

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Clear speech enhancements of Mandarin tones by L1, Heritage, and L2 speakers

Permalink

<https://escholarship.org/uc/item/35k9b181>

ISBN

9798290615226

Author

Vonessen, Julia

Publication Date

2025-06-15

Peer reviewed|Thesis/dissertation

Clear speech enhancements of Mandarin tones by L1, Heritage, and L2 speakers

By

JULIA SOUZA VONESSEN
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Linguistics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Georgia Zellou, Chair

Santiago Barreda

Katharine Graf Estes

Committee in Charge

2025

Abstract

Becoming a competent language user involves the acquisition of different speaking styles, including the acquisition of a clear speech style wherein speakers hyperarticulate in order to make it easier for a listener to understand. The present dissertation investigates the acoustic properties and perceptual benefits of clear speech in first language (L1), heritage language (HL), and second language (L2) speakers of Mandarin Chinese.

The first experiment compares the global clear speech strategies of longer duration, increased intensity, higher f_0 , and expanded f_0 range in both English and Mandarin. 8 L1 Mandarin-L2 English speakers, 11 English-dominant heritage Mandarin speakers, and 10 L1 English-L2 Mandarin speakers produced monosyllabic Mandarin and English words. Longer duration and expanded f_0 range were the primary acoustic enhancements of clear speech across speaker groups and languages, although HL Mandarin speakers did not expand f_0 range in Mandarin clear speech.

The second experiment compares the pitch contour enhancement of the pitch contours of the same monosyllables in Mandarin and English clear and casual speech. The findings show that the contour of the pitch is enhanced in both English and Mandarin, with characteristic aspects of the tone contour being enhanced in both languages. HL and L2 speakers' enhancements were sometimes smaller than L1 speakers' enhancements, and HL and L2 speakers sometimes enhanced different aspects of the contour than L1 speakers.

The third experiment investigates the perceptual benefits of L1, HL, and L2 Mandarin clear speech. There was a clear speech intelligibility benefit for Mandarin in all speaker groups,

though results differed by tone. HL and L2 Mandarin speakers' Tone 3 productions were often perceived as Tone 2.

The results of these experiments suggest 1) that HL and L2 Mandarin speakers make acoustic enhancements in Mandarin clear speech, including enhancing the tone contour; 2) that these enhancements result in intelligibility benefits for L1 Mandarin listeners; and 3) that there is a great deal of inter-talker variation in which clear speech strategies are used by each individual speaker.

Acknowledgements

This dissertation would not have been possible without the support of so many people.

First and foremost, I want to thank my advisor Georgia Zellou for her steadfast support, guidance, and encouragement. I have learned so much from her. She always knew when to offer advice and when to push me to grow. I couldn't have asked for a better advisor.

I want to thank my committee members Santiago Barreda and Katharine Graf Estes for their thoughtful critique and insightful feedback at all stages of this project. I also want to thank my research assistant, David Dai, for his many hours of hard work and insight. A thank you also to Julia Menard-Warwick, whose course on Pedagogical Applications of SLA Theory has shaped my understanding of what it means to be a teacher. And a heartfelt thank you to my first mentor in linguistics, Rachel Hayes-Harb, for her unwavering support and commitment to building up her students.

To Melissa, Eve, Ethan, Ashley, Sophia, and all my friends and cohort in the department, thank you for many great conversations, board game nights, and painting in the grad lounge. Thank you especially to Melissa, my writing partner for the last few years, without whom I might still be staring at a blank sheet of paper. Thank you to Michelle and Sarah for camping trips, good food, and many outings. Thank you to Sonika, Imani, Madison, Charlie, V, Hannah, and everyone else – learning from you all has been one of my greatest joys. Thank you to my best and oldest friend, Leila, for all the letters.

Thank you also to my family, for always encouraging me, for cheering me on, and for helping me stay grounded. And finally, thank you to my partner Yvonne, my greatest support.

Table of Contents

Abstract	ii
Acknowledgements	iv
Chapter 1. Introduction and Literature Review	1
1.1 Introduction	1
1.2 Mandarin tones	3
1.2.1 Theories of second and heritage language acquisition	5
1.2.2 L2 acquisition of Mandarin tones	8
1.2.3 HL acquisition of Mandarin tones	10
1.3 Clear speech	11
1.3.1 Types of clear speech	12
1.3.2 Clear speech cross-linguistically	15
1.3.3 L2 and HL production of clear speech	18
1.3.4 Perception of clear speech	20
1.4 Gaps in the literature	22
1.5 Roadmap of the thesis	23
Chapter 2. Global strategies for clear speech enhancement: A cross-linguistic comparison of Mandarin and English clear speech produced by L1, HL, and L2 Mandarin speakers	25
2.1 Introduction	25
2.2 Methods	28

2.2.1 Participants.....	28
2.2.2 Materials	30
2.2.3 Methods	32
2.2.4 Analysis	34
2.3 Results.....	37
2.3.1 Duration	37
2.3.2 Intensity	40
2.3.3 Mean f0	43
2.3.4 F0 range	46
2.4 Discussion	49
Chapter 3. Enhancement of tonal contours in the clear speech productions of L1, HL, and L2	
Mandarin speakers	56
3.1 Introduction.....	56
3.2 Methods.....	60
3.2.1 Participants.....	60
3.2.2 Materials	60
3.2.3 Methods	60
3.2.4 Analysis	61
3.3 Results.....	62
3.3.1 f0 mean	62

3.3.2 f0 slope.....	66
3.3.3 f0 curve	69
3.4 Discussion	74
Chapter 4. Perception of L1, HL, and L2 Mandarin speakers' clear speech productions	79
4.1 Introduction.....	79
4.2 Methods.....	81
4.2.1 Talkers	81
4.2.2 Listeners	84
4.2.3 Materials	85
4.2.4 Methods	86
4.2.5 Analysis	88
4.3 Results.....	89
4.3.1 Intelligibility	89
4.3.2 Error analysis and accuracy of Tone 3 identification	91
4.4 Discussion	96
Chapter 5. Discussion	100
5.1 Summary of Findings.....	100
5.2 Overall Discussion	102
5.2.1 Cross-linguistic clear speech strategies	102
5.2.2 HL and L2 acquisition of Mandarin tones	105

5.2.3 HL and L2 acquisition of a clear speech register.....	108
5.2.4 Implications for language teaching.....	111
5.2.5 Speech production theories.....	113
5.3 Future Directions	115
5.4 Concluding Remarks.....	118
References.....	120
Appendix A: Model Outputs.....	133

Chapter 1. Introduction and Literature Review

1.1 Introduction

This dissertation investigates strategies for clear speech enhancement in first language (L1), second language (L2), and heritage language (HL) speakers of Mandarin Chinese. First, we compare the global clear speech strategies used in Mandarin to the strategies used in English. Then, we investigate tone contour enhancement in Mandarin. Finally, we investigate the intelligibility of speech from these speakers to L1 Mandarin listeners.

People are able to produce speech in a wide variety of speaking styles, and are able to perceive speech that was produced in a variety of ways. Learning to produce and perceive speech in different styles is fundamental to becoming a competent language user, whether in a first or a second language. Not only are learners required to learn new sounds and sound combinations, but also to produce and perceive these new sounds in a variety of contexts: in casual conversation, during communication breakdowns, and with different types of communication partners.

Clear speech is a speaking style that speakers use when they want to ensure that a listener understands them (Bradlow & Bent, 2002; Picheny et al., 1985). Much research has investigated the acoustic characteristics and intelligibility of clear speech; however, most of this research has focused on English or other European languages, and most research focuses on clear speech produced by first language talkers. A very narrow band of research has investigated clear speech in tonal languages, or clear speech produced by second language or heritage speakers.

Languages with lexical tones use pitch patterns over a vowel or syllable to differentiate words from each other. These tones may be level tones, where the pitch is constant across the syllable and distinguishable from other tones based on relative pitch height, or contour tones, where the pitch changes over the course of the word: falling, rising, or both. It is estimated that 60-70% of the world's languages are tonal (Maddieson, 2013; Yip, 2002).

Mandarin Chinese, the world's second-most spoken language, has over one billion speakers and almost 200 million second language (L2) speakers (Eberhard et al., 2023). In the United States, an estimated 3.5 million households speak Chinese (including Mandarin and Cantonese) at home, including 770,000 individuals who were born in the United States and speak Chinese at home (Dietrich & Hernandez, 2022).

Despite the large number of Mandarin speakers, there is a disproportionately small amount of research examining different speech styles in Mandarin, in comparison to English. Understanding clear speech in tonal languages is vital for our understanding of the ways speakers are able to modulate their speech during communication, given the presence of an entire acoustic dimension that does not exist in the same way in non-tonal languages. The large population of Mandarin speakers worldwide provides both a rich availability of data for exploring speaking styles in tonal languages, as well as an enhanced need for understanding of the different speaking styles used by so many people.

Furthermore, it is important to understand the different ways that people acquire languages, whether that is a home language that is not the dominant language in society (heritage language), or an additional language acquired in adulthood (second language). One challenging aspect of acquiring Mandarin for speakers from many different language backgrounds is the acquisition of lexical tones (Pelzl et al., 2019). Not only are learners and speakers of Mandarin

required to learn to produce and perceive these tones, but they must also learn to produce and perceive these tones in speaking styles that are appropriate for different communicative situations. Investigation of the strategies used by heritage and second language speakers of Mandarin to produce tones in clear and casual speaking styles can inform us on the versatility of these speakers' ability to produce lexical tones, and can suggest areas of difficulty in tone production that can be targeted and supported by language teachers.

The remainder of this chapter reviews the literature on clear speech and Mandarin lexical tones. Section 1.2 discusses Mandarin tones in more depth, including heritage and L2 acquisition of Mandarin tones. Section 1.3 focuses on the nature of clear speech, including types of clear speech, clear speech in Mandarin, and L2 and heritage production of clear speech. Section 1.4 discusses gaps in the literature and how the current dissertation addresses these gaps. Finally, Section 1.5 presents an outline of the dissertation.

1.2 Mandarin tones

Mandarin Chinese is a lexical tonal language with four tones plus a neutral tone (Cheng, 2011). The four main tones, as realized on a single syllable /ma/, are summarized in Table 1. Each syllable carries a tone. T1, the high tone, is a level tone with high pitch. T2 and T4 are contour tones with rising and falling pitch, respectively. T3, the lowest tone, is often realized with a falling-rising contour when produced utterance-finally or in isolation. T3 may also be realized as a low-falling tone, called half-T3, which does not have a rise at the end. T3 is often characterized by creaky voice.

Table 1.1: Mandarin tones

Tone	Name	Pinyin	IPA	Character	Meaning
T1	High-Level	ma1	ma ⁵⁵	妈	mother
T2	High-Rising	ma2	ma ³⁵	麻	hemp
T3	Low-Dipping	ma3	ma ²¹⁽⁴⁾	马	horse
T4	High-Falling	ma4	ma ⁵¹	骂	scold

Tupper et al. (2020) suggest that Mandarin tonal contours may be best characterized by three f0 dimensions: mean pitch, slope of the pitch, and curve of the pitch. These values may be calculated from f0 contours by fitting a second-degree quadratic equation to the contour. Tupper et al. (2020) report which features of the four Mandarin tones distinguish each tone from the other tones, as produced in isolation. As a high level tone, T1 is characterized by a high mean pitch and slope of zero, while rising T2 is characterized by a positive slope. The low-dipping T3 is characterized by a low mean pitch and a positive curve (a curve in the same direction as a U shape). Finally, the falling T4 is characterized by a negative slope and a negative curve (a curve similar to the second half of an upside-down U shape). While both T3 and T4 have contours that curve, the curve of T3 and T4 differ in the direction of the curve. T3 possesses a convex curve similar to a U-shape or a parabola that opens upward, and T4 possesses a concave curve, similar to the second half of a parabola that opens downward. Like T3, T2 also has a convex curve, with a slight decrease in pitch before a larger increase. However, the turning point for T2 is earlier in the syllable than the turning point for T3 (Shen & Lin, 1991).

There is variation in the realization of tones across utterances. The relative pitch height and contour shape of tones in Mandarin may be influenced by a variety of factors, including tonal context – tones are produced differently in isolation than when adjacent to other tones (Xu,

1994). Tone sandhi rules may also turn a T3 into a T2 when it precedes another T3. Additional factors such as prosody (Chen & Tseng, 2019), emotion (Röhr et al., 2022), informational load (Androsova et al., 2022), and speech rate (Tang and Li, 2020; Xu, 1994) also impact tones' pitch contour and pitch height. As a first investigation into L2 and HL clear speech in Mandarin, the current dissertation will focus on tones in isolation.

1.2.1 Theories of second and heritage language acquisition

Second language (L2) acquisition is the acquisition of an additional language after childhood. Second language speakers differ systematically from first language speakers of a language in many linguistic aspects, including perception and production of the sounds of that language (Flege, 1995). Heritage language (HL) acquisition refers to the acquisition of a home language that differs from the dominant or majority language of a given society. Lohndal et al. (2019) outlines three common criteria that define a heritage speaker: that the heritage language is a minority language that differs from the majority language used in the society; that the heritage speaker is to some extent bilingual in both the heritage language and in the majority language used in the society; and that, by the time they are an adult, the heritage speaker becomes more proficient in the majority language used by society. In the United States context, HL speakers of Mandarin primarily acquire Mandarin at home with family members, but they may also be part of immigrant communities that speak the language, or they may take part in community-organized weekend language classes (Montrul, 2016; Carreira & Kagan, 2011).

