# Research Statement

*Liang Zhao*

Substantial variability exists in the phonetic realization of speech sounds across languages and dialects, on both segmental and suprasegmental level. My primary research interest is in understanding not only the extent of this variation, but also how it can be processed and potentially adapted to by listeners who are not necessarily familiar with the variation. Specifically, my doctoral project focuses on the phonetic variation of lexical tones across Mandarin dialects, seeking to understand the extent to which the phonetic properties of lexical tones vary across dialects, and how this phonetic variation is processed by non-native listeners of the dialect, and the possible conditions for adapting to an unfamiliar tone system. Speech of Mandarin dialects is used as the source of tonal variability in my current study as they have similar segmental inventories, but disparate phonetic realizations of the lexical tone. My work draws on a range of empirical methods and techniques for speech production and perception study, including corpus-based phonetic analysis, automatic acoustic measurement, and behavioural data analysis. In the following research statement, I describe my research methods and findings thus far.

## 1. To what extent the lexical tones differ in their phonetic realization across the Mandarin dialects?

Mandarin dialects are mutually intelligible regional dialects differing in their phonetic tone inventories (Tang & van Heuven 2008; Wu, Chen, Van Heuven, & Schiller, 2016b; Gooskens, 2018; Li, Best, Tyler, & Burnham, 2020). Perception of an unfamiliar Mandarin dialect essentially comes down to the task of processing an unfamiliar tone system with familiar segmentals. Before delving into the perception domain, it is crucial that we have solid knowledge of these distinct tone systems. Besides, previous records of the lexical tone systems of Mandarin dialects primarily relied on Chao tone transcriptions from field studies in the traditional impressionistic approach and lacking in acoustic-phonetic detail in temporal series. With the purposes of obtaining first-hand speech data and updating the acoustic-phonetic status of the lexical tone systems, I constructed the ManDi Corpus, a spoken corpus of regional Mandarin dialects and Standard Mandarin, and conducted acoustic-phonetic analysis of the six tone systems (Zhao & Chodroff, LREC 2022).

F0 contours were plotted to represent the phonetic realization of the lexical tones (Jongman et al., 2006; Tupper et al., 2020). As tone realizations are subject to the adjacent tonal contexts due to anticipatory and carry-over effect (Xu, 1994, 1997), F0 measurement was taken from the monosyllabic words. To normalize duration, ten equally spaced F0 values were extracted over the sonorant portion of the word and converted to semitones (formula from Yuan & Liberman, 2014), and Chao tones (formula from Shi, 2018).

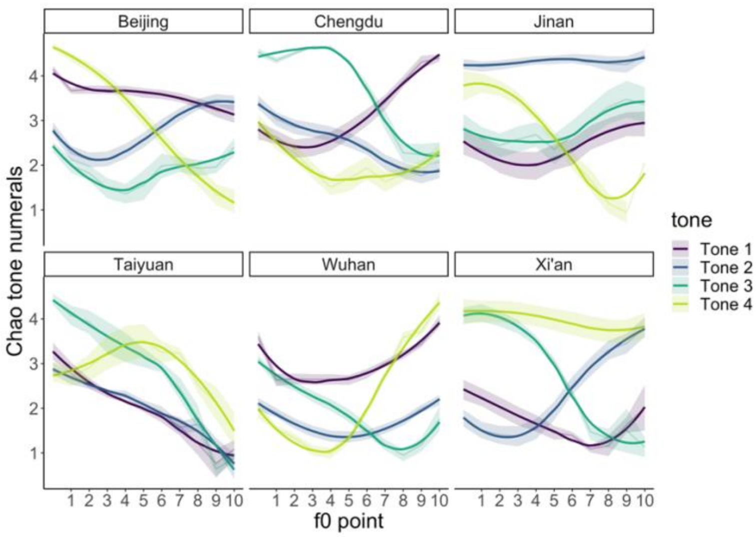


Figure 1. Smoothed mean F0 contours (converted to Chao tones) of the lexical tones in the six dialects. Ribbons represent 0.5 standard error deviation from the mean.

