

# The ManDi Corpus: A Spoken Corpus of Mandarin Regional Dialects

Liang Zhao<sup>1</sup> and Eleanor Chodroff<sup>1</sup>

University of York<sup>1</sup>

YO10 5DD, Heslington, York, UK

liang.zhao@york.ac.uk, eleanor.chodroff@york.ac.uk

## Abstract

In the present paper, we introduce the ManDi Corpus, a spoken corpus of regional Mandarin dialects and Standard Mandarin. The current version of the corpus contains 357 recordings (about 9.6 hours) of monosyllabic words, disyllabic words, short sentences, a short passage and a poem, each produced in Standard Mandarin and in one of six regional Mandarin dialects: Beijing, Chengdu, Jinan, Taiyuan, Wuhan, and Xi'an Mandarin from 36 speakers. The corpus was collected remotely using participant-controlled smartphone recording apps. Word- and phone-level alignments were generated using Praat and the Montreal Forced Aligner. 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. The corpus is available on OSF (<https://osf.io/fgv4w/>) for non-commercial use under a CC BY-NC 3.0 license.

**Keywords:** speech, Mandarin Chinese dialects, remote data collection, smartphone recordings

## 1. Introduction

Speech science has traditionally relied on high quality speech recordings, collected in person with a standardized recording setup in laboratories/studios or in field settings. The strictly controlled environment for speech recording allows the acoustic-phonetic analysis to be focused on differences in the speech alone. However, recent global developments ranging from COVID-19 to climate change have triggered a comprehensive re-evaluation of our approach to research and travel, while researchers have drastically rethought data collection approaches and turned to varieties of remote speech data collection. Remote audio collection hereby refers to the scenario where the researcher delivers an experiment to a participant virtually and remotely, who in turn records his/her speech using a personally available recording device, such as a smartphone or computer, and sends the recording back to the experimenter.

Early attempts at remote audio collection were inevitably confined to the state of technological development at the time. Previous methods in the 1990s exploited access to landline telephones to collect speech corpora (e.g. CALLHOME, Canavan and Zipperlen, 1996). This method increased the range of participant enrolment, but the recordings were limited by the telephone bandwidth, and were thus of low quality for phonetic research.

Thanks to current smartphone technology, high quality recording baseline and its widespread accessibility can advance remote audio collection for research in basic and forensic speech sciences, and speech engineering. As approximately 50% of the world's population is estimated to own a smartphone and this figure is likely to rise, remote audio collection with smartphone devices could increase both the amount and range of speech data. With a robust method, we can simultaneously reach a vast number of diverse communities around the world: anyone with access to a smartphone could contribute. Speech technologies such as speech-to-text systems and text-to-speech synthesis might also benefit from large quantities of data which allows for more precise estimates of true population parameters.

Moving forward with any type of remote data collection requires reasonable control of the potential variability

introduced by the recording environment, recording devices and uncertainty in implementation (Leemann et al., 2020). Previous work (de Decker and Nycz, 2011) tested the consistency of recording devices within a laboratory but on older technology (e.g., 2010-era mobile phones and laptops). Bird et al. (2014) investigated the use of smartphones for recording in isolated indigenous communities, but focused on the feasibility of general speech collection as opposed to between-device or environmental effects on acoustic-phonetic measures. More recently, Grillo et al. (2016) identified significant variation between smartphone devices in measurements of voice quality in a laboratory setting. Work from Leemann et al. (2020) has also investigated remote speech collection via a smartphone application, but in the same country with simultaneous videoconferencing. In contrast, we collected audio recordings from a culturally and geographically remote region (UK to China), without videoconferencing.

Speech resources for Mandarin Chinese have been substantially expanded over the past few decades through telephone conversation collection (e.g. CALLFRIEND, Canavan and Zipperlen, 1996) and lab recording (e.g. ALLSTAR, Bradlow), but most related corpora have primarily collected speech data in Standard Mandarin and only a few contained regional varieties as part of the research (e.g. NCCU, Chui and Lai, 2008). Although regional dialects of the Mandarin branch (in contrast with the non-Mandarin branch, such as Wu or Cantonese) are spoken by over 70% of the population in mainland China, recordings of these dialects that are available for speech science research remain scarce.