Like L2 speakers, HL speakers differ systematically from L1 speakers of a language (Montrul, 2010). In particular, thanks to early experience with the heritage language, HL speakers' production and perception of target language sounds are often similar to those of L1

speakers, especially when compared to L2 speakers (Au et al., 2002). Yet, HL speakers also show a great deal of variation in the production and perception of HL speech sounds (Chang & Yao, 2016; Polinsky & Kagan, 2007). Ethnic and cultural identity can affect the speech production of HL speakers (Kissling, 2018), and HL speech shows differences from L1 speech, sometimes thought to result from transfer from the dominant language to the heritage language (Godson, 2004; Chang, 2024), although the extent of transfer from the dominant language to the heritage language is limited (Montrul, 2016).

Heritage language and bilingual phonological acquisition are often described in comparison to monolingual L1 acquisition: do young HL speakers acquire the target language in a similar way to monolingual L1 speakers? Paradis and Genesee (1996) defined three processes affecting the language development of bilingual language speakers in comparison to monolingual speakers: acceleration, delay, and transfer. Acceleration refers to an earlier acquisition of language features than monolingual speakers, while delay refers to a later acquisition of some linguistic feature as compared to monolingual speakers. Transfer refers to the effect of the dominant, societal language on the heritage language. While these terms were originally used to describe the acquisition of syntax in bilingual children, this model has since been used to describe the phonological acquisition of heritage language speakers, especially studies investigating transfer (Lleó, 2018). Many studies have found effects of transfer between a heritage speaker's heritage language and the dominant language of society (Lleó, 2018; Yang & Fox, 2017). Yet, when it comes to lexical tone, when the dominant language of society is non-tonal, there may be little to transfer to the heritage tonal language in this respect.

One of the prominent characteristics of heritage language is the great variation across HL speakers (Chang & Yao, 2024), both in terms of speakers' production and perception of the

language (Chang & Yao, 2016), but also in terms of speakers' life experiences, linguistic experiences, and motivations, which may shape their HL development (Polinsky & Kagan, 2007). Lu and Li (2008) found that a range of motivations shape HL speakers' desire to learn Mandarin, including integrative goals of associating more closely with Chinese culture or connecting with one's heritage and identity, as well as instrumental goals of getting a better job or meeting a university requirement. He (2008) proposes that success in HL attainment is associated with a variety of factors, including an individual's "rootedness" in the heritage culture, a desire to communicate, the development of complex identities, and various factors in the individual's learning environment, including access to diverse input and the attitudes of the surrounding community towards the heritage language. All these factors may vary widely across different HL speakers. When research studies control for certain HL speaker characteristics, for example by limiting participants to students of one language course, a reduced amount of variation between HL talkers has been found (Chang, 2024; Chang & Weiss-Cowie, 2021).

L2 phonological acquisition has been broadly characterized in terms of two models: Flege's Speech Learning Model (SLM; Flege, 1995) and the Perceptual Assimilation Model (Best, 1994). The SLM predicts which L2 sounds will be easy for someone to acquire, based on the sounds of their L1. Under the SLM, L2 sounds are compared to existing L1 sound categories and can be categorized as the same, new, or similar. The SLM predicts that L2 sounds that are the same as L1 sounds will be easiest to acquire, while L2 sounds that are new will be initially difficult but can be fully acquired. On the other hand, sounds that are similar to an L1 sound (but not identical) will be initially easy but may never be fully acquired (Flege, 1995).

The Perceptual Assimilation Model (PAM) predicts which L2 phonological contrasts will be easier or harder for L2 speakers to discriminate perceptually, based on whether those sounds

are perceived as L1 sounds by learners. The PAM categorizes contrasts between L2 sounds based on whether those sounds assimilate into L1 categories. Pairs of sounds which fall into different L1 categories will be easy to discriminate; pairs of sounds which fall into the same L1 category will be difficult to discriminate. Pairs of sounds where one is categorized and one is uncategorized will be easy to discriminate, while the ease of discrimination of pairs of sounds where both are uncategorized will depend on the degree of phonetic similarity between those sounds (Best, 1994). As L2 learners gain experience with the language, they develop new categories for L2 sounds over time (Best & Tyler, 2007).

1.2.2 L2 acquisition of Mandarin tones

Tone acquisition presents significant challenges to L2 learners. Indeed, even highly proficient L2 Mandarin speakers continue to make errors in tone production, although overall accuracy is high (Song, 2021). Part of this can be attributed to the fact that learners without knowledge of a tonal language do not have L1 tonal categories to which they can associate L2 tones; L1 prosody patterns may be the closest linguistic resource they can draw on in order to develop lexical representations of pitch. L2 learners are indeed successful at creating L2 tonal categories, to the extent that L2 tone experience is beneficial for learning additional tone languages (Schaeffer & Darcy, 2020).

Both the SLM and the PAM models of L2 phonological acquisition have been used to model L1 English-L2 Mandarin tone acquisition. Hao (2014) investigated L1 English-L2 Mandarin tone acquisition with respect to the SLM. L1 English speakers of different Mandarin proficiency levels rated Mandarin tones for their similarity to English and then completed a tone identification task. As predicted by the SLM, the tone with the least English-like rating, T3, was

hardest for inexperienced learners to identify, but more experienced learners showed improvement in T3 identification accuracy, which is characteristic for sounds that are dissimilar from L1 sounds. Similarly, Song (2021) found that very advanced L2 speakers (who had been living in China for about a decade) had the fewest errors in production of T3. On the other hand, Hao (2014) found that T2 was the tone rated to be most English-like. T2 was identified well by inexperienced learners, but there was no benefit of experience in T2 identification accuracy. Very advanced L2 speakers in Song's (2021) study also showed the most errors in T2 production. Given T2's perceptual similarity to English prosody patterns, the SLM predicts this pattern of initial ease but lack of full acquisition.

Meanwhile, So and Best (2008) found that L1 English speakers with no Mandarin experience assimilate Mandarin tones to L1 English prosodic intonation patterns in ways that allow predictions about tone acquisition to be made with the PAM. While T1 and T4 mapped reliably onto "flat tone" and "statement" prosody categories, respectively, and T2 mapped onto the "question" prosody category, both T2 and T3 mapped onto an "uncertainty" category; the fact that these two tones map onto the same prosodic category means that the PAM predicts that this contrast should be difficult for L1 English-L2 Mandarin learners to acquire.

Indeed, several studies have found that T2 and T3 are especially difficult for learners to acquire (Pelzl, 2019). This difficulty discriminating between T2 and T3 is not limited to those with an English L1: Hao (2012) found that both English and Cantonese learners of Mandarin found it especially difficult to discriminate between T2 and T3, and that L1 Cantonese speakers were not more accurate at perceiving and producing Mandarin tones than the L1 English speakers. The L1 Cantonese speakers also had difficulty distinguishing Mandarin T1 and T4, which map onto the same Cantonese tone. Tsukada et al. (2015) also found that heritage

Cantonese speakers learning Mandarin performed worse than L1 English speakers learning Mandarin at distinguishing Mandarin T1 and T4. This is a case where phonological assimilation of a tone to an L1 Cantonese tone category is a barrier to learning L2 Mandarin tone categories, as predicted by the PAM. On the other hand, Hao (2012) suggests that L1 English and L1 Cantonese speakers' difficulty with T2 and T3 may be due to the phonetic similarity between T2 and T3.

Overall, Yang (2015) proposes that L1 English-L2 Mandarin learners perceive tones phonologically (according to tone categories; see also Shen & Froud, 2016) but produce tones phonetically. This may result from a focus on pitch contour over level during production, but level over contour during perception.

1.2.3 HL acquisition of Mandarin tones

HL Mandarin speakers' Mandarin tone productions have largely been found to be intermediary between L1 and L2 productions, with HL speakers showing similarities to both groups, while also possessing some unique characteristics. Yang (2015) found that HL speakers have more stable production patterns than L2 speakers, but that HL speakers showed less distinct T1 and T4 productions and T2 and T3 productions than L1 speakers. HL speakers also produced an intermediate pitch range and tone space, smaller than an L1 pitch range, but larger than an L2 pitch range. Chang and Yao (2016) found that HL Mandarin talkers patterned with L1 talkers in producing a larger distinction between the turning points of T2 and T3 as compared to L2 talkers; however, HL speakers showed more variability in the turning points of these tones than either L1 or L2 speakers. High-exposure HL speakers patterned with L1 talkers in producing a shorter T1, while low-exposure HL speakers patterned with L2 talkers and produced T1 with a

longer duration. HL talkers patterned in between L1 and L2 talkers for speech rate differences between monosyllabic and bisyllabic speech. Yet, HL speakers also showed production patterns distinct from L1 and L2 speakers, as HL speakers tended to shorten phrase-final T3.

As for perceptual properties of HL tones, Chang and Yao (2016) found that monosyllabic HL and L2 tones were identified less accurately than L1 tones by L1 listeners. Yet, in a multisyllabic context, HL tones were more intelligible than L2 tones to L1 listeners. T3 was the least intelligible tone for L1 listeners, an effect that was enhanced for HL and L2 T3 productions. In terms of category goodness, L1 listeners rated HL tones intermediate to L1 and L2 tones. It was difficult for L1 listeners to correctly identify HL speakers' language background.

In terms of HL perception, Yang (2015) found that HL tone perception also patterns intermediate to L1 and L2 Mandarin tone perception, with HL Mandarin speakers having stable tone category spaces that were larger than those of L2 listeners but smaller than those of L1 listeners. However, like L2 listeners, HL listeners focus on pitch level more than pitch contour during perception.

Overall, HL Mandarin speakers largely pattern in between L1 and L2 Mandarin speakers in terms of tone production and perception (Chang & Yao, 2016; Yang, 2015), but show a great deal of variation in their production and perception of Mandarin tones (Chang & Yao, 2016).

1.3 Clear speech

Becoming a competent language user includes acquiring knowledge of how changing or enhancing one's production may impact intelligibility by a listener. The Hypo- and Hyper-articulation (H&H) theory (Lindblom, 1990) holds that in speech production, phonetic variation can be explained as a trade-off between speaker-oriented factors (i.e., reducing articulatory

effort) and listener-oriented factors (i.e, making the speech signal as intelligible as possible). In situations where speakers believe listeners may not be easily able to understand them, they may increase their articulatory effort to enhance their speech, producing hyper-articulated or enhanced speech. On the other hand, in situations where listener-oriented factors are less important, speaker-oriented forces of decreasing articulatory effort take precedence, and speakers produce hypo-articulated or casual speech (Lindblom, 1990).

Clear speech is the term for a type of speech style where speakers purposefully increase their articulatory effort in order to increase their intelligibility to a listener (Bradlow & Bent, 2002; Picheny et al., 1985). This has been most widely studied in English. Broadly, English clear speech is characterized by slower speech rate, louder voice, and a larger vowel space (Bradlow et al., 2003; Ferguson & Kewley-Port, 2007, Krause & Braida, 2004). As for pitch, various studies have found f_0 raising and greater f_0 range in clear than in casual speech (Hazan & Baker, 2011; Han et al., 2021). Clear speech may also be characterized by segmentally-targeted enhancements. For example, while vowels are lengthened in clear speech as compared to casual speech (Ferguson & Kewley-Port, 2007; Picheny et al., 1986), English tense vowels are lengthened more than lax vowels, which increases a length distinction between these two types of vowels (Leung et al., 2016).

1.3.1 Types of clear speech

Enhanced speech may be split into two broad categories: speech produced in difficult environments, and speech addressed to different types of listeners. Speech produced in difficult environments can include Lombard speech, which is speech produced in a noisy environment (Lombard, 1911; Junqua, 1996; Marcoux et al., 2022; Tang et al., 2017), masked speech (Badh

& Knowles, 2023; Cohn et al., 2021), and speech produced through a vocoder (Hazan & Baker, 2011; Granlund et al., 2012). Speakers may also produce listener-directed speech, enhancing their speech for different types of listeners, including individuals who are hard of hearing (HoH-DS; Picheny et al., 1986), second language speakers or foreigners (often called L2-DS, nonnative-DS, or FDS; Uther et al., 2007; Scarborough et al., 2007), computers (computer-DS, Burnham et al., 2010; Raveh et al., 2019) or voice AI systems (voice AI-DS; Cohn et al., 2022), children (child-DS; Cychosz et al., 2021), and infants (IDS; Fernald, 1989; Uther et al., 2007; Tang et al., 2017; Zhang et al., 2024). These listener-directed speaking styles show acoustic differences from speech directed to a normal-hearing, L1 adult (often called casual speech, conversational speech, plain speech, or adult-directed speech (ADS)).

In all these enhanced speech styles, speakers make acoustic adjustments to their speech with the goal of communicating better. However, the exact acoustic adjustments may differ between these speaking styles. Hazan et al. (2015) compared L2-DS (elicited through an interactive task with an L2 speaker) to clear speech elicited through two acoustic barrier conditions, one where an L1 listener heard speech through a vocoder, and one where an L1 listener heard speech through multitalker babble. In L2-DS, talkers primarily enhanced their speech by producing longer durations and an expanded vowel space. Yet, the size of these acoustic adjustments was smaller for L2-DS than for the acoustic barrier conditions. Uther et al. (2007) compared L2-DS and IDS, finding that vowels were hyperarticulated in both conditions; yet, only IDS showed longer durations and higher pitch, while L2-DS and ADS were equivalent in terms of duration and pitch. IDS may be a special case of listener-DS, as attention and affect are of high value in interactions with infants (Fernald et al., 1989), whereas intelligibility is the primary driving force behind acoustic modifications in clear speech registers.