Figure 1 shows the dialect-specific tone contours scaled in Chao tone numerals, which are highly disparate among the dialects. In fact, many of these contours conformed to the previous descriptions in terms of the contour type, but differed in the relative pitch values (Zhao & Chodorff, 2022). For example, the textbook transcription of Tone 1 in Beijing Mandarin is [55] as a high-level tone, but it is not exactly realized with the highest values within the F0 range given the speech data and more likely a [44] tone. Updated Chao tone numerals for dialect-specific tone systems were proposed in my thesis and quantified in Levenshtein distance. Moreover, sub-parameters were also measured to capture finer F0 characteristics, including F0 onset, the turning point and ΔF0 (Moore & Jongman, 1997; Wang et al., 2003). The corpus-based phonetic analysis has important implications for the phonetics-phonology interface, as well as for generalized perceptual adaptation to novel tone systems.

**2. What are the perceptual mechanisms for processing the phonetic tone variation?**

The key question to the perception of phonetic tone variation is how listeners deal with the vast phonetic variance in naturally produced utterances and still understand the intended meaning. Several prominent theoretical frameworks of speech perception have identified two high-level mechanisms: top-down processing and bottom-up processing (McClelland & Elman, 1986; Stevens, 2002). Bottom-up processing refers to the processing of acoustic properties from the immediate incoming signal. Top-down mechanism refers to processing based on prior knowledge. My research specifies sentential context as the top-down information and tone acoustics as the bottom-up information.

While researchers generally agree that both top-down and bottom-up information types are used in lexical tone processing, there is little consensus on the relative weighting of these sources of information and how they interact. To what extent do speaker expectations guide (or indeed, override) attendance to bottom-up segmental and pitch information? To test this, I manipulated the reliability of tonal information in high and low surprisal (lexical expectedness) sentences by altering the tone category of a critical word in either sentence-medial or sentence-final position (Zhao, Sloggett & Chodroff, 2022). The high-surprisal sentences were supposed to be judged as semantically plausible with the default tone category of the critical word, while the low-surprisal sentences were semantically implausible with the tone mismatch. Apart from surprisal manipulation, the stimulus sentences also contrasted in dialect: Standard Mandarin as the familiar dialect and Chengdu Mandarin as the unfamiliar dialect to native speakers of Standard Mandarin ($N$ = 21). The experiment was hosted using *Gorilla* Experiment Builder (Anwyl-Irvine et al., 2018). Accuracy and response time were measured in the sentence semantic-plausibility judgment task, where the participants were instructed to register a "yes" or "no" response elicited by the question "*Does this sentence make sense*?" for each trial. Bayesian mixed-effects regression models were used to assess the effects of surprisal (low vs. high) and dialect (familiar vs. unfamiliar) on accuracy and response time.

For the familiar tone system (Standard Mandarin), speakers are expected to use both top-down and bottom-up information, as the sentential context and tone representations are both reliable cues for listeners. For the unfamiliar tone system, I expect potential dominance of top-down information from sentential context, and little or no use of lexical tone due to the unfamiliarity of the tone system. But the findings suggest that speakers seem to attend to bottom-up tone acoustics even if they do not always use it in determining word identity. Specifically, accuracy was high for the familiar dialect in both surprisal conditions as expected (Figure 2), indicating the validity of the surprisal manipulation, and an integrated top-down and bottom-up processing as the tone mismatch was acoustically processed and identified as leading to an implausible meaning given the sentential context. For the unfamiliar Chengdu Mandarin, sentences were judged as plausible regardless of the tone mismatch, which drastically drove down accuracy in the high-surprisal Chengdu condition where the sentences were supposed to be reported as implausible. This may suggest that the unfamiliar tone acoustics was not processed, and listeners relied on top-down information to access sentence meaning.

However, the response time result says otherwise. Consistent slowdown for high-surprisal sentences was statistically credible for both dialects (Figure 3). It was quite surprising to find that listeners responded differently between high- and low-surprisal condition for Chengdu Mandarin sentences, which suggests they were aware of the tone mismatch in the unfamiliar speech and this awareness was present as early as the experiment started as there was no reliable effect of trial. It is likely that the novel tone acoustics was processed to construct impoverished category-contour mappings to the extent that listeners' attention was modulated by high-surprisal realizations (response times), but it is not sufficient or reliable enough to be used for semantic decision (accuracy).

The major generalization is that the phonetic tone variation is processed with integrated top-down and bottom-up mechanisms for both familiar and unfamiliar tone systems. There is an overriding top-down influence on deciding the word identity that is congruent with the prior context. The bottom-up processing of the novel tone acoustics is present early on as the speech signal becomes available, which strongly suggests a rapid adaptation to the unfamiliar tone system even without explicit exposure to the dialect.