In this paper, we present our methods for remote speech data collection using smartphone recording applications and describe the ManDi Corpus, a spoken corpus of six Mandarin dialects and Standard Mandarin. The six Mandarin dialects (Beijing, Chengdu, Jinan, Taiyuan, Wuhan, Xi'an) each represent a mainstream variety spoken in a major province of mainland China. Figure 1 demonstrates the approximate locations of cities where the six dialects are spoken.



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

## 2. Speech Data Collection

A speech production experiment was created and hosted online using the *Gorilla* Experiment Builder ([www.gorilla.sc](http://www.gorilla.sc)) (Anwyl-Irvine et al., 2018). Speech data were collected between August and October 2020. The participants were required to read through the given material presented through the *Gorilla* web-based project in both Standard Mandarin and their regional dialect that was one of the six dialects—Beijing, Chengdu, Jinan, Taiyuan, Wuhan, and Xi'an dialect. They simultaneously recorded themselves using the designated smartphone recording application and then uploaded recordings to a cloud drive folder upon finishing all the tasks.

### 2.1 Participants

Thirty-six speakers have thus far been recruited for the corpus. The participants were native speakers of one of the six dialects (Table 1), aged between 18 and 45, proficient in Standard Mandarin and literate in the Standard Simplified Chinese writing system with no auditory impairment or reading difficulty. The participants were expected to be digitally literate to complete the online experiment. More participants will be recruited to balance genders and dialects.

	Beijing	Chengdu	Jinan	Taiyuan	Wuhan	Xi'an
M	3	1	2	3	2	2
F	6	4	3	4	4	2
Total	9	5	5	7	6	4

Table 1: Participants in the speech production experiment

### 2.2 Reading Materials

Five sets of reading materials were constructed for the study: monosyllabic words (Word List 1), disyllabic words (Word List 2), short sentences, the North Wind and the Sun passage, and a modern Chinese poem, *Wo Chun* (“卧春”). Each type of material was read first in Standard Mandarin and then again in the participant's dialect. There were altogether ten tasks for the five sets of reading material per participant.

The list of monosyllabic words consisted of 40 unique characters that were constructed by pairing ten unique monosyllables each with one of the four lexical tone categories in Standard Mandarin (10 syllables  $\times$  4 tone categories). The chosen monosyllables were the ones that

allow realizations of all four tone categories. For example, the syllable “ma” (in Pinyin) could be realized as “ma1”, “ma2”, “ma3”, “ma4” (numbers as Chao tone numerals). The neutral tone was excluded in this task as it does not typically occur in monosyllabic words in isolation. Production of lexical tones carried by monosyllabic words is particularly useful for mapping out the tone system of a given dialect as it avoids tonal variation due to contextual influence (Xu, 1997). Four of the syllables had a CVC template, four a CV template and two a V template.

Twenty disyllabic words were included in the disyllabic word list. Each word represented a unique combination of lexical tones over two syllables. For disyllabic words, the second syllable can optionally be assigned a neutral tone, so there were altogether 20 tone combinations (4 tone categories for the first syllable  $\times$  5 tone categories for the second syllable). Again, all the syllables in the disyllabic words allowed a full set of tone realizations in Standard Mandarin.

The short sentences were originally designed as stimuli for a perception experiment on the effect of semantic plausibility on tone perception. There were 24 pairs of sentences each containing a target word at either sentence-final (12 pairs of sentences) or sentence-medial position (12 pairs of sentences). Each pair of sentences shared the same sentence structure with the only difference in the tone realization of one target word. Specifically, one sentence from the pair was grammatically correct and semantically plausible, while the other sentence altered the tone category of the target word (sentence-final or sentence-medial), thus making the sentence semantically implausible. Overall, half of the sentences were semantically plausible, and half were semantically implausible.

The passage material in the experiment was the short story, *The North Wind and the Sun*. This was to provide recordings of the six dialects as connected speech. The Standard Mandarin translation of the story was used for recording in both Standard Mandarin and regional dialects as the words in Standard Mandarin are to a large extent shared in the lexicon of those dialects. One native speaker of each dialect helped to check the materials before the experiment.

The poem was adapted from a modern poem by Chinese novelist, Han Han. All the characters used in the poem were homophonic in Standard Mandarin: by the written form, the poem depicted a tranquil scene of early spring, but if read aloud with the same segmental sequence and a different tonal sequence, the poem could be perceived as a monologue of a man mocking himself for being silly. We recorded the original spring scene version of the poem in both Standard Mandarin and the regional dialects as potential stimuli for future studies on speech perception.