A further consideration in the investigation of enhanced speech are the methodological choices used to elicit enhanced speech. These speaking styles are often investigated as read speech, elicited by having participants read words, sentences, or passages aloud, while imagining they are speaking to a certain type of listener (e.g., Bradlow & Bent, 2002). Often, these are read sequentially in a list, while some studies also elicit enhanced speech through a task with a simulated misunderstanding (e.g., Tupper et al., 2021; Cohn et al., 2022) or in a situation where a phonetic competitor to the target word is present (Kato & Baese-Berk, 2021). The exact instructions given to participants may result in acoustic differences as well: Lam et al. (2012) found that the size of acoustic modifications as compared to a habitual speaking style were typically largest when speakers were instructed to “overenunciate,” followed by instructions to speak to a hearing-impaired listener. Instructions to simply speak “clearly” resulted in the smallest acoustic modifications.

Furthermore, acoustic differences have been found between clear speaking styles in read speech, versus clear speaking styles in spontaneous speech or interaction with an actual person. Speech produced in interaction with a real (normal-hearing) listener is enhanced less than speech produced to an imaginary hard of hearing listener (Scarborough & Zellou, 2013). Similarly, Scarborough et al. (2007) found that speakers spoke slower and with expanded vowel spaces in both real and imagined L2-DS. Yet, speech to an imagined listener was also slower and with an expanded vowel space as compared to speech directed to a real conversation partner. Likewise, Hazan and Baker (2011) found greater acoustic enhancement in read clear speech as opposed to interactive speech in various conditions, with talkers producing higher f_0 in read speech, and greater enhancement of pitch range in read clear speech, as compared to interactive clear speech. They propose that talkers fine-tune their productions more effectively for real listeners than for

imagined listeners, only enhancing speech for real listeners when there is a miscommunication, whereas read clear speech results in consistent levels of enhancement.

While it is important to acknowledge the differences between read laboratory speech and interactive, conversational speech, the fact that speakers produce acoustically different speech styles when reading speech to different types of imagined listeners is of linguistic interest and demonstrates the incredible versatility and agency that speakers have when producing speech.

1.3.2 Clear speech cross-linguistically

Cross-linguistically, there are commonalities among the strategies used by speakers in clear speech. Granlund et al. (2012) found similar levels of intensity and f_0 increases in enhanced speech produced by Finnish and English speakers in a task where communication was hindered by the use of a vocoder. Similarly, Smiljanić and Bradlow (2005) found that clear speech produced in English and Croatian showed similar levels of slowed speech, increased pitch range, and vowel space expansion. However, clear speech patterns of vowel lengthening were language-specific to English and Croatian, with the English tense-lax distinction enhanced in English clear speech, and the Croatian vowel length contrast enhanced in Croatian clear speech.

Taken together, prior work on non-tonal languages suggests that global clear speech strategies include higher intensity (Granlund et al., 2012), slower speech (Smiljanić & Bradlow, 2005), expanded vowel space (Smiljanić & Bradlow, 2005; Bradlow, 2002), higher f_0 (Granlund et al., 2012), and greater pitch range (Smiljanić & Bradlow, 2005). Language specific strategies could include vowel length (Smiljanić & Bradlow, 2005) or other potential targeted segment or inventory-specific enhancements.

Investigations of enhanced speech in tonal languages often focus on Lombard speech or infant-directed speech (IDS). In tonal languages, as in non-tonal languages, Lombard speech has been consistently found to have higher pitch (Tang et al., 2017; Zhao and Jurafsky, 2009; Kasisopa et al., 2014). Yet, the picture is more complicated when it comes to enhancement of tonal contrast. In investigations of tone enhancement in Lombard speech, Zhao and Jurafsky (2009) and Tang et al. (2017) found no tone space expansion in Cantonese and Mandarin respectively. However, Kasisopa et al. (2014) did find tone space expansion in an investigation of Thai Lombard speech, which contrasts with the findings from Mandarin and Cantonese and may suggest language-specificity in tone enhancements.

Studies investigating infant-directed speech (IDS) also provide evidence for the ways in which lexical tones may be enhanced. In non-tonal languages, IDS exhibits a great amount of pitch modification, including greater f_0 range and variation than ADS (e.g., Knoll & Costal, 2015). In tonal languages, IDS displays hyper-articulated tones that differ from ADS in having higher f_0 , wider f_0 range, longer duration, and exaggerated contour; yet, the hyper-articulation of these tones does not impact the identifiability of the tone (Xu & Burnham, 2010; Cheng & Chang, 2014). In Mandarin specifically, IDS also features higher pitch and exaggerated tonal contours compared to ADS (Tang et al., 2017). However, although the contours are exaggerated, the overall shape of the contour, including the relative timing of the turning point in the falling-rising contour of T3, is the same in IDS and ADS, demonstrating the preservation of acoustic cues for tone identity in IDS (Liu et al., 2007). Tang et al. (2017) found that pitch height increased in both IDS and in Lombard speech as compared to ADS, while only IDS showed tone hyperarticulation.

A few studies have specifically investigated Mandarin clear speech. Mandarin clear speech is consistently characterized by an increase in duration, across a variety of stimulus types and contexts (Rittenberry & Shport, 2024; Zhang et al., 2024; Tupper et al., 2021; Papoušek & Hwang, 1991; Chien & Brajot, 2025). Higher intensity is also associated with Mandarin clear speech (Rittenberry & Shport, 2024; Tupper et al., 2021).

As for differences in pitch and pitch contour, the picture is more complicated. Across-the-board increases in mean f_0 in Mandarin clear speech have been found in some studies (Zhang et al., 2024; Rittenberry & Shport, 2024). In other studies, only certain tones show an increase in mean f_0 . Zhang et al. (2024) and Chien and Brajot (2025) found increases in mean f_0 in T1 and T2, while Zhang et al. (2024) also found a decrease in mean f_0 in T3. Tupper et al. (2021) found increases in mean f_0 in T2 and T3 in clear speech. Pitch range increases in Mandarin clear speech (Papoušek & Hwang, 1991; Rittenberry & Shport, 2024), with increases in pitch range in clear speech being found specifically for the contour tones T2, T3, and T4 (Zhang et al., 2024; Chien & Brajot, 2025). Pitch maximum increases in clear speech (Papoušek & Hwang, 1991; Rittenberry & Shport, 2024), while pitch minimum either does not change (Papoušek & Hwang, 1991) or increases just in T1 and T2 (Rittenberry & Shport, 2024).

When the contour shapes of individual tones are analyzed, previous research has found steeper slopes for the contour tones in clear speech, including steeper slopes for T2 (Zhang et al., 2024), T3 (Chien & Brajot, 2025), and T4 (Zhang et al., 2024; Tupper et al., 2021; Chien & Brajot, 2025). Chien and Brajot (2025) found that tone space area increases in Mandarin clear speech. Previous research has not identified an effect of clear speech on the curve of the contour of any Mandarin tone. Overall, Papoušek & Hwang (1991) found that tonal contours in foreigner-directed speech (FDS) displayed more f_0 fluctuations, with the number of local

minimums or maximums in the f_0 contour corresponding closely to the number of syllables in the utterance. This is different from ADS and IDS, which are characterized by a reduced number of f_0 fluctuations, and likely reflects the intelligibility-driven nature of FDS.

1.3.3 L2 and HL production of clear speech

Speech production in an L2 is more effortful than in an L1. This has the potential to impact the balance between talker- and listener-oriented forces in the H&H model, where talkers must balance their own articulatory effort with listener intelligibility (Lindblom, 1990). For instance, L2 talkers report prioritizing their own pronunciation when planning utterances to a Voice-AI system which is likely to misunderstand them (Wu et al., 2020).

Even though L2 speech is more effortful, there is work suggesting that L2 speakers are able to acoustically enhance their speech in similar ways and to similar extents as L1 speakers (Smiljanić & Bradlow, 2007; Smiljanić & Bradlow, 2011; Granlund et al., 2012). However, work measuring the precise acoustic properties of L2 productions show that the size of these enhancements may be modulated by proficiency (Kato & Baese-Berk, 2021; 2022; 2023) and age of immersion (Rogers et al., 2010). Kato (2020) found that both L1 and L2 English speakers (the non-native speakers' L1 was Mandarin) produce slower speaking rate, higher mean f_0 , wider f_0 range, higher mean intensity, and an expanded vowel space in clear speech as opposed to casual speech. L1 and high proficiency L2 speakers made these modifications to a similar extent, while lower proficiency L2 speakers made these modifications to a lesser extent. Talkers from all three participant groups also showed the ability to enhance specific segmental contrasts (e.g., the contrast between word-initial /p/ and /b/). Lower-proficiency talkers showed comparable acoustic enhancements for contrasts that were more familiar to them, while

differences in language background only emerged for contrasts that were not present in the L2 talkers' L1 (Kato & Baese-Berk, 2021). Thus, this work provides evidence that acquiring the specific production patterns of clear speech in one's second language is a skill that increases with proficiency.

Research in L2 enhanced speech in Mandarin specifically is limited. Ji et al. (2022) investigated Lombard speech produced by L1 Korean-L2 Mandarin speakers, and found that L1 and L2 Mandarin speakers both made changes to tonal contours in noisy environments as compared to quiet environments, but that L1 and L2 speakers differed in the specific contour changes. For example, in Lombard speech, L1 Mandarin speakers increased the pitch range of T4 more than L2 Mandarin speakers did. Additionally, in Lombard speech, these L2 Mandarin speakers' T2 and T3 productions became acoustically more similar, which the authors attribute to L2 speakers' increased difficulty in using auditory feedback to adjust pronunciation in a noisy environment. In a follow up study with more participants, Ji et al. (2025) found that L2 Mandarin speakers generally made greater duration and intensity changes in Lombard speech, while L1 Mandarin speakers made no intensity changes, and only increased duration for T3 and T4. On the other hand, both L1 and L2 speakers increased f_0 in noisy conditions. The greatest differences between L1 and L2 speakers was in T3. While both groups raised f_0 for T3 in white noise, L2 speakers disproportionately raised the middle portion (which typically has the lowest pitch), resulting in a T2-like contour shape in Lombard speech.

As for heritage clear speech, HL speakers are also able to enhance their speech, using the same strategies as L1 speakers (Chang, 2024; Chang & Weiss-Cowie, 2021). However, HL speakers show distinct patterns from L1 speakers in some clear speech strategies. Chang and Weiss-Cowie (2021) found that HL Korean speakers use the same strategies of longer vowel

duration, expanded vowel space, and offset f0 raising as L1 Korean talkers, but that the HL talkers actually produce larger vowel space enhancements than L1 talkers during glide production (Chang & Weiss-Cowie, 2021). Additionally, HL Korean talkers show an effect of transfer from their dominant language in the characteristics of one Korean vowel in reduced speech, although the vowel is produced in a more L1 Korean-like way in enhanced speech (Chang, 2024). L2 English speakers with an early age of immersion in Rogers et al.'s (2010) study could also be characterized as heritage speakers of Spanish. The ability of these speakers to enhance clear speech in English in the same ways as monolingual English speakers demonstrates the ability of heritage speakers to enhance speech in their dominant, non-heritage language.

1.3.4 Perception of clear speech

The intention behind clear speaking styles is to produce speech that is more intelligible: in other words, to produce speech that is easier for a listener to understand. In perceptual studies where listeners transcribe words or phrases that are played to them in noise, clear speech produced by L1 talkers consistently results in an intelligibility benefit for listeners, as measured by transcription accuracy rates of clear speech versus casual speech (Smiljanić & Bradlow, 2005; Bradlow & Bent, 2002; Picheny et al., 1985).

Some talkers produce speech that is more intelligible than others (Ferguson, 2012). A perceptual benefit from clear speech may be modulated by talker factors such as language background (Bradlow & Bent, 2002) and age (Smiljanić, 2013); by listener factors such as familiarity with the speaker (Domingo et al., 2020) and familiarity with the lexical items (Gonçalves & Silveira, 2015); and by lexical properties of the speech in question, such as lexical frequency (Gonçalves & Silveira, 2015) and semantic context (Kennedy & Trofimovich, 2008).

Different types of clear speech may also produce different intelligibility benefits. Aoki and Zellou (2023) found that HoH-DS was more intelligible than L2-DS, which did not differ in intelligibility from casual speech. Similarly, Hazan et al. (2015) found that listeners rate L2-DS as less clear than speech produced to a listener in a noisy environment, and Vonessen et al. (2024) found that voice AI-DS did not show any intelligibility benefit when compared to human-DS.

In Mandarin specifically, L1 clear speech has been found to be more intelligible to L1 listeners than casual speech. Yet, this perceptual benefit for clear speech is modulated by tone, with increases in identification accuracy for Tone 1 and Tone 4 and a decrease in accuracy for Tone 2 in clear speech (Zeng et al., 2023). Listeners with no Mandarin experience also benefit from Mandarin clear speech, as clear speech improves tone identification accuracy for these very new learners (Han et al., 2018; Zeng et al., 2023), though only some talkers appear to provide a benefit to these listeners (Han et al., 2019). Mandarin Lombard speech also results in increased intelligibility (Chen et al., 2025). On the other hand, Mandarin IDS does not result in increased intelligibility as compared to ADS, reflecting the lesser emphasis on intelligibility in IDS (Tang et al., 2025).