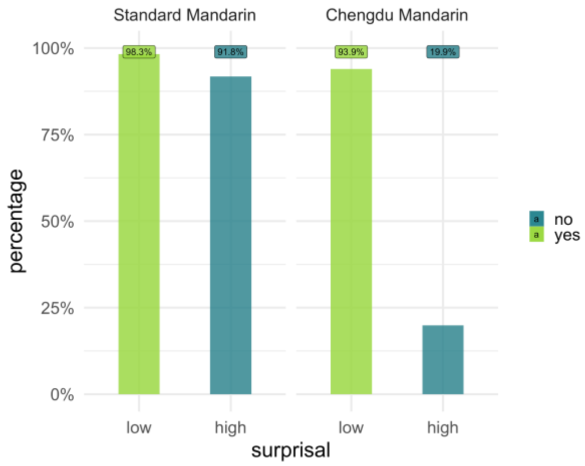


Figure 2: Percentage of “correct” responses across dialect and surprisal conditions. “Yes” (plausible; lime green) is treated as correct for low-surprisal conditions, and “no” (not plausible; blue green) for high-surprisal conditions.

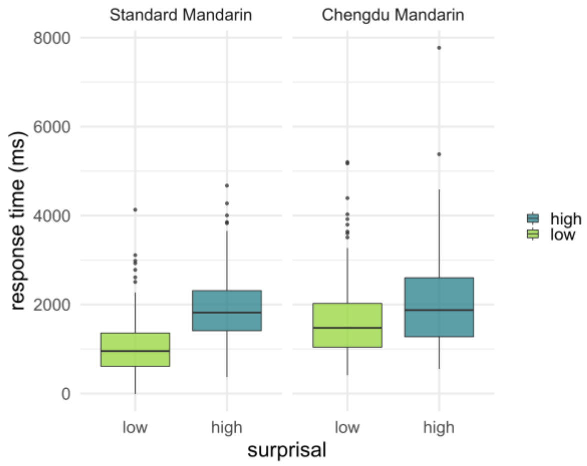


Figure 3: Response times across dialect and surprisal conditions.

Additionally, this experiment was replicated on another dialect, Jinan Mandarin, as an attempt to examine whether the perception of unfamiliar tone systems is effected by the relative phonetic similarity/dissimilarity to the listeners’ native tone system (Zhao & Chodroff, BAAP 2022). Chengdu tone system is considered more dissimilar to the native Standard Mandarin tone system than Jinan Mandarin. The interim results showed that greater phonetic dissimilarity between the unfamiliar tone system and the listeners’ native tone system led to better discrimination of sentence plausibility (Chengdu Mandarin > Jinan Mandarin), as predicted by the perceptual assimilation model (So & Best, 2011, 2014; Reid et al., 2015). Further comparisons with other Mandarin dialects can be done to extensively understand across-dialect perception of lexical tone variation.

**3. What are the factors affecting perceptual adaptation to an unfamiliar tone system?**

Based on the findings in the previous section, I took further interest in investigating the potential factors that may impact adaptation to an unfamiliar tone system. Current literature has provided ample evidence that the perceptual system is highly flexible and adaptive to variability in speech signals (Munro & Derwing, 1995; Weil, 2001; Norris, McQueen, & Cutler, 2003a; Zheng et al., 2005; Bradlow & Bent, 2008). But much more research is still needed to uncover the factors that might prompt, facilitate or inhibit perceptual adaptation.

A critical aspect in the design of the experiment (Zhao, Sloggett & Chodroff, 2022) was that the participants heard both the low- and high-surprisal versions of the sentence in each dialect, which provided the minimal-pair sentences contrasting in semantic plausibility as well as the lexical tone. I wonder whether the presentation of the minimal-pair sentences may facilitate the rapid adaptation to a novel tone system and whether this can be achieved when the minimal pairs are removed. If not, would increasing ambient exposure to the dialect facilitate adaptation when minimal-pair sentences are absent?

To test the potential conditions for adapting to a novel tone system with ambient exposure, I removed the minimal-pair contrast in the dialect-specific stimuli and introduced three-time repetitions of the trials in the new experiment, and compared the data with the previous experiment introduced in the previous section (Zhao, Sloggett & Chodroff, 2023a, in preparation). The absence of minimal-pair sentences is expected to impede adaptation. Increased ambient exposure over repeated trails is expected to redeem the impediment imposed by the removal of minimal-pair contrast.