### 2.3 Procedure

Once recruited, the participants received a web link to the experiment. After a brief introduction to the study and presentation of the consent form, participants were given a detailed set of instructions for completing the experiment and at-home recording using smartphone devices. It was recommended that they use an extra device, such as a desktop, laptop, or tablet to display the *Gorilla* page with the reading materials and meanwhile use a smartphone to make the recordings.

As preparation, the participants attended to specific requirements of the recording environment and operational

instructions. If possible, the recording was to be made in a quiet room with plenty of soft furnishing to reduce reverberation, such as a bedroom, and with the doors and windows closed. The recording device, the participant's own smartphone, was to be placed on a hard surface, such as a table or a desk for better recording quality. Participants were also suggested to maintain a distance of approximately 20 to 30 cm from the microphone and to keep this distance throughout the experiment.

To prepare the recording device, participants installed a freely available sound recording app on their smartphones. iPhone users installed the app called Awesome Voice Recorder by Newkline Co., Ltd. with version 8.0.4 or later. Users of Android devices installed the app called ASR recorder by NLL APPS. Participants were then instructed to open the “settings” menu from AVR on iPhone or ASR on Android device, and set the following entries as specified and leave other unspecified settings by default: file format as WAV files, encode quality at medium, sample rate at 44100 Hz, bit rate at 128kbps, and channels as stereo. (These were later converted to mono recordings.) These two apps were chosen because of their similar setting options so that the settings for both apps were kept identical.

In addition, a unique and anonymous experiment ID was collected from each participant to link the recording to the transcript from Gorilla. This was done to minimize the use of any identifiable information in the file naming convention.

In the practice trials, the participants were asked to read aloud an additional list of 10 disyllabic words presented in characters in Standard Mandarin, record themselves, check the recording quality and contact the experimenter if needed. The practice trials were not included in the final corpus. After practice, the participants started recording in sequence—Word List 1, Word List 2, sentences, the passage, and the poem. For each type of material, they were instructed to read the given material first in Standard Mandarin as one task and then in their regional dialect as another task for a total of ten tasks.

All items were presented in characters at the center of the webpage. For the word lists and sentences, participants proceeded through each item by clicking the “next” button on the screen. These items were fully randomized. For the passage and poem, the text was presented on one screen. The participants were instructed to make an individual recording for each task so that the size of each file was easy to upload and send. At the start of each task, participants began a new recording, and then read aloud each item. By the end of the task, the participants saved the recording and named the file as “experiment-ID\_task-number” (e.g. lz0916x1h\_01). The task number was provided at the end of each task.

After completing all recording tasks, the participant uploaded the ten WAV files by scanning a QR code that directed to a Tencent (Weiyun) cloud folder. Files could then be uploaded anonymously and were accessible only to the experimenter. In addition to the recording files, the participants also uploaded a special QR payment code generated from WeChat Pay—a secure payment app widely used in China and similar in function to PayPal. We used this QR payment code to make a payment to the participants. For a full recording (completion of all tasks and expected quality of the recording), the participant was compensated with ¥45 CNY (approximately £5). For

receipt of any partial recording (partial completion of the task or very poor recording quality due to technical issues), the participant received ¥30 CNY (approx. £2.70). Thirty-five participants thus far have provided full recordings; one participant provided nine recordings out of ten. The uploaded recordings were then downloaded, immediately de-identified, and transferred to a secure folder on a password-protected computer owned by the university.

A post-experiment survey was then conducted to record the participants' dialectal background, demographic information, devices used for recording and general feedback on the experiment. For dialectal background, we recorded the participants' regional dialect to include both city- and county-level information; the participants also rated their fluency and proficiency in speaking Standard Mandarin and the regional dialect, and how frequently the regional dialect was used. For demographic information, the participants reported gender, age, and educational level. Information on the recording device included device type (mostly Huawei and iPhone) and how long the device has been in usage. Feedback questions were presented as rating scales for the participant's opinion on the experimental design, plus one optional text box for suggestions. In addition, *Gorilla* automatically collected information on local timestamps and the devices that were used to open the web link, device information including device type, operating system and browser.