As for the intelligibility of clear speech produced by L2 talkers, there are mixed findings on whether clear speech enhancements produced by L2 speakers provide an intelligibility benefit for listeners. Although Kato and Baese-Berk (2023) found sizeable differences in acoustic modifications between higher and lower proficiency L2 talker groups for clear as compared to casual speech, neither group's acoustic modifications resulted in an intelligibility benefit for L1 listeners. Yet, in another study, Kato and Baese-Berk (2022) found that clear speech produced by L2 talkers did provide an intelligibility benefit for L1 listeners. This benefit was larger for speech

produced by higher proficiency L2 speakers than for speech produced by lower proficiency L2 speakers. Similarly, Rogers et al. (2010) found that speech produced by L1 Spanish-L2 English speakers who were immersed in an English-speaking culture before the age of 12 had a similar clear speech benefit as L1 English speakers, while speakers who were immersed after the age of 15 displayed a reduced clear speech benefit. Hence, when L1 vs. L2 clear speech is assessed perceptually, via the size of the boost for L1 listeners, L2 speakers do not necessarily produce enhancements that benefit native listeners as much as L1 speakers do; instead, factors such as proficiency and age of immersion play a role in the intelligibility benefit of clear speech. To the author's knowledge, the intelligibility of clear speech produced by HL speakers has not yet been assessed.

1.4 Gaps in the literature

Previous research has proposed that signal-enhancing strategies such as increased duration, higher pitch, expanded pitch range, and greater intensity are global clear speech strategies, commonly used across many different languages (Smiljanić & Bradlow, 2005; Granlund et al., 2012). However, much of clear speech research has been conducted on non-tonal languages, and cross-linguistic studies of clear speech have also only directly compared non-tonal languages. It is important to conduct cross-linguistic work that includes tonal languages. In Chapter 2, we seek to compare the global clear speech strategies used in Mandarin and in English.

Previous investigations of L2 and HL clear speech have also only looked at non-tonal languages such as English and Korean (e.g., Jung & Dmitrieva, 2023; Chang & Weiss-Cowie, 2021), or Lombard speech (Ji et al., 2022; Ji et al., 2025). The present study is (to the author's

knowledge) the first investigation of clear speech enhancements in a tonal L2 or HL.

Examinations of global speech enhancements in L2 and HL Mandarin are investigated in Chapter 2, while tone contour enhancement in L2 and HL Mandarin is investigated in Chapter 3.

To the author's knowledge, no prior work has investigated the perceptual consequences of clear speech produced by L2 speakers in a tonal language, or by HL speakers in any language. In Chapter 4, the present study investigates the perceptual benefit of Mandarin clear speech produced by L1, HL, and L2 Mandarin speakers.

Rittenberry and Shport (2024) noticed a great deal of variation in the strategies used by different talkers in Mandarin clear speech, and call for further investigation of this individual variation. The present study uses a Bayesian approach to statistical modeling, which allows for detailed analysis of group means, simple effects, and between-talker variation. In each chapter, we examine individual talker differences: individual differences in global clear speech enhancement strategies in Chapter 2, individual differences in tone contour enhancement in Chapter 3, and individual differences in the perceptual benefits of clear speech in Chapter 4.

1.5 Roadmap of the thesis

In Chapter 2, we focus on global strategies for the enhancement of clear speech in English and Mandarin as produced by L1, HL, and L2 Mandarin speakers. We investigate the clear speech strategies of enhancing duration, intensity, mean pitch, and pitch range in both English and Mandarin. In Chapter 3, we focus on tone-specific enhancements of the pitch contour in L1, HL, and L2 Mandarin clear speech. Then, in Chapter 4, we examine the perception of L1, HL, and L2 Mandarin clear speech by L1 Mandarin listeners. Finally, in

Chapter 5, we discuss the findings and lay out conclusions and future directions for research in L2 and HL Mandarin clear speech.

Chapter 2. Global strategies for clear speech enhancement: A cross-linguistic comparison of Mandarin and English clear speech produced by L1, HL, and L2 Mandarin speakers

2.1 Introduction

Clear speech is a speaking style used when speakers modify their speech to be more intelligible to a listener (Bradlow, 2002). In clear speech, speakers may use a variety of strategies to enhance the overall speech signal, including increasing the duration, intensity, overall pitch, pitch range, and vowel space of their speech (Picheny et al., 1986; Bradlow et al., 2003; Smiljanić & Bradlow, 2005; Ferguson & Kewley-Port, 2007; Hazan & Baker, 2011; Han et al., 2021; Lam & Tjaden, 2016). These overall signal-enhancing speech strategies are said to be global clear speech strategies, and are thought to be used across languages (Granlund et al., 2012; Smiljanić & Bradlow, 2005). Yet, speakers may also enhance speech in language-specific ways that enhance language-specific phonetic contrasts. For example, the duration of language-specific vowel length contrasts in English and Croatian are enhanced in language-specific ways (Smiljanić & Bradlow, 2008).

L2 and HL speakers are able to acoustically enhance their speech in similar ways and to similar extents as L1 speakers (Smiljanić & Bradlow, 2007; Smiljanić & Bradlow, 2011; Granlund et al., 2012). For L2 speakers, acoustic clear speech modifications become larger in

size as proficiency increases (Kato & Baese-Berk, 2021; 2022; 2023), as well as with an earlier age of immersion (Rogers et al., 2010). Research on HL clear speech is sparse. HL Korean-L1 English learners use strategies of longer duration, vowel space expansion, and offset f0 raising in Korean clear speech (Chang, 2024; Chang & Weiss-Cowie, 2021). Chang (2024) found that these HL clear speech patterns were similar to those produced by L1 speakers, although HL speakers showed an influence of their dominant language (English) in the vowel quality of one vowel in reduced speech. Chang and Weiss-Cowie (2021) similarly found that HL Korean-L1 English speakers used the same clear speech strategies as L1 talkers; however, the size of vowel space enhancements produced by HL talkers was larger than values found for L1 talkers in a previous study (Chang, 2017).

Much clear speech research comes from clear speech in non-tonal languages, where the signal-enhancing strategies of increased mean pitch and increased pitch range do not conflict with any code-based strategies that would enhance language-specific contrasts. However, in tonal languages such as Mandarin Chinese, different pitch patterns are used lexically to distinguish words. Thus, there is a potential for pitch to be used as a code-enhancing strategy, to make tones more distinguishable, such that pitch may not be used as a signal-enhancing strategy in the same way that it would in a non-tonal language. An increase in pitch range would increase the distinctiveness of different tones as higher tones become higher and lower tones become lower, so the overall global strategy of increase in pitch range would not conflict with code-based strategies to increase pitch range to differentiate tones. However, an overall increase in mean pitch, which might make lower tones higher, would conflict with a code-based strategy that makes low tones lower to increase discriminability.

In Mandarin, longer duration is consistently used as a clear speech strategy (Rittenberry & Shport, 2024; Zhang et al., 2024; Tupper et al., 2021; Papoušek & Hwang, 1991; Chien & Brajot, 2025). Where investigated, higher intensity has also been found to be a clear speech strategy (Rittenberry & Shport, 2024; Tupper et al., 2021). When it comes to pitch, the picture is more complicated. Some studies have found higher pitch used as a clear speech strategy (Rittenberry & Shport, 2024; Papoušek & Hwang, 1991), but other studies have found no overall difference in mean pitch between Mandarin clear and casual speech, or have found differences only for specific tones (Zhang et al., 2024; Tupper et al., 2021; Chien & Brajot, 2025). On the other hand, pitch range has been consistently found to increase in Mandarin clear speech (Zhang et al., 2024; Rittenberry & Shport, 2024; Papoušek & Hwang, 1991; Chien & Brajot, 2025). Rittenberry and Shport (2024) suggest that their finding of a difference in pitch range of 13-15 Hz between Mandarin clear and casual speech is comparable to the difference of 18 Hz in English and Croatian, which are non-tonal languages (Smiljanić & Bradlow, 2005). However, to the author's knowledge, no study has yet directly compared clear speech in a tonal and non-tonal language.

The present study investigates the signal-enhancing clear speech strategies of duration, intensity, f_0 mean, and f_0 range in Mandarin and English, as produced by L1 Mandarin-L2 English speakers, HL Mandarin-L1 English speakers, and L2 Mandarin-L1 English speakers. We ask whether the size of any clear speech enhancements are the same across both language and the talker's status as an L1, HL, or L2 speaker.

If these signal-enhancing strategies are indeed global clear speech strategies, then we would expect all talkers to apply these strategies equally across languages. On the other hand, conflicting signal-based and code-based strategies, especially with mean pitch, might result in no

effect of mean pitch for Mandarin. Additionally, we may see an effect of language background. Because of the increased effort involved in producing speech in a second language, we might expect to see smaller acoustic modifications produced by L2 speakers than by L1 speakers. We might also see larger acoustic modifications produced by HL speakers as compared to L1 speakers (Chang & Weiss-Cowie, 2021; Chang, 2017).

2.2 Methods

2.2.1 Participants

8 L1 Mandarin speakers, 11 HL Mandarin speakers, and 10 L2 Mandarin speakers were recruited to participate in this study. This sample size is consistent with previous research on L2 and HL clear speech which have had sample sizes of four to ten talkers in each language group (Chang, 2024; Chang & Weiss-Cowie, 2021; Jung & Dmitrieva, 2023). The participants in this study were all undergraduate or graduate students at an American university. Most participants were recruited from a joint psychology-linguistics research participant pool and were granted partial course credit for participation. 4 HL and 3 L2 Mandarin speakers were recruited from flyers distributed to students taking Chinese classes by Chinese instructors or by word of mouth. These participants were compensated with gift cards.

The 8 L1 Mandarin-L2 English speakers recruited for the study had a mean age of 19.4 (age range 18-21; 6 female, 2 male). Seven of these participants lived in China until adulthood before coming to the U.S. for college; one moved to the U.S. at age 10. Six participants spoke a different dialect of Chinese at home as a child in addition to or instead of Mandarin (Wu/Shanghainese, Sichuanese, Gan Chinese, Wuhanese). These speakers started learning

English in school between the ages of 5 and 10 (mean 7.4). Two participants also spoke a third language (Arabic or Japanese).

The 11 HL Mandarin-L1 English speakers recruited for the study had a mean age of 20.1 (age range 18-23; 2 non-binary, 8 female, 1 male). These talkers reported starting to learn Mandarin between the ages of 0 to 5 (mean 0.8) and spoke Mandarin and/or Taiwanese Mandarin in the home as children. Two of these talkers also spoke Cantonese in the home as children. These talkers report that Mandarin or Taiwanese Mandarin is their strongest dialect of Chinese, and English is their most dominant language in general. All talkers lived in the U.S. until adulthood, apart from one talker who lived in China for one month at the age of 8. These talkers attended an average of 2.0 years of after school or weekend Chinese classes, 1.5 years of elementary or middle school classes, 1.3 years of high school classes, and 0.4 years of college courses.

The 10 L2 Mandarin-L1 English speakers recruited for the study had a mean age of 20.5 (age range 18-26; 8 female, 2 male). They began acquiring Mandarin from the ages of 5-18 (mean 11.3). These talkers did not speak Mandarin in the home as children. Five of these talkers spoke Cantonese and/or Taishanese in the home growing up; the other five did not speak any tone languages at home. One talker lived in China between the ages of 8 and 12. Apart from that talker, all the L2 talkers lived in the U.S. until adulthood. English is the most dominant language for all of these talkers. The most dominant variety of Chinese for these talkers is Mandarin (7 talkers) or Cantonese (3 talkers). On average, these talkers participated in 1.5 years of weekend or after school Chinese instruction, 1 year of elementary or middle school classes, 0.5 years of high school classes, and 0.8 years of college classes. Two speakers studied abroad in China, for four months and 1.5 years, respectively.

For the purposes of this study, heritage and L2 speakers were categorized based on whether they spoke Mandarin in the home as children, such that participants who spoke Mandarin in the home as children were categorized as HL Mandarin speakers. Participants who did not speak Mandarin in the home as children, but who instead learned Mandarin in school or in after-school programs, were categorized as L2 Mandarin speakers (Wu, 2023). We expect that there is a great deal of variation in the Mandarin proficiency of talkers in both HL and L2 Mandarin groups. As a first study investigating HL and L2 clear speech in Mandarin, it is worth including a diverse population in order to see the range of variation across talkers. Future studies can decrease variation by controlling for more participant characteristics such as proficiency and age of acquisition. For example, Chang & Weiss-Cowie (2021) showed success at getting less variation in heritage speaker performance by studying one class of learners; this technique may be useful for controlling for the proficiency of HL and L2 speakers in future work in this area.

2.2.2 Materials

The word list for the experiment is made up of 12 Mandarin monosyllabic words - 3 different vowels /i, ʌ, u/ carrying four different lexical tones. These are shown in Table 2.1. These twelve syllables were also used by Tupper et al. (2021) in their investigation of L1 Mandarin clear tones. For the /ʌ/ syllables, the same characters as Tupper et al. (2021) were used. For the /i, u/ syllables, the characters chosen to represent these syllables for the task were selected from a series of beginner textbooks used by UC Davis Chinese classes (Liu et al., 2017; 2018a; 2018b).

Table 2.1: Word list

Pinyin	IPA	Char	Meaning
yi1	i ⁵⁵	一	one
yi2	i ³⁵	姨	aunt
yi3	i ²¹⁴	以	with
yi4	i ⁵¹	意	meaning
e1	ɤ ⁵⁵	婀	graceful
e2	ɤ ³⁵	鹅	goose
e3	ɤ ²¹⁴	恶	nauseous
e4	ɤ ⁵¹	饿	hungry
wu1	u ⁵⁵	屋	house
wu2	u ³⁵	无	without
wu3	u ²¹⁴	五	five
wu4	u ⁵¹	物	thing

Three English hVd words with the vowels /i, ʌ, u/ were also used: *heed*, *hud*, and *who'd*. These words were chosen to match the three Mandarin vowels, with a high front vowel, a central vowel, and a high back vowel.

