The ManDi Corpus: A Spoken Corpus of Mandarin Regional Dialects

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

Introduction

Recent global developments ranging from COVID-19 to climate change have triggered a comprehensive re-evaluation of our approach to speech data collection, from the traditional lab set-up to remote date collection.

Remote audio collection: audio collection delivered virtually with participant-controlled recording process using available personal devices

Essential to remote audio collection: reasonable control of the potential variability introduced other than speech itself (Leemann et al., 2020)

e.g. recording environment, recording devices and uncertainty in implementation is essential to remote audio collection, etc.

- Different findings on the influence of recording devices on f0 values and vowel formants (De Decker & Nycz 2011, Grillo et al. 2016)
- General assumption of possible uncertainty in absolute values of acoustic measurements, but reliable in relative patterns

Mandarin Chinese corpus:

- Mostly targeted Standard Mandarin speech (e.g. CALLFRIEND, Canavan and Zipperlen, 1996; ALLSSTAR, Bradlow),
- Mandarin-branch dialects resources remain scarce despite the fact that they are spoken by over 70% of the population.

Goals:

- To present our methods for remote speech data collection using smartphone recording applications
- To introduce the ManDi Corpus, a spoken corpus of six Mandarin dialects (Beijing, Chengdu, Jinan, Taiyuan, Wuhan, Xi'an) and Standard Mandarin.

Figure 1. Locations of cities where the six Mandarin dialects are spoken.

Speech Data Collection

Production experiment conducted online using the *Gorilla* Experiment Builder (Anwyl-Irvine et al., 2018)

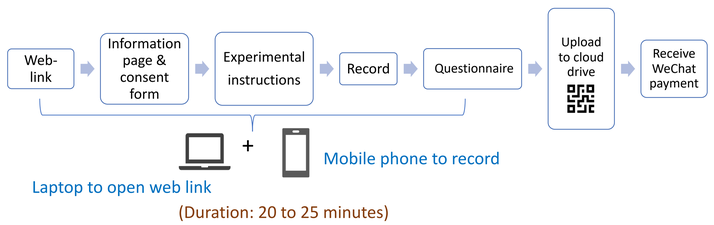
Participants: 36 native speakers of Mandarin dialects

- Beijing (9), Chengdu (5), Jinan (5), Taiyuan (7), Wuhan (6), Xi'an (4)

Reading materials:

- Word list 1: 40 monosyllabic words (10 unique syllables × 4 tones)
- Word list 2: 20 disyllable words (4 tone categories for the first syllable × 5 tone for the second syllable)
- Short sentences: 24 pairs of semantically plausible & implausible sentences
  - Implausibility was created by altering the tone of one target word in either the sentence-final or –medial position.
- *North Wind and the Sun* passage: script translated in Standard Mandarin
- *Wo Chun* homophonic poem:
  - Depicting a tranquil spring scene in the original written form
  - If read aloud, can be perceived as a man mocking himself to be silly, with word-wise tones altered
- For each type of material, participants were instructed to read them either in Standard Mandarin or their native dialect. (10 tasks in total; trials randomized in each task)

Procedure:



Pilot Study

Goals:

- To measure the acoustic-phonetic realizations of lexical tones of the six Mandarin dialects
- To verify data reliability by comparing measured tone systems to previous records, especially Standard Mandarin, a well-documented variety (Ho, 2003; Figure 3)

Rational:

- Mandarin branch dialects: comparable (similar) segmental inventories, but distinct tone systems (Norman, 2003; Tang, 2017)
- Existing documentation mostly done in the traditional impressionistic approach through fieldwork and using Chao tone numerals for description (Table 2; Li, 2002, *Modern Dictionary of Chinese Dialect*).
- The current state of knowledge regarding Mandarin dialect tone systems should be updated and supported by acoustic-phonetic analysis.

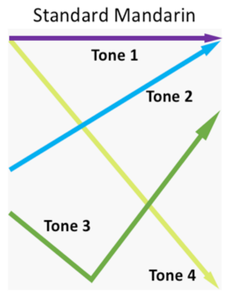


Figure 3: Schematic tone contours of Standard Mandarin.

	source	tone 1	tone 2	tone 3	tone 4
BEI	Dict.	55	35	214	51
	data	44	24	213	51
	perception	level	rising	dipping	falling
CHD	Dict.	55	21	53	213
	data	*25	31	52	212
	perception	rising	low-falling	falling	dipping
JNN	Dict.	213	42	55	21
	data	323	*55	*34	41
	perception	dipping	level	rising	falling
TYN	Dict.	11		53	45
	data	*31		51	34(2)
	perception	low-falling		falling	rising
WHN	Dict.	55	213	42	35
	data	*34	212	31	*215
	perception	rising	low-dipping	falling	dipping
XIA	Dict.	21	24	53	44
	data	21	24	41	44
	perception	low-falling	rising	falling	level

Table 2: Comparisons of tone systems between dictionary records, measured date and native speaker's perception based on the data. Differences are marked by an asterisk.