### 3. Corpus Annotation

#### 3.1 File Naming

Before annotation, the collected recordings were renamed with the following format: native dialect, speaker sequential number, gender, recorded dialect, and task, each separated by an underscore. The native dialect codes were BEI for Beijing dialect, CHD for Chengdu, JNN for Jinan, TYN for Taiyuan, WHN for Wuhan, and XIA for Xi'an. The speaker sequential number was a three-digit code automatically generated by Gorilla in the temporal order of participation in the experiment, e.g. the first participant was 001. Gender was coded as F for female speakers and M for male speakers. The recorded dialect referred to whether the material was recorded in Standard Mandarin, coded as CMN, or the participant's own dialect, coded the same way as for the native dialect. Tasks were coded as WL1 for the monosyllabic word list, WL2 for the disyllabic word list, SST for sentences, NWS for *North Wind and the Sun*, and WCH for the poem, *Wo Chun*.

An example file name with the .wav extension is as follows: CHD\_012\_F\_CMN\_WL2.wav, meaning this is a recording of Word List 2 (disyllabic words) spoken in the Standard Mandarin dialect by a female native Chengdu dialect speaker who was the 12<sup>th</sup> participant in the experiment.

#### 3.2 Transcripts

A total of 317 transcripts were obtained from Gorilla. A few transcripts were missing or did not match the content of the recording. This data loss arose from a few issues. If a participant restarted the experiment at any point, then Gorilla regenerated transcripts for all tasks, meaning any previously completed task was overwritten. We also speculate some data loss was due to internet instability during the experiment or other potential technical errors from Gorilla.

### 3.3 Forced Alignment

All recordings were force aligned at the utterance-, word- and phone-levels using a combination of Praat (Boersma and Weenink, 2020) and the Montreal Forced Aligner (MFA) (McAuliffe, et al., 2017).

The recording-level transcripts were first aligned to the audio at the utterance level using a custom-made Praat script and then manually checked. All recordings were aligned using the pretrained Mandarin acoustic model released with the MFA. This model was trained on the Mandarin subset of the GlobalPhone corpus. Mandarin pronunciation dictionaries were created using the corresponding pretrained Mandarin G2P model, also released with the MFA. These were then manually corrected by a native speaker. The dialect recordings were aligned using MFA v1.0 and the Standard Mandarin recordings using MFA v2.0.0b9. Figure 2 shows an example output TextGrid with word- and phone-level annotations from the MFA along with its corresponding WAV file.

We manually checked and corrected misaligned sonorant boundaries for the recordings of monosyllabic words (Word List 1). Additional manual corrections of the TextGrid alignments will be uploaded and documented on the OSF website.

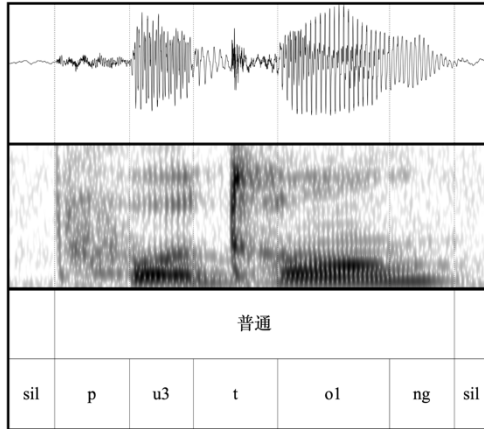


Figure 2: Part of the WAV file for the disyllabic word “普通” <pu3 tong1> and its corresponding TextGrid in Praat.

## 4. Pilot Study

The recording quality in general was sufficiently suitable for acoustic-phonetic analysis. Still, we spotted several acoustic conditions and background noise in some of the recordings. These included mouse clicking, unexpected ringtones or audible vibration from the smartphone device, insect noise, e.g. crickets, and clock ticking. To test whether the remotely collected speech data was nevertheless reliable for acoustic-phonetic analysis, we investigated the acoustic-phonetic realizations of lexical tone in each of the six Mandarin dialects, and compared these to previous descriptions of the tone systems. We especially rely on the Beijing analysis for validation purposes as this dialect most directly correlates to the well-documented Standard Mandarin system (Ho, 2003). The schematic tone contours of Standard Mandarin are shown in Figure 3.

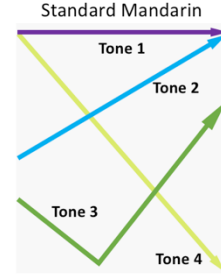


Figure 3: Schematic tone contours of Standard Mandarin.