Two word order lists were created for each set of words (English and Mandarin separately). For Mandarin, each list contained two instances of each of the 12 target words, for a total of 24. For English, each list contained two instances of each of the 6 target words, for a total of 12. Each word appeared once in the first half of the list and once in the second half of the list. During the experiment, the lists were randomly assigned to each participant such that participants saw the first list for clear speech and the second list for casual speech, or vice versa, for both English and Mandarin.

2.2.3 Methods

Before beginning the recording task, participants were shown a list of the three English words. To ensure that each participant knew the target pronunciation of the words, each word was accompanied by a more common English rhyme (*seed* for *heed*, *mud* for *hud*, and *food* for *who'd*). Participants were asked to say the three English words aloud. If they pronounced the vowel incorrectly, the experimenter prompted the participant with the rhyming word (e.g., “This one rhymes with food”). Once the participant had produced each target word, the experiment continued.

Participants completed the experiment in a soundproof booth with an Audio-Technica AT2020USBi cardioid condenser microphone. Participants first completed informed consent. HL and L2 Mandarin participants also completed a character familiarity task where they rated their familiarity with the 12 Mandarin characters in the study on a four-point scale (0 = “I do not know this word”, 1 = “I think I have seen this word before”, 2 = “I recognize this word as a Chinese word, but I do not know its meaning”, 3 = “I know this word”; Bent & Bradlow, 2003). For any character that participants rated 2 or lower, participants were shown information about that character’s meaning and pronunciation. All participants were also shown information about the character 惡 which can have both a third tone (e3) and fourth tone (e4) pronunciation.

Participants were asked to use the third tone pronunciation of this character during the study.

For task instructions, participants were told that they would read words aloud in Chinese and English in different ways. First, they were asked to “Speak clearly as if you are talking to someone who is having a hard time understanding you” (Zellou et al., 2022), and they were told that their recordings would be played to language learners to help them learn. For casual speech,

they were asked to “Speak as if you are talking to a friend or family member you have known for a long time who has no trouble understanding you, and speak quickly” (Zellou et al., 2022). For a practice task, participants were presented with three common characters (the first person, second person, and third person singular pronouns) and asked to read them aloud in the two different speech styles.

Participants first completed a recording task for clear Mandarin productions. Each ordering list was randomly assigned to either the clear speech block or the casual speech block. For each trial, they were presented with a character (一) and its Pinyin representation, both with the accent (yī) and the tone number (yī1). After reading the word aloud, participants pressed a button to proceed to the next trial. In total, they read each target word aloud twice (24 clear Mandarin productions). Then, they completed the casual Mandarin productions, with the order of the other ordering list. They read 24 casual Mandarin productions.

After the Mandarin recording blocks, participants immediately completed a clear English block and a casual English block. Each block contained 6 productions (2 repetitions of each target word). In total, participants produced 60 productions (48 Mandarin, 12 English; 30 clear, 30 casual), and the entire recording task took less than ten minutes.

After the recording task, participants completed the LEAP-Q questionnaire with information about their language backgrounds (Marian et al, 2007; Kaushanskaya et al, 2020). They also answered additional questions about their experience with Mandarin and other Chinese languages.

2.2.4 Analysis

Vowels were segmented in Prosody Pro (Xu, 2013), and duration measures were extracted as the difference between the timepoints where periodicity and higher-level formant structure started and ended in the waveform. Specifically, boundaries for each vowel were set when the shape of the waveform became different from the overall pattern of the vowel, often when the waveform changed to a simple sine wave pattern, or when the waveform changed to a very low amplitude. If the syllable ended with creaky, the end boundary was placed before the creaky began (this was often the case with productions of Tone 4). The very first and last identifiable cycle in the syllable were excluded from the boundaries, as they were usually very quiet and irregular. ProsodyPro was also used to extract mean intensity of each vowel. In order to get a more normal distribution of duration measurements, we analyze the log of duration in this analysis.

To get f_0 , pulses were also corrected in ProsodyPro (Xu, 2013). Octave errors and missing pulses during creaky voice were corrected. When pulses were manually added, effort was made to follow the same pattern as correct auto-assigned pulses from that syllable (i.e., if pulses were auto-assigned to occur at the positive peak in the cycle, then missing pulses were also added following that pattern). If no reference was present (as was often the case for creaky voice in Tone 3 productions), pulses were added at the point in the cycle with the largest amplitude (this usually occurred at the negative peak in the waveform for each cycle). A phonetically trained research assistant completed initial segmentation and pulse correction, and the author reviewed and further corrected boundaries and pulses as needed.

ProsodyPro extracts a pitch measurement for each pulse point, applies a trimming function to remove any measurements that are too distant from the points that come before and

after, samples pitch at 10 ms intervals, and then applies a triangular smoothing function (Xu, 1999). F0 measurements were then taken at 10 evenly-spaced timepoints across the vowel (0% to 90%).

To normalize f0 across talkers, we took the log of all the f0 measurements, then subtracted a talker-specific mean from each log f0 measurement. This normalization method is ideal for comparing across speakers as it first transforms the values to a logarithmic scale, which corresponds to the way that pitch is perceived. Scaling each talkers' values to a talker-specific mean by subtracting the mean for that talker allows comparison across talkers. Some studies further this normalization by dividing this centered value by the standard deviation of a particular talker's values, thus standardizing to each talkers' pitch range (Zhu, 2004); however, since variation of pitch range between talkers is of interest, we use only this centering method and do not divide by the standard deviation.

Mean f0 was calculated by taking the average of these 10 values. F0 range was calculated by taking the difference between maximum and minimum f0 of the ten values.

To investigate the four acoustic cues of duration, intensity, mean f0, and f0 range, we ran separate Bayesian multilevel models for each cue. Each model included all possible interactions between Language (Mandarin or English), Language Group (L1 Mandarin-L2 English speakers, heritage Mandarin-L1 English speakers, and L2 Mandarin-L1 English speakers), and Speaking Style (Clear or Casual speech). Each model also included random effects of Language and Style for each individual talker. Equation 2.1 shows the model formula.

$$\begin{aligned} \text{Acoustic Cue} \sim & \text{Language} * \text{LanguageGroup} * \text{Style} \\ & + (\text{Language} * \text{Style} | \text{Talker}) \end{aligned} \quad (2.1)$$

We use t-distributions to model the data, including the priors. To calculate the priors for the model for duration, we use means and standard deviations of the four Mandarin tones in monosyllabic speech as reported by Yang et al. (2017). Since we analyze log-duration instead of duration in milliseconds, we sampled a normal distribution with the mean and standard deviation of each tone 1000 times each, then took the log of these values. Then, we used the mean and standard deviation of all these sampled log-duration values (mean = 5.962, s.d. = 0.325) as the mean and scaling for the prior for the intercept. For the priors for population-level effects, between-talker variation, and sigma, we used t-distributions centered on 0 with the same scaling as for the intercept (0.325).

To calculate the prior for the intercept for the model for intensity, we use the average intensity for Mandarin clear and casual speech as reported by Rittenberry and Shport (2024) in Appendix B (mean=72.8 dB). For the scaling, we use the larger of the two reported standard deviation values (7.0 dB) as the scaling for the intercept. For the priors for population-level effects, between-talker variation, and sigma, we used t-distributions centered on 0 with the same scaling as for the intercept (7.0 dB).

As for the priors for the intercept for the model for mean f_0 , since we center f_0 to the mean f_0 for each talker, we expect the intercept to be 0. As for the scaling for the prior for the intercept, we use the standard deviation values for Mandarin clear and casual speech reported by Rittenberry and Shport (2024, Appendix B); however, we use log values instead of Hertz. This resulted in a mean of 0 and a scaling of 0.112 for the prior for the intercept. For the priors for population-level effects, between-talker variation, and sigma, we used t-distributions centered on 0 with the same scaling as for the intercept (0.112 log-Hz).

To calculate the prior for the intercept for the model for f0 range, we use Rittenberry and Shport's (2024) reported average and standard deviation of pitch range values for Mandarin tones in clear and casual speech. To convert from the reported Hz values to log-Hz values, half the reported pitch range for each tone was added and subtracted from the mean pitch for each tone to create a minimum and maximum pitch for clear and casual speech. Then, these Hertz values were converted into log values; then, we took the difference between these min and max values to get a mean pitch range and standard deviation. These values (mean=0.150 log-Hz, SD=0.084 log-Hz) were used as the mean and scaling for the prior of the intercept. For the priors for population-level effects, between-talker variation, and sigma, we used t-distributions centered on 0 with the same scaling as for the intercept (0.084 log-Hz).

2.3 Results

2.3.1 Duration

The model for duration resulted in an intercept of 5.57 (s.d. = 0.03, 95% C.I. = [5.51, 5.63]) with a residual error of 0.24 (s.d. = 0.01, 95% C.I. = [0.22, 0.27]). Results are reported in log-ms, as effect sizes do not scale consistently when converted to ms. Figure 2.1 shows estimates for group means of duration for each language group, language, and speaking style. Overall, clear speech productions were longer than casual speech productions (mean = 0.19, s.d. = 0.02, 95% C.I. = [0.15, 0.23]), and Mandarin productions were longer than English (mean = 0.28, s.d. = 0.03, 95% C.I. = [0.23, 0.33]). Compared to the overall mean, HL Mandarin talkers produced overall longer productions (mean = 0.10, s.d. = 0.04, 95% C.I. = [0.02, 0.17]), an effect which was stronger for English productions and reduced for Mandarin productions (mean = -

0.06, s.d. = 0.03, 95% C.I. = [-0.13, 0]), while L2 Mandarin talkers produced overall shorter productions (mean = -0.08, s.d. = 0.04, 95% C.I. = [-0.16, 0]), especially in Mandarin (mean = -0.05, s.d. = 0.04, 95% C.I. = [-0.12, 0.02]).

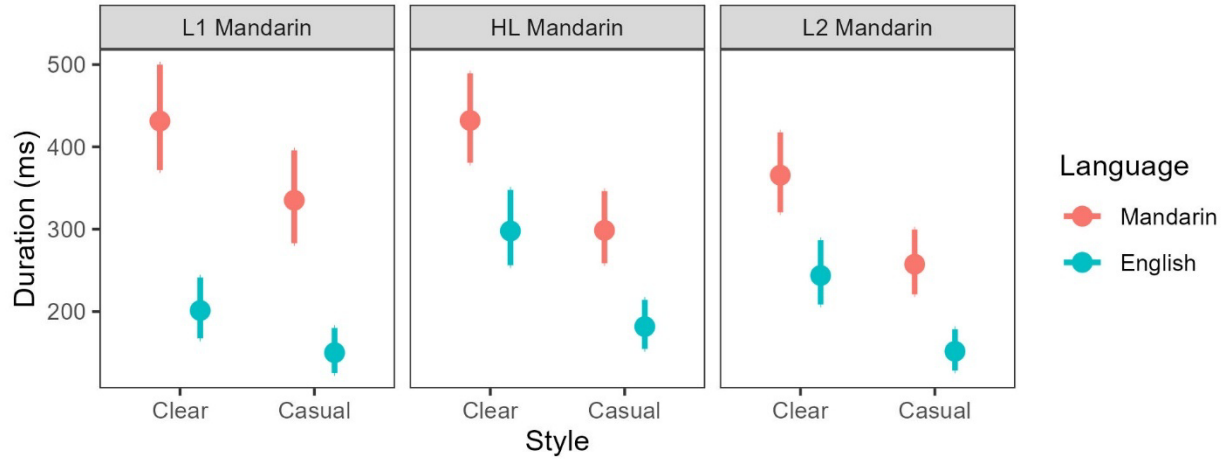


Figure 2.1. 95% credible intervals for group means for duration, for each combination of Talker Language Group (L1 Mandarin-L2 English, HL Mandarin-L1 English, and L2 Mandarin-L1 English), Language (English or Mandarin), and Speaking Style (Casual or Clear).

Figure 2.2 shows the size of the clear speech effect for different talker groups. All talker groups show a positive effect of speaking style on duration, for both languages, with clear speech productions longer than casual speech productions. We see a slightly smaller clear speech effect for Mandarin than for English (mean = -0.02, s.d. = 0.01, 95% C.I. = [-0.05, 0]). We also see an interaction between Language and Language Groups, with L1 Mandarin speakers producing even longer Mandarin speech (mean = 0.11, s.d. = 0.04, 95% C.I. = [0.04, 0.19]) and HL Mandarin speakers producing shorter Mandarin speech (mean = -0.06, s.d. = 0.03, 95% C.I. = [-0.13, 0]). There were no interactions between Language Group and Speaking Style; there were also no three-way interactions between Language, Language Group, and Speaking Style.

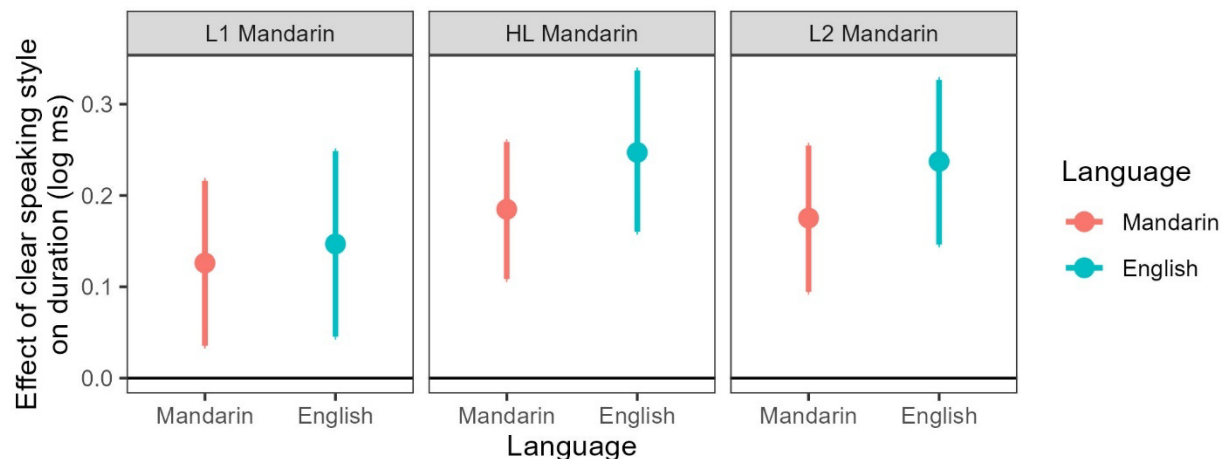


Figure 2.2. 95% credible intervals for the effect of Clear speech on duration, for each talker language group and for both Mandarin and English.