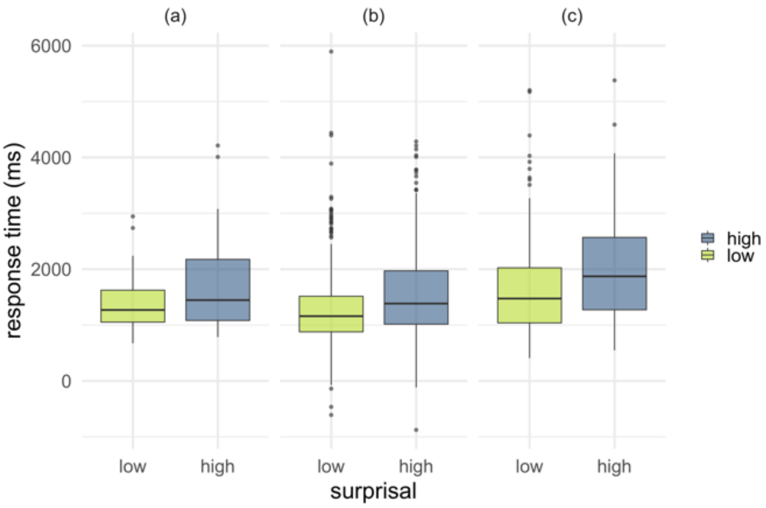


Figure 4: Response times between low and high surprisal across the experiments:
(a) no minimal pairs & no repetition, (b) no minimal pairs & with repetition, and (c) minimal pairs & no repetition.
(a) and (b) were from the current experiment in the paper; (c) was from the previous experiment.

To single out the effect of minimal-pair presentation, the response-time data of Chengdu sentence trials in the first block of repetition was extracted from the newest experiment as in Figure 4 (a), and compared to the data from the previous experiment in Figure 4 (c). For the effect of repetition, response times with all the repetitions in the newest experiment in Figure 4 (b) were compared to the previous experiment (c). Statistical models were run for each comparison with the fixed effects of surprisal, presentation (with-minimal-pair design vs. no-minimal-pair design) and repetition (with repetition vs. no repetition). If there were considerable difference in response times between (a) and (c), the presentation of minimal-pair sentences may be crucial for adapting to a novel tone system. If there were no difference between (b) and (c), increased ambient may be sufficient to compensate for the absence of minimal pairs in the stimuli.

The models showed no reliable difference between the two sets of data within each comparison, which pointed towards the conclusion that rapid adaptation to the novel tone system was persistent under rather adverse conditions, where minimal-pair sentences were absent in the stimuli and ambient exposure was limited to one-time repetition. Since minimal ambient exposure was sufficient for adaptation to the unfamiliar ton system with (c) or without minimal pairs (a) in the stimuli, it was not the case that listeners would rely on the increased ambient exposure to become aware of the surprisal tone manipulation. In fact, repetitions did not reliably improve sensitivity to the surprisal tones, but generally accelerated responses in both surprisal conditions (b). Repetition was more likely a consolidating factor as the mappings between the phonological categories and the novel phonetic tone realizations were reinforced over repetitions. Nevertheless, both factors may facilitate discrimination between the low- and high-surprisal meanings and may potentially ease the process of adaptation or learning of the new tone system, evidence by the increased estimated mean difference in response time between high and low surprisal.

The above experiments have shown that unfamiliar tone systems can be rapidly adapted to with ambient exposure. What about the effect of explicit exposure? A follow-up experiment was done replicating the experiment (a) with uninterrupted explicit exposure of the North Wind and Sun passage using Chengdu and Jinan Mandarin stimuli. Not to expand here, the general conclusion was that two-minute explicit exposure to the connected speech of the dialect reliably boosted accuracy in the sentence semantic-plausibility judgment when the unfamiliar tone system was phonetically distant from the listeners' native tone system. When the unfamiliar tone system was phonetically similar to the listeners' native tone system, the effect of exposure seemed to diminish and even inhibit sensitivity to the surprisal manipulation after exposure (Zhao, Sloggett & Chodroff, 2023b, in preparation).