Beijing, Chengdu, Jinan, Taiyuan, Wuhan and Xi'an all belong to the group of Mandarin branch dialects which have distinct tonal realizations and comparable segmental inventories (Norman, 2003; Tang, 2017). With the exception of Taiyuan, each dialect has the same four phonological tone categories; Taiyuan has a merger between Tone 1 and Tone 2. Existing documentation of these tone systems was mostly done in the traditional impressionistic approach through fieldwork and described using the Chao tone numeral system (Table 2; Li, 2002, *Modern Dictionary of Chinese Dialect*). As speech communities constantly interact and develop, the current state of knowledge regarding Mandarin dialect tone systems should be updated and supported by acoustic-phonetic analysis.

Since tonal realizations are largely affected by the adjacent tonal contexts in terms of anticipatory effects and carry-over effects (Xu, 1994, 1997), we focused our analysis on the monosyllabic word list production data. F0 contours were used to represent the phonetic realization of the lexical tone (Jongman et al., 2006; Tupper et al., 2020). We automatically extracted ten equally spaced F0 values over the sonorant portion of the word and converted F0 values in hertz to semitones. Semitones were calculated with the following formula (Yuan and Liberman, 2014):

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

where  $F_{0\_base}$  was the speaker-specific F0 value in the 5<sup>th</sup> percentile. For each of the ten points, the by-speaker mean was calculated; the grand mean and standard error were then derived over speakers in each dialect group. As shown in Figure 4, each dialect indeed has a relatively unique acoustic-phonetic realization of the lexical tone categories.

One straightforward way to examine whether the measurements from our corpus data were reliable for acoustic-phonetic analysis was to compare our plot of Beijing tone system with Standard Mandarin which are supposed to have the same phonological categories and similar pitch contours. The measured tone categories of Beijing dialect in the top-left plot in Figure 4 showed the same relative patterns as in Standard Mandarin (Figure 3) with tone 1 as a level tone, tone 2 a rising tone, tone 3 a dipping tone, and tone 4 a falling tone. As we were able to map out the tone patterns of Beijing dialect using smartphone recordings, we would expect to reliably capture the tone inventories of the other dialects as well.

In fact, our tone plots conformed to a large extent to the previously documented tone categories of the other dialects and provided visualized acoustic-phonetic detail of the tone

systems, which may help to update Chao tone numerals for current speech. Table 2 lists the tone systems of the six dialects in Chao tone numerals from the *Modern Dictionary of Chinese Dialect* (Li, 2002), compared with the measured data in Figure 4 and the experimenters' perception. There were a few cases (marked by an asterisk) where the general contour patterns from data did not match previous records. This might be attributed to variation across time and community or due to the relatively small sample size.

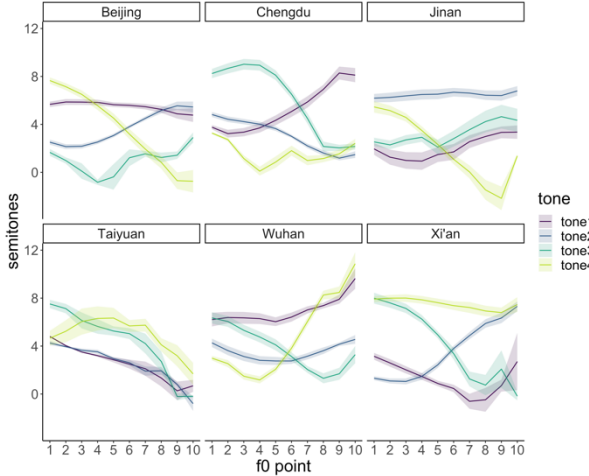


Figure 4: Tone contours of the six Mandarin dialects. Ribbon represents  $\pm 0.5$  standard error from the mean.

## 5. Suggestions for Future Studies

In collecting speech data for acoustic-phonetic analysis, it is often case that not all researchers have access to or resources to build software or apps specifically for audio data collection, such as *The English Dialect App* (Leemann et al. 2018) and *Swiss Voice App* (Kolly et al. 2014). Some workflow for collecting audio data with a basic set-up and reliable recording quality would be worthwhile. Based on some difficulties encountered during data collection and data processing, we present a few suggestions for future corpus collection, particularly when collecting larger amounts of speech data and when using existing recording apps or other devices.

In our design, the instructions were presented in the written form through a web link; we would instead suggest using a video demonstration either in addition to or even instead of written text. As for communication with participants, although we did not necessarily need real-time supervision of the speech production, it would be good that participants get immediate assistance if needed. We do recommend that an experimenter be available for contact during the experiment.