There was variation in duration among talkers; systematic between-talker variation in duration averaged about 0.14 log ms (s.d. = 0.02, 95% C.I. = [0.10, 0.19]). Talkers also varied in the magnitude of effects on duration, as indicated by random effect standard deviations: 0.13 log ms for Language (s.d. = 0.02, 95% C.I. = [0.09, 0.18]), 0.10 log ms for Speaking Style (s.d. = 0.02, 95% C.I. = [0.07, 0.15]), and 0.06 log ms for the Language $\times$ Speaking Style interaction (s.d. = 0.01, 95% C.I. = [0.03, 0.09]). Figure 2.3 shows estimates for individual talkers' clear speech effect on duration for English and Mandarin. Almost all the talkers show a positive effect of clear speech on duration, with clear speech productions longer in both English and Mandarin. However, some talkers show an effect on duration for only one language, and a few talkers do not seem to use duration as a clear speech cue.

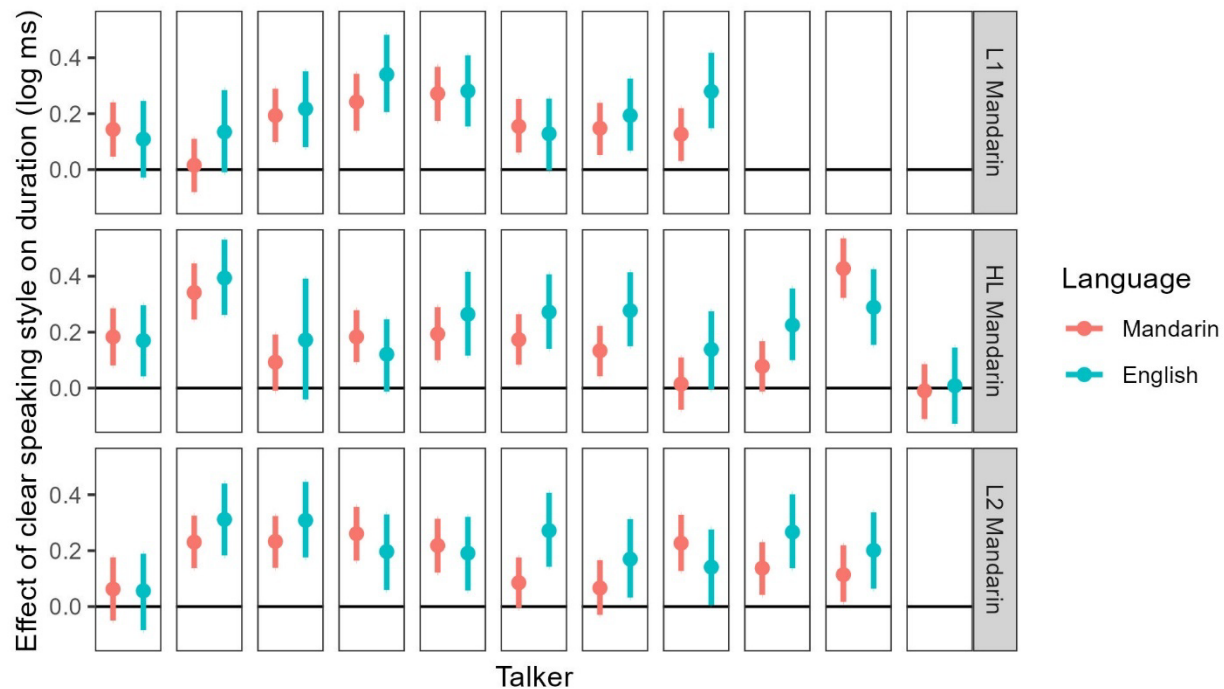


Figure 2.3. 95% credible intervals for estimates of the clear speech effect on duration for each individual talker, for Mandarin (red) and English (teal). Each facet panel represents an individual talker, and panels are ordered with L1 Mandarin speakers in the first row, HL Mandarin speakers in the second row, and L2 Mandarin speakers in the last row.

2.3.2 Intensity

The model for intensity resulted in an intercept of 67.1 dB (s.d. = 1.07, 95% C.I. = [64.99, 69.19]) with an estimated magnitude of residual error of 2.65 dB (s.d. = 0.07, 95% C.I. = [2.51, 2.77]). Figure 2.4 shows estimates for group means of intensity for each language group, language, and speaking style. Overall, clear speech productions were slightly louder than casual speech productions (mean = 0.75, s.d. = 0.19, 95% C.I. = [0.38, 1.11]), an effect size which is under the just noticeable difference in speech perception (Long, 2014). There was no difference

between the loudness of Mandarin productions as compared to English productions (mean = -0.31, s.d. = 0.22, 95% C.I. = [-0.73, 0.12]). There were no differences between the overall loudness of productions made by L1, HL, or L2 Mandarin talkers.

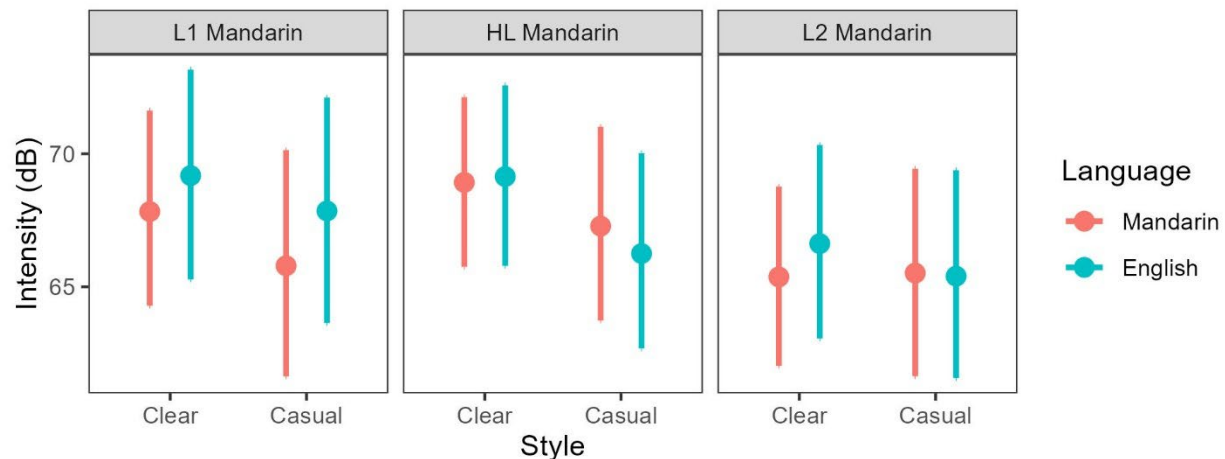


Figure 2.4. 95% credible intervals for estimates of intensity for each combination of Talker Group, Language, and Speaking Style.

Figure 2.5 shows the effect of clear speech on intensity for each speaker group. Differences in intensity across speaking styles were very small, usually less than the just noticeable difference in intensity (1 dB; Long, 2014). There were statistically detectable differences in intensity for L1 and HL Mandarin talkers in English and Mandarin clear speech, and for L2 talkers for English clear speech, but not for L2 talkers in Mandarin clear speech. All effect sizes and interactions between Language, Language Group, and Speaking Style were very small.



Figure 2.5. 95% credible intervals for estimates of the effect of clear speaking style on intensity in Mandarin and English, for L1, HL, and L2 Mandarin talkers.

There was a great deal of variation in intensity among talkers; systematic between-talker variation in intensity averaged about 5.6 dB (s.d. = 0.79, 95% C.I. = [4.31, 7.37]). This between-talker variation is much larger than the size of the fixed effects of Speaking Style, reflecting the tendency of different talkers to speak at different overall decibel levels. Talkers also varied in the magnitude of effects on intensity, as indicated by random effect standard deviations: 1.05 dB for Language (s.d. = 0.19, 95% C.I. = [0.74, 1.49]), 0.86 dB for Speaking Style (s.d. = 0.16, 95% C.I. = [0.6, 1.21]), and 0.25 dB for the Language $\times$ Speaking Style interaction (s.d. = 0.13, 95% C.I. = [0.03, 0.52]). Figure 2.6 shows estimates for individual talkers' clear speech effect on intensity for English and Mandarin. About half of talkers in all language groups show small increases in intensity in both English and Mandarin clear speech. A few L2 Mandarin talkers show small increases in intensity in English clear speech, but not in Mandarin clear speech. Many talkers do not use intensity as a clear speech enhancement strategy in either language. Most effect sizes are small, below the just noticeable difference in intensity (Long, 2014), but some individual talkers show effect sizes that are larger.

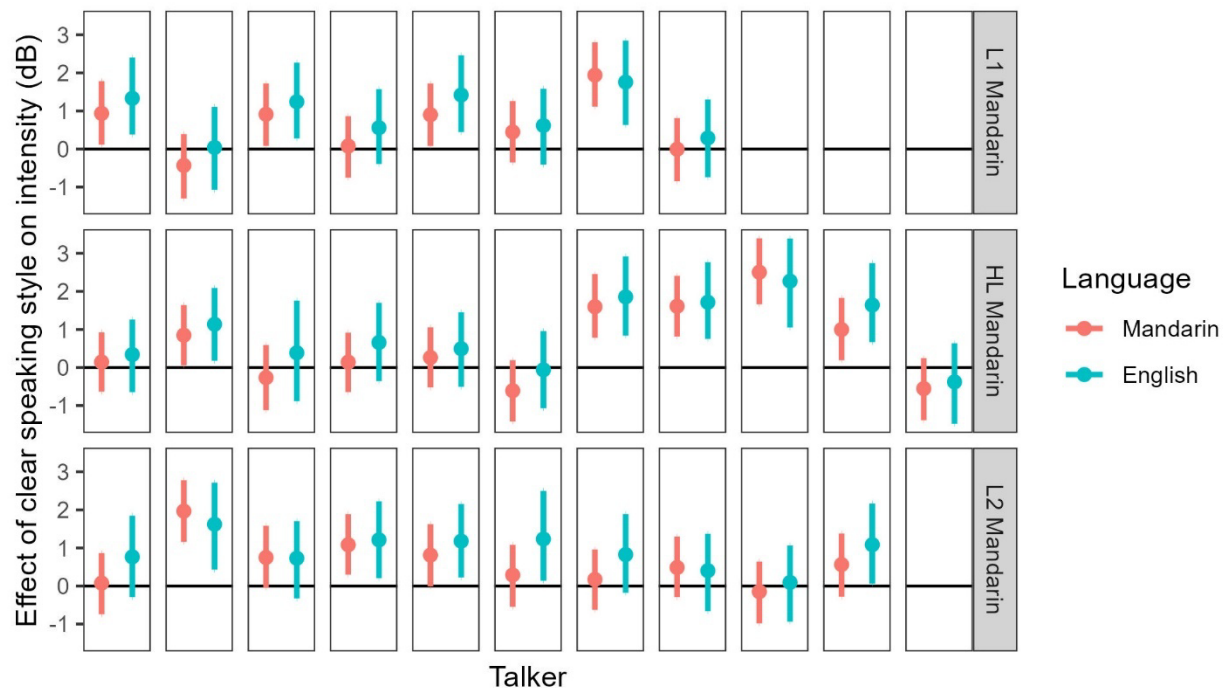


Figure 2.6. 95% credible intervals for estimates for the effect of clear speaking style on intensity for L1, HL, and L2 Mandarin speakers in Mandarin and English.

2.3.3 Mean f0

The model for mean f0 resulted in an intercept of 0.01 log Hz (s.d. = 0.00, 95% C.I. = [0.00, 0.02]) with an estimated magnitude of residual error of 0.11 log Hz (s.d. = 0.00, 95% C.I. = [0.10, 0.11]). Figure 2.7 shows estimates for group means of mean f0 for each language group, language, and speaking style. Overall, the effect of a clear speaking style on f0 mean was very small (mean = 0.01, s.d. = 0.00, 95% C.I. = [0.00, 0.02]), and there was no difference between the mean f0 of Mandarin productions as compared to English productions (mean = 0.00, s.d. = 0.00, 95% C.I. = [-0.01, 0.01]). L1 Mandarin talkers produced higher pitched productions (mean = 0.03, s.d. = 0.01, 95% C.I. = [0.02, 0.04]), while HL and L2 Mandarin talkers produced slightly lower pitched productions (HL: mean = -0.01, s.d. = 0.01, 95% C.I. = [-0.02, 0]; L2:

mean = -0.02, s.d. = 0.01, 95% C.I. = [-0.03, -0.01]). L1 Mandarin talkers seem to overall use higher pitch in English than Mandarin, while L2 Mandarin talkers seem to use higher pitch in Mandarin than English.