Method

- Measured F0 contours were used to represent the phonetic realization of the lexical tone (Jongman et al., 2006; Tupper et al., 2020)
- Ten equally spaced F0 values over the sonorant portion of the word and converted F0 values in hertz to semitones with the following formula (Yuan and Liberman, 2014):

$$\text{Semitone} = 12 \times log_2\left(\frac{F_0}{F_{0\_base}}\right)$$

(F_{0_base} was the speaker-specific F0 value in the 5th percentile)

- Grand mean values were calculated for each point by tone category and dialect

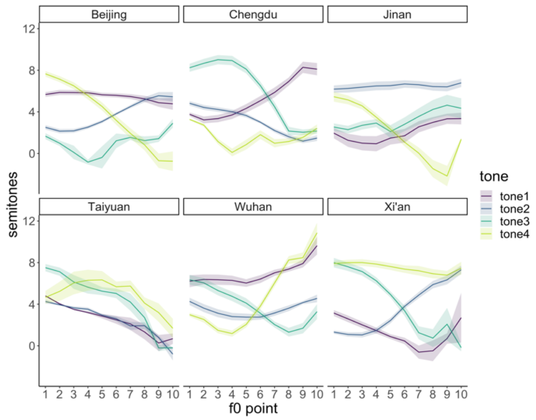


Figure 4: Tone contours of the six Mandarin dialects. Ribbon represents ±0.5 standard error from the mean.

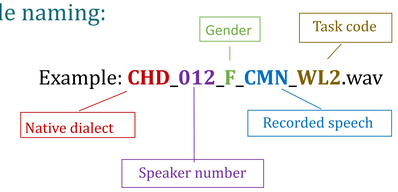
Results

- Each dialect indeed has a relatively unique acoustic-phonetic realization of the lexical tone categories (Figure 4).
- Our tone plots conformed to a large extent to the previously documented tone categories of the other dialects; particularly, the same contour patterns with Standard Mandarin
- Observed difference between measured data and previous descriptions may inform possible variation across time and community or due to the relatively small sample size.

Corpus Annotation

Altogether 357 recordings (9.6 hours) from 36 participants (ManDi corpus available on OSF https://osf.io/fgv4w/).

File naming:



Transcripts:

- 317 transcripts (one transcript per speaker per task) were generated using a R script from Gorilla
- Missing transcripts due to Gorilla system error or unstable internet connection

Forced alignment

- Utterance alignments first created by a Praat script
- Word-level and phone-level annotation automatically created by running Montreal Forced Aligner (MFA) (McAuliffe, et al., 2017).
- Annotations of Word list 1 recordings were manually checked

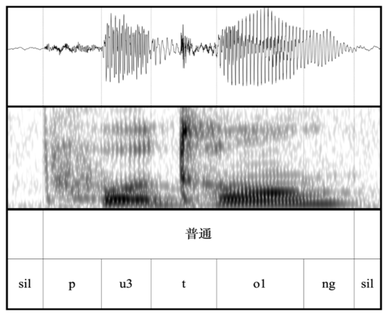


Figure 2: Part of the WAV file for the disyllabic word "普通" <pu3 tong1> and its corresponding TextGrid in Praat.

Conclusion & Suggestions

The pilot study of dialect-specific tone systems showed that with practicable design and decent recording quality, remotely collected speech data can be suitable for analysis of relative patterns in acoustic-phonetic realization.

Some workflow for collecting audio data with a basic set-up and reliable recording quality would be worthwhile.

- Experiment instruction: video demonstration in addition to written instruction
- Supervision: Pre-registered time slot to receive immediate response from the researcher if needed, but not necessarily real-time supervision of the whole experiment
- Recording: Separate recordings for different tasks or different types of stimuli, which makes it easier for data processing.
- Data missing: To avoid overriding data from multiple attempts, participants were expected to complete all the production tasks in one attempt instead of several attempts on different days.
- File uploading: Multiple options for participants to share, upload or send the files can also be helpful in case of technical difficulties.