In the recording process, we would suggest making separate recordings for different tasks or different types of stimuli, which makes it easier for data processing. Also, individual audios with smaller size are easier to send or upload. In additional, participants were expected to complete all the production tasks in one attempt instead of several attempts on different days. This, to some extent, reduces variability from speakers themselves, and it also helps avoid potential missing data. For collecting recordings, multiple options for participants to share, upload or send the files can also be helpful in case of technical difficulties.

	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, measurements from ManDi corpus data and native speaker's perception based on the data. Deviations in general contour patterns between our data and the dictionary are marked by an asterisk.

## 6. Conclusion

We presented an annotated spoken corpus of six Mandarin dialects for acoustic-phonetic analysis. The corpus consists of 357 recordings of monosyllabic and disyllabic words, and connected speech including short sentences, a short passage and a poem, each recorded in Standard Mandarin and in one of six regional Mandarin dialects. The speech data for the corpus were collected remotely using participant-controlled smartphone recording apps. The corpus was further aligned using Praat and the Montreal Forced Aligner. 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.

## 7. Acknowledgements

We would like to thank Paul Foulkes and Phil Harrison for helpful suggestions and discussion on remote audio collection.

## 8. Bibliographical References

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., and Evershed, J. (2018). Gorillas in our Midst: *Gorilla.sc*, a new web-based Experiment Builder. *bioRxiv*, 438242.



- Bird, S., Gawne, L., Gelbart, K., and McAlister, I. (2014, August). Collecting bilingual audio in remote indigenous communities. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers (pp. 1015-1024).
- Boersma, W. and Weenink, D. (2020). Praat: Doing phonetics by computer [Version 6.1.16].
- De Decker, P., and Nycz, J. (2011). For the record: Which digital media can be used for sociophonetic analysis? University of Pennsylvania Working Papers in Linguistics. 17(2), 51– 59.
- Grillo, E. U., Brosious, J. N., Sorrell, S. L., and Anand, S. (2016). Influence of smartphones and software on acoustic voice measures. *International journal of telerehabilitation*, 8(2), 9.
- Ho, D. A. (2003). *The characteristics of Mandarin dialects*. The Sino-Tibetan languages, 126-130.
- Jongman, A., Wang, Y., Moore, C. B., and Sereno, J. A. (2006). *Perception and production of Mandarin Chinese tones* (pp. 209-217).
- Kolly, M. J., Leemann, A., Dellwo, V., Goldman, J. P., Hove, I., & Ibrahim, A. (2014). ‘Swiss Voice App’: A smartphone application for crowdsourcing Swiss German dialect data.
- Leemann, A., Kolly, M. J., & Britain, D. (2018). The English Dialects App: The creation of a crowdsourced dialect corpus. *Ampersand*, 5, 1-17.
- Leemann, A., Jeszenszky, P., Steiner, C., Studerus, M., and Messerli, J. (2020). Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing. *Linguistics Vanguard*, 6(s3).
- 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).
- McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., and Sonderegger, M. (2017, August). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In Interspeech (Vol. 2017, pp. 498-502).
- Norman, J. (2003). The Chinese dialects: phonology. The Sino-Tibetan Languages, 3, 72-83.
- Tang, C. (2017). Dialects of Chinese. *The Handbook of Dialectology*, 547-558.
- Tupper, P., Leung, K., Wang, Y., Jongman, A., and Sereno, J. A. (2020). Characterizing the distinctive acoustic cues of Mandarin tones. *The Journal of the Acoustical Society of America*, 147(4), 2570-2580.
- Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95(4), 2240-2253.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of phonetics*, 25(1), 61-83.
- Yuan, J., and Liberman, M. (2014). F0 declination in English and Mandarin broadcast news speech. *Speech Communication*, 65, 67-74.
- Canavan, A., and Zipperlen, G. (1996). CALLHOME Mandarin Chinese Speech LDC96S34. Philadelphia: Linguistic Data Consortium.
- Canavan, A., and Zipperlen, G. (1996). CALLFRIEND Mandarin Chinese-Mainland Dialect LDC96S55. Philadelphia: Linguistic Data Consortium.
- Chui, K., and Lai, H. L. (2008). The NCCU Corpus of Spoken Chinese: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics*, 6(2).

## 9. Language Resource References

- Bradlow, A. R. (n.d.) ALLSSTAR: Archive of L1 and L2 Scripted and Spontaneous Transcripts and Recordings. Retrieved from <https://oscar3.ling.northwestern.edu/ALLSSTARcentral#!/recordings>.