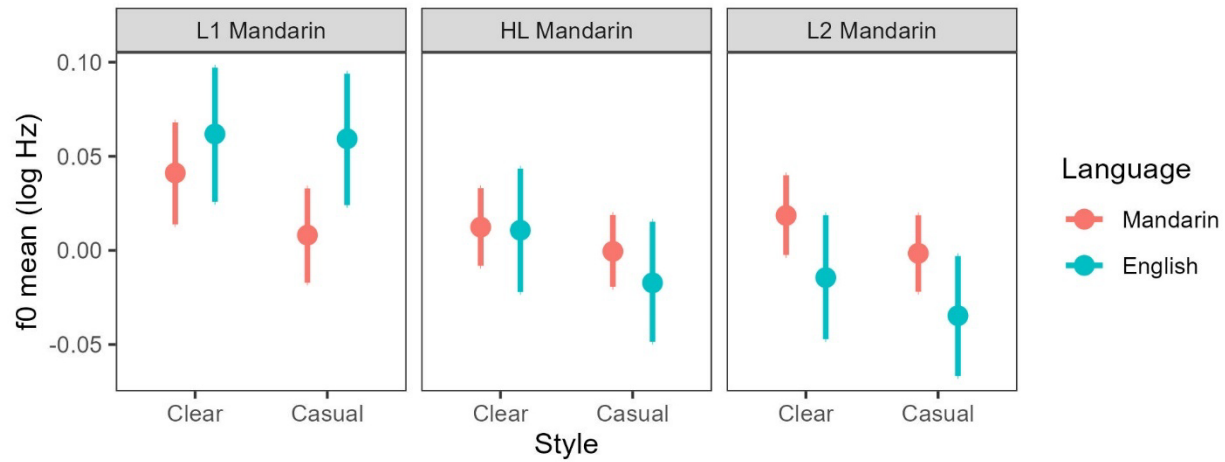


Figure 2.7. 95% credible intervals for estimates of mean f_0 for each combination of Talker Language Group, Language, and Speaking Style.

Figure 2.8 shows credible intervals for estimates of the clear speech effect for each talker group and language. While there is a slight qualitative trend to use higher pitch as a clear speech cue for most talker groups and languages, the effect size is very small, and at a level that might not be noticeable to a listener (Long, 2014). There is little to no effect of clear speech on overall mean pitch for any talker group or language.

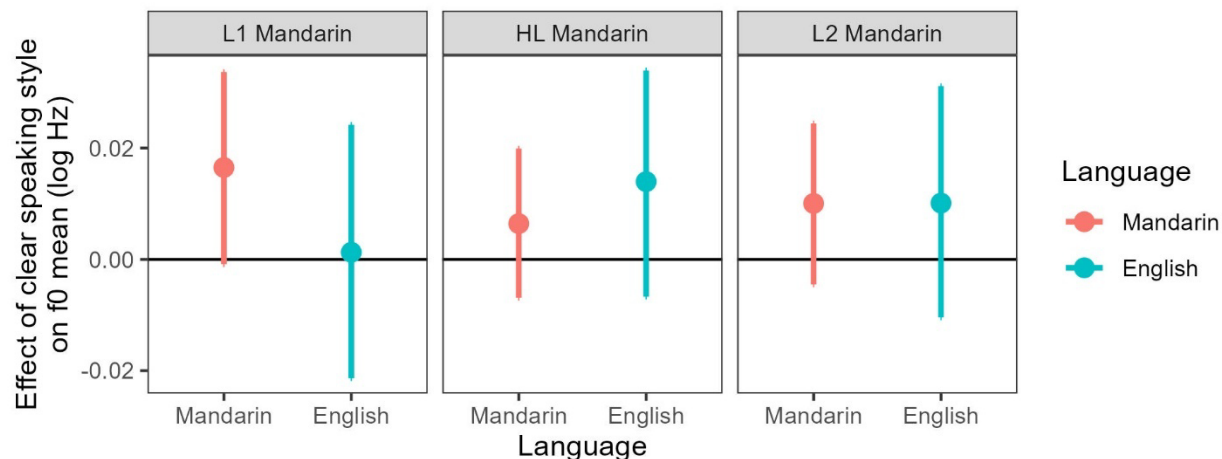


Figure 2.8. 95% credible intervals for the effect of clear speaking style on mean f_0 , for every combination of Talker Language Group, Language, and Speaking Style.

There was a small amount of variation in mean f_0 among talkers; systematic between-talker variation in mean f_0 averaged about 0.01 log Hz (s.d. = 0.00, 95% C.I. = [0.00, 0.02]). Talkers also varied in the magnitude of effects on mean f_0 , with random-effect standard deviations of 0.02 log Hz for Language (s.d. = 0.01, 95% C.I. = [0.01, 0.03]) and 0.01 log Hz for Speaking Style (mean = 0.01, s.d. = 0.01, 95% C.I. = [0.00, 0.02]). There was little to no variation in each individual talker's interaction between Language and Speaking Style (mean = 0.00, s.d. = 0.00, 95% C.I. = [0.00, 0.01]). Figure 2.9 shows estimates for individual talkers' clear speech effect on mean f_0 for English and Mandarin. Only a few talkers seem to show an f_0 raising effect in clear speech. Notably, all talkers show a great deal of similarity across Mandarin and English, suggesting that the size of any clear speech effect for f_0 raising might be similar across languages.

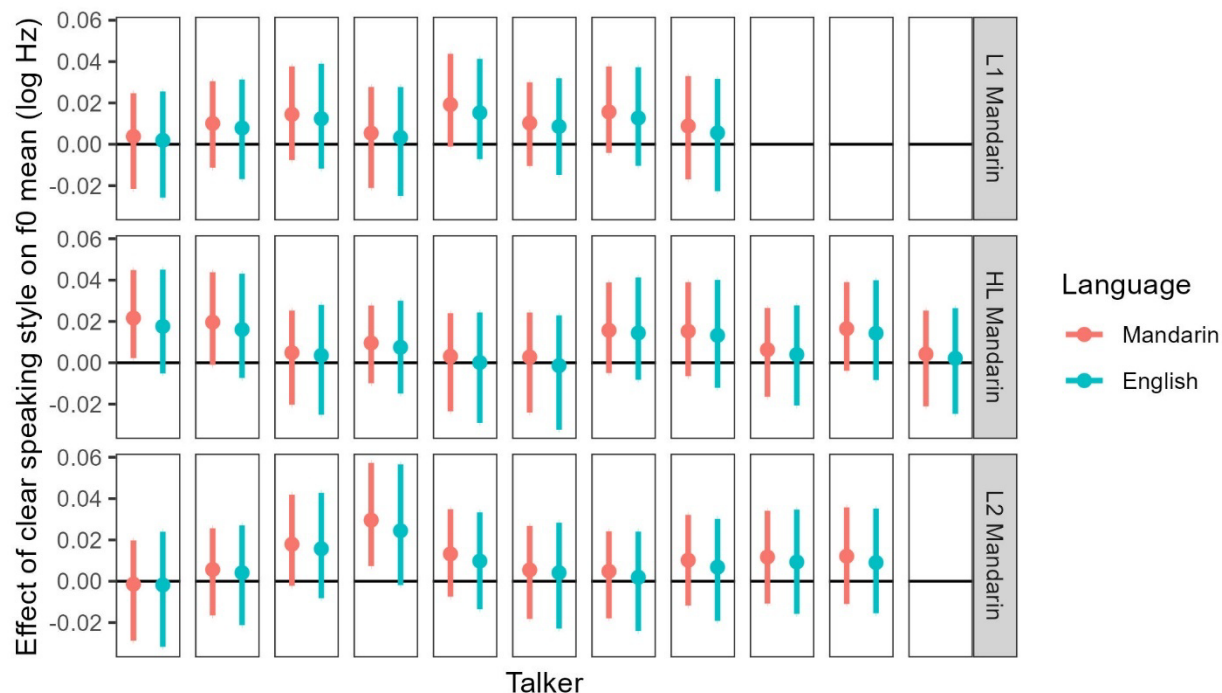


Figure 2.9. 95% credible intervals for the effect of clear speaking style on mean f0, for each individual talker.

2.3.4 F0 range

The model for f0 range resulted in an intercept of 0.25 log Hz (s.d. = 0.01, 95% C.I. = [0.22, 0.28]) with an estimated magnitude of residual error of 0.11 log Hz (s.d. = 0.00, 95% C.I. = [0.10, 0.12]). Figure 2.10 shows estimates for group means of mean f0 for each language group, language, and speaking style. Overall, clear speech productions had a wider f0 range than casual speech productions (mean = 0.04, s.d. = 0.01, 95% C.I. = [0.03, 0.05]), while there was no difference between the f0 range of Mandarin productions as compared to English productions (mean = 0.00, s.d. = 0.01, 95% C.I. = [-0.01, 0.02]). Overall, L1 Mandarin talkers produced speech with a wider f0 range (mean = 0.04, s.d. = 0.02, 95% C.I. = [0.01, 0.08]), while L2

Mandarin talkers produced speech with narrower f0 ranges (mean = -0.05, s.d. = 0.02, 95% C.I. = [-0.08, -0.01]).

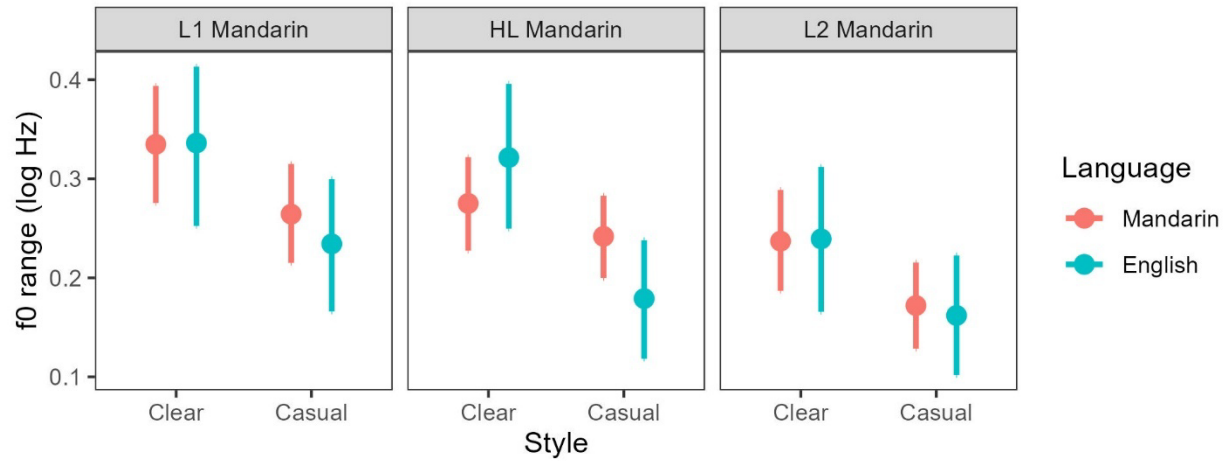


Figure 2.10. 95% credible intervals for the f0 range of L1 Mandarin, HL Mandarin, and L2 Mandarin talkers, in Clear and Casual speech, and in Mandarin (red) and English (teal).

Figure 2.11 shows credible intervals for estimates of the clear speech effect on f0 range for each talker group and language. For all talker groups, we see an expanded pitch range in English clear speech. We also see an expanded pitch range in Mandarin clear speech for L1 and L2 Mandarin speakers. However, HL Mandarin speakers show little or no effect of clear speech on pitch range in Mandarin, despite showing a large effect for English clear speech.

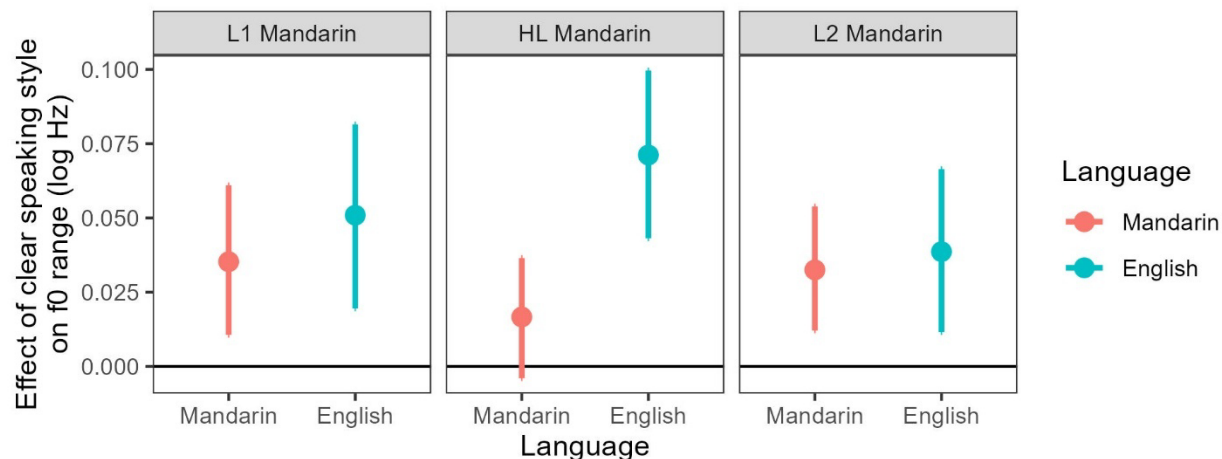


Figure 2.11. 95% credible intervals for the clear speech effect on f0 range of the three talker groups in both Mandarin (red) and English (teal).