## 4. Corpus phonetics methodology

Contributing to my research on lexical tone variation was a methodological project for remote audio collection and speech data annotations such as the ManDi corpus. The current version of the corpus contains 357 recordings (about 9.6 hours) of monosyllabic words, disyllabic words, short sentences, *the North Wind and Sun* passage and a Chinese anecdotic 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 during the pandemic. The recordings were force-aligned to create TextGrid annotations at the utterance-, word- and phone-levels, using custom-made Praat scripts and the pretrained Mandarin acoustic model released with the Montreal Forced Aligner (McAuliffe, et al., 2017). Sonorant boundaries for the recordings of monosyllabic words were manually checked and corrected in preparation for acoustic analysis. Throughout this process, I acquired a unique skillset for large-scale phonetic analysis that includes using pretrained acoustic models for phonetic forced alignment, employing corpus phonetics tools, as well as scripting in Praat for acoustical analysis of speech data, and R for data processing and statistical analysis. This skillset has been critical to my own research, and I believe it will be increasingly vital for future research in experimental linguistics and cognitive science.

**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.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729. https://doi.org/10.1016/j.cognition.2007.04.005

Gooskens, C. (2017). Dialect intelligibility. *The handbook of dialectology*, 204-218.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones (pp. 209-217). na.

Jóźwik, A., & Shi, F. (2018). Analysis of monosyllabic tones in Mandarin Chinese produced by Polish students. In Proc. 9th International Conference on Speech Prosody 2018 (pp. 957-960).

Li, Y., Best, C. T., Tyler, M. D., & Burnham, D. (2020). Tone Variations in Regionally Accented Mandarin. In INTERSPEECH (pp. 4158-4162).

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, 18(1), 1-86.

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017, August). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In Interspeech (Vol. 2017, pp. 498-502).

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America,* 102(3), 1864–1877. https://doi.org/10.1121/1.420092

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech*, 38(3), 289-306.

Norman, J. (1988). *Chinese*. Cambridge University Press.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. Cognitive Psychology, 47(2), 204–238.

Reid, A., Burnham, D., Kasisopa, B., Reilly, R., Attina, V., Rattanasone, N. X., & Best, C. T. (2015). Perceptual assimilation of lexical tone: The roles of language experience and visual information. *Attention, Perception, & Psychophysics*, 77(2), 571-591.

So, C. K., & Best, C. T. (2011). Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties. Poznań Studies in Contemporary Linguistics, 47, 133–145.

So, C. K., & Best, C. T. (2014). Phonetic Influences on English and French Listeners' Assimilation of Mandarin Tones to Native Prosodic Categories. *Studies in Second Language Acquisition*, 36(2), 195-221.

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.

Tang, C., & Van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*, 119(5), 709-732.

Tang, C., & van Heuven, V. J. (2008). Mutual intelligibility of Chinese dialects tested functionally. Linguistics in the Netherlands 2008, (25), 145-156.

Tupper, P., Leung, K., Wang, Y., Jongman, A., & Sereno, J. A. (2020). Characterizing the distinctive acoustic cues of Mandarin tones. *The Journal of the Acoustical Society of America*, 147(4), 2570-2580.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, 113(2), 1033–1043. https://doi.org/10.1121/1.1531176

Weil, S. (2001). Foreign accented speech: Encoding and generalization. *Journal of the Acoustical Society of America*, 109(5), 2473.

Wu, J., Chen, Y., van Heuven, V. J., & Schiller, N. O. (2016). Predicting tonal realizations in one Chinese dialect from another. *Speech Communication*, 76, 1-27.

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.

Zhao, L., & Chodroff, E. (2022, June). The ManDi Corpus: A Spoken Corpus of Mandarin Regional Dialects. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 1985-1990).

Zhao, L., Sloggett, S., & Chodroff, E. (2022). Top-down and Bottom-up Processing of Familiar and Unfamiliar Mandarin Dialect Tone Systems. Proc. Speech Prosody 2022, 842-846.

Zhao, L., Sloggett, S., & Chodroff, E. (2023a). The Conditions for Adapting to an Unfamiliar Lexical Tone System without Explicit Exposure. In preparation.

Zhao, L., Sloggett, S., & Chodroff, E. (2023b). The Effect of Exposure on the Adaptation to Unfamiliar Mandarin Dialect Tone Systems. In preparation.

Zheng, Y., Sproat, R., Gu, L., Shafran, I., Zhou, H., Su, Y., Jurafsky, D., Starr, R., & Yoon, S. Y. (2005). Accent detection and speech recognition for Shanghai-accented Mandarin. 9th European Conference on Speech Communication and Technology, 217–220.