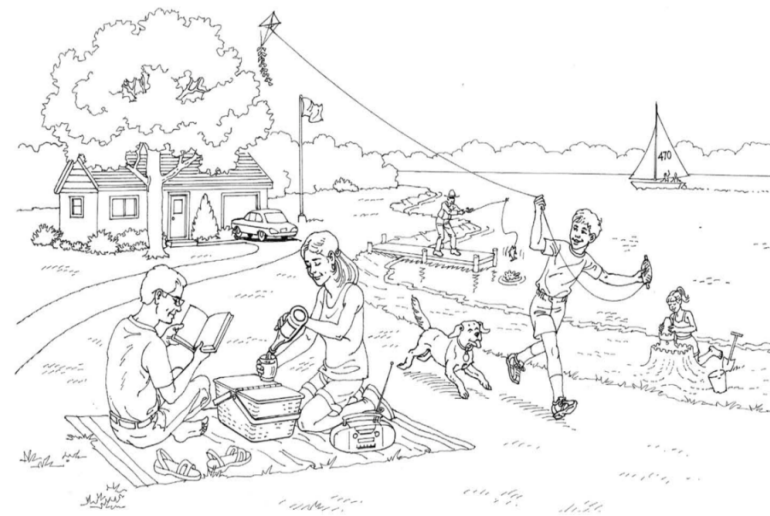o One hour per speaker

❖ Materials

o Same for all controlled groups

❑ **Picture description**

The 'cookie-theft' picture

From the Boston Diagnostic Aphasia Examination - Goodglass & Kaplan, 1983

Extract from the Western Aphasia Battery, Revised (WAB-R). Copyright (C) 2006 NCS Pearson, Inc. Used with permission. All rights reserved. 56
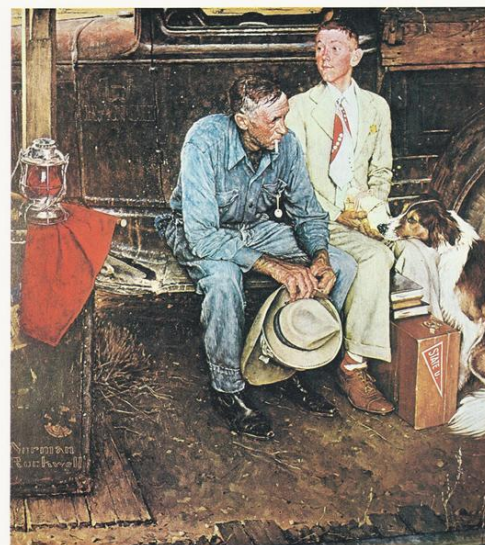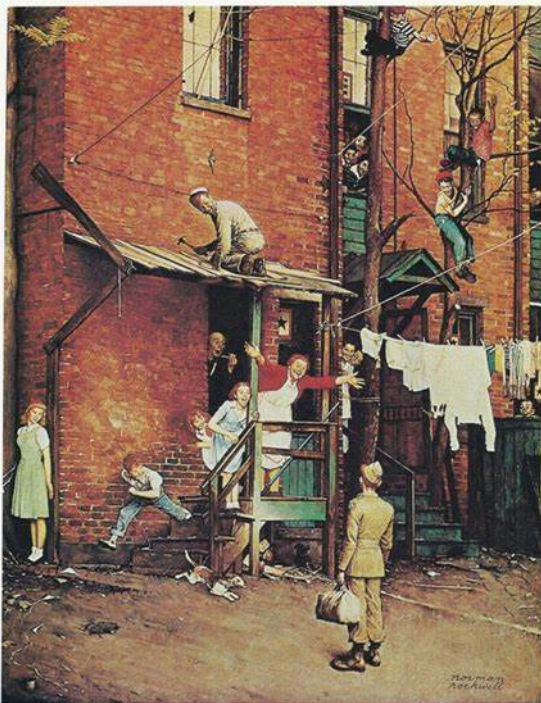
Figure 3. The "Picnic Scene" from the Western Aphasia Battery, Revised (WAB-R). Copyright c 2006 NCS Pearson, Inc. Used with permission. All rights reserved.

➢ Yes, they have experiences

❑ **Picture description**

The Runaway, 1958

➢ No, little experience

❑ **Questions**

❖ Time
  o One hour per speaker?

❖ Materials
  o Same for all controlled groups?

❖ Picture description?
  o Better pictures (with Chinese elements, and for specific ages) ?
  o New pictures from Mark Liberman

❖ Mode
  o Audio only or Audio + Visual?

Much of the work has been collaborated with

Linguistic Data Consortium
University of Pennsylvania

Department of Chinese and Bilingual Studies,
Hong Kong Polytechnic University

Thanks to all those who have contributed in the work

[1] Ding, H. and Y. Li: Tonal Adaptation of Disyllabic Letter-Character Pattern in Mandarin Alphabetical Words. In: Speech Prosody 2020. DOI: 10.21437/SpeechProsody.2020-101

[2] Ding, H., Y. Zhang, H. Liu and C.-R. Huang: A Preliminary Phonetic Investigation of Alphabetic Words in Mandarin Chinese. In: Interspeech 2017, p.3028-3032.

[3] Ding, H., O. Jokisch and R. Hoffmann: A Phonetic Investigation of Intonational Foreign Accent in Mandarin Chinese Learners of German. In: Speech Prosody 2012, p.374-377

[4] Ding, H., R. Hoffmann, H. Hussein and O. Jokisch: An Acoustic Investigation of Chinese Tone 4 Produced by German Learners. In: IALP 2012. IEEE, p.5-8.

[5] Grabe, E.: Pitch accent realization in English and German. In: Journal of Phonetics, 26(2): 129-143, 1998.

[6] Hou, R. and Huang, C.-R.: Robust stylometric analysis and author attribution based on tones and rimes. In: Natural Language Engineering, vol. 26, p. 49–71, 2020.

[7] Keating, P. and Kuo, G.: Comparison of speaking fundamental frequency in English and Mandarin. In: The Journal of the Acoustical Society of America, vol. 132, no. 2, pp. 1050–1060, 2012.