There was some variation in f0 range among talkers; systematic between-talker variation in mean f0 averaged about 0.07 log Hz (s.d. = 0.01, 95% C.I. = [0.05, 0.09]). Talkers also varied in the effects on f0 range, with random-effect standard deviations of 0.04 log Hz for Language (s.d. = 0.01, 95% C.I. = [0.03, 0.06]), 0.02 log Hz for Speaking Style (s.d. = 0.01, 95% C.I. = [0.01, 0.03]), and 0.02 log Hz for the Language $\times$ Speaking Style interaction (s.d. = 0.01, 95% C.I. = [0.00, 0.03]). Figure 2.12 shows estimates for individual talkers' clear speech effect on mean f0 for English and Mandarin. Most talkers in all language groups have an expanded f0 range in English clear speech, while some also show an effect for Mandarin. There are some talkers who show an effect for English but not Mandarin.

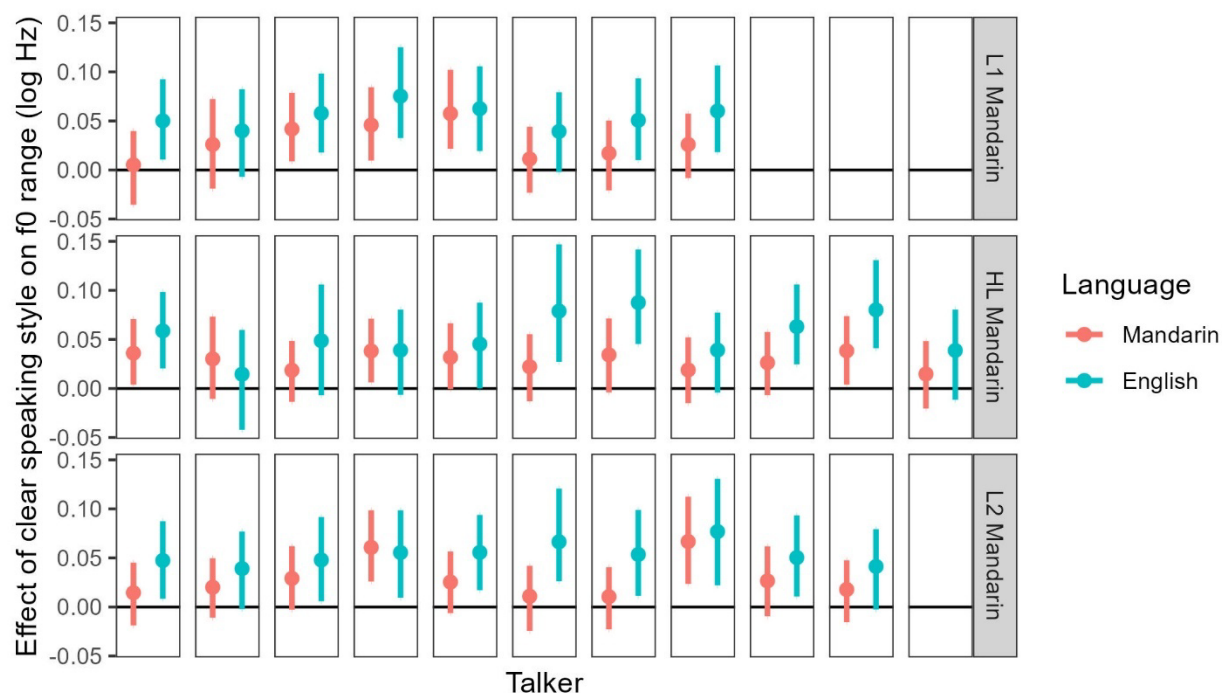


Figure 2.12. 95% credible intervals for the effect of a clear speaking style on f_0 range, for each individual talker.

2.4 Discussion

The present study investigated the signal-enhancing clear speech strategies used by L1, HL, and L2 Mandarin speakers in Mandarin and in English. Table 2.2 shows an overview of the presence of clear speech effects used by each talker group in each language. Overall, all speaker groups used longer duration as a clear speech strategy in both Mandarin and English. Expanded f_0 range was also a clear speech cue for all speaker groups and languages, with the exception that HL Mandarin speakers did not expand f_0 range in Mandarin clear speech. Some statistically-detectable differences were found in the intensity and mean f_0 of clear speech in English and Mandarin, but these differences were smaller than would be detectable to a listener in conversation (Long, 2014).

Table 2.2: Overview of clear speech effects for each talker group, language, and acoustic cue. A plus sign + indicates a large positive effect in clear speech, a plus sign in parentheses (+) indicates a statistically detectable but very small positive effect in clear speech, and blank cells indicate no effect.

Cue	L1 Mandarin, L2 English		HL Mandarin, L1 English		L2 Mandarin, L1 English	
	Mandarin	English	Mandarin	English	Mandarin	English
Duration	+	+	+	+	+	+
Intensity	(+)	(+)	(+)	(+)		(+)
Mean f0	(+)		(+)	(+)	(+)	(+)
F0 range	+	+		+	+	+

Consistent with previous studies which have overwhelmingly found longer duration to be a clear speech strategy across languages (Granlund et al., 2012; Smiljanić & Bradlow, 2005; Tupper et al., 2021; Rittenberry & Shport, 2024; Zhang et al., 2024; Chien & Brajot, 2025; Papoušek & Hwang, 1991; Picheny et al., 1986), all talker groups used longer duration as a clear speech effect in both Mandarin and English. The fact that all groups and languages show an effect of duration suggests that duration is indeed a global strategy for enhancing clear speech, in both Mandarin and English. Additionally, most individual talkers used longer duration as a clear speech strategy; the clear speech effect for duration seems to be one that is common across talkers. Yet, the task instructions likely also influenced the clear speech increase in duration, as the instructions for the casual speaking style included “speak quickly.” Future work in this area could omit this instruction, to allow participants to decide how to enhance or reduce speech without focus on a particular acoustic cue. As for differences across language, productions in

Mandarin were longer than productions in English; this is likely due to the differences in stimuli syllable structure: open V syllables for Mandarin, with closed hVd syllables for English.

As for intensity, L1 Mandarin talkers showed a small increase in intensity as a clear speech enhancement when speaking Mandarin, consistent with previous studies examining intensity in Mandarin clear speech (Rittenberry & Shport, 2024; Tupper et al., 2021). Small differences in higher intensity were also seen in HL Mandarin clear speech, as well as in English clear speech by all speaker groups. However, the size of these effects was small, at or below the just noticeable difference (Long, 2014). When examining individual talkers, we find increases in intensity in clear speech in about half of talkers in each language group. Some L2 Mandarin talkers displayed increased intensity in English clear speech, but not in Mandarin clear speech, displaying a difference in clear speech cues used in their first versus their second language. However, only a few talkers showed differences in intensity above the just noticeable difference, perhaps reflecting a tendency of a few talkers to use a “speak louder” strategy (Rittenberry & Shport, 2024).

While there is a qualitative trend toward increasing mean f_0 for clear speech in Mandarin and English, the effect size of increase in mean f_0 is very small, very close to the just noticeable difference for change in pitch (Long, 2014). When examining individual speakers, only a few showed larger increases in mean f_0 for clear productions in either English or Mandarin. This small effect of mean f_0 clear and casual speech may result from the “speak quickly” instruction in the casual speech condition, as pitch increases in fast speech (Tang and Li, 2020; Cooper & Sorenson, 2012). It is also possible that an increase in mean f_0 in L1 Mandarin clear speech is driven by tone-specific enhancements where some tones are raised but not others, resulting in the small effect found by the present study. Tone-specific f_0 enhancements are examined in Chapter

3. As differences in intensity and mean f0 were typically below the size that would be noticeable to a listener, these differences may reflect increased underlying effort from the speaker in clear speech, instead of listener-oriented clear speech adjustments.

As for f0 range, all groups showed an increase in f0 range for English clear speech productions. Both L1 and L2 talkers also showed an increase in f0 range for Mandarin clear speech productions, consistent with previous research on f0 range in Mandarin clear speech (Rittenberry & Shport, 2024; Zhang et al., 2024; Chien & Brajot, 2025). However, HL Mandarin talkers did not increase the range of clear speech in Mandarin. When examining individual talkers, almost every single talker in every speaker group increased the f0 range of English clear speech. However, only some of the talkers increased pitch range in Mandarin, perhaps reflecting a tendency for some speakers to preserve the same pitch range across different speech styles instead of hyper-articulating tones to increase the pitch range.

For all cues, we see variation between talkers in the size of acoustic modifications. This is consistent with previous work that has found variation in which cues are enhanced by different speakers in clear speech (Rittenberry & Shport, 2024; Perkell et al., 2002). For the present study, duration seems to be the most consistently used cue, while pitch range is also a common cue. Increased intensity is also common, but the size of intensity difference is small. Mean f0 is used very little or not at all.

The extent of individual variation between cues used by different talkers with the same language background in this study is an interesting finding. Previous research has demonstrated that the size of L2 clear speech enhancements may be modified by factors such as proficiency (Kato & Baese-Berk, 2021; 2022; 2023) and age of immersion (Rogers et al., 2010). The present study did not examine or control for factors like language proficiency or amount of current or

previous exposure to a second or heritage language. The L2 speakers in the study had a wide range of experience with tone languages, including some with no prior tone language experience, several who are heritage speakers of a tone language, and one who lived in China for a few years as a child. This variation in experience could influence which clear speech cues are used by the different L2 talkers in this study. Studies that have controlled for HL speaker characteristics have found a reduced amount of intertalker variability (Chang, 2024; Chang & Weiss-Cowie, 2021). However, variation within clear speech effects is larger than just the question of L2 or HL language experience and exposure, as L1 talkers also show variation in which cues they choose to enhance in clear speech (Rittenberry & Shport; Perkell et al., 2002). Investigating which factors might influence a speaker's use of certain acoustic cues but not others is an important direction for future clear speech research.

One limitation of the current study is the monosyllabic nature of the stimuli. Talkers might be expected to use different clear speech strategies in a monosyllabic context, as compared to a sentential context or a conversational context. For example, Zhang et al. (2024) found that L1 Mandarin speakers enhanced f_0 range, but not f_0 mean, when reading a Mandarin passage in clear speech; yet, when reading individual words, L1 Mandarin speakers enhanced both f_0 mean and f_0 range, but only for some of the Mandarin tones. Some of this variation in results may stem from the degree of individual variability in clear speech strategies used by the different talkers in different studies. Another limitation of the present study is that the syllable structure differed for Mandarin (V) as compared to English (hVd) syllables; in future research, either comparing across similar syllable structures (e.g., CV) or comparing across a longer passage of text would be better to compare across these two languages. Additionally, the present study had the clear speech block ordered before the casual speech block, which we theorized to allow for a more

casual speaking style after the participants had gained familiarity with the stimuli. However, this ordering diverges from common practice in clear speech research which usually orders casual speech blocks before clear speech blocks (e.g., Bradlow & Bent, 2002), which may impact direct comparison of the results between studies. A further limitation to the study's design is the use of separate models to investigate the different acoustic cues, as this approach may miss correlations between these cues.

Since Mandarin tones differ in their mean pitch and pitch contours, we can hypothesize that different tones might call for different code-based enhancement strategies. In the present study, we did not account for differences between the four Mandarin tones. In Chapter 3, we extend these findings with an investigation into code-specific tone enhancements in Mandarin clear speech, specifically examining the clear speech strategies used for each tone.

Although we have found some acoustic signal-based differences in the clear speech of L1, HL, and L2 Mandarin speakers in both English and Mandarin, it is unclear whether these acoustic differences will actually result in an increase in intelligibility for a listener. In Chapter 4, we investigate the impact of a clear speaking style on intelligibility of Mandarin tones produced by L1, HL, and L2 talkers.

In conclusion, producing slower speech with a wider pitch range is the primary strategy for clear speech production. Increases in duration and pitch range seen for all speaker groups for English and Mandarin, with the exception of HL Mandarin speakers, who expanded f_0 pitch range in English but not in Mandarin. Some statistically detectable differences in mean pitch and intensity were found for clear speech, but these effects were mostly very small (i.e., below the threshold that a human listener could notice in normal conversation; Long, 2014). Yet, these differences in intensity and mean f_0 may still reflect increased effort during clear speech

production. There is a great deal of variation between which cues talkers use to enhance clear speech, with duration being the most consistently used cue.

Chapter 3. Enhancement of tonal contours in the clear speech productions of L1, HL, and L2 Mandarin speakers

3.1 Introduction

Clear speech is a speaking style where speakers enhance their speech to be more intelligible to a listener (Bradlow, 2002). In general, clear speech is slower, longer, louder, and has higher pitch, greater pitch range, and an expanded vowel space as compared to casual speech (Picheny et al., 1986; Bradlow et al., 2003; Smiljanić & Bradlow, 2005; Ferguson & Kewley-Port, 2007; Hazan & Baker, 2011; Han et al., 2021; Lam & Tjaden, 2016; Bradlow et al., 2003). However, speakers may also enhance phonetic contrasts in clear speech in language-specific ways. For example, Smiljanić and Bradlow (2008) found that stop closure duration of voiceless stops is increased in English clear speech, enhancing the difference between voiced and voiceless stops, while Croatian did not show the same enhancement for its voiced and voiceless stops. Both languages showed longer vowel duration in clear speech, with language-specific vowel length contrasts enhanced and preserved in clear speech. What about in a tonal language like Mandarin: with a balance between the global clear speech strategy of greater pitch range and the need to preserve tone contour, how are tones enhanced in clear speech?

Mandarin Chinese is a lexical tonal language with four lexical tones (see Table 1.1). The contours of these tones may be modelled using growth curve analysis as second-degree polynomials, with parameters of mean, slope, and curve (Tupper et al., 2020; Mirman, 2017). As a high level tone, Tone 1 is characterized by a large mean f_0 and an f_0 slope that is close to zero. The rising Tone 2 is characterized by a positive f_0 slope. In isolation, falling-rising Tone 3 is

characterized by a low mean f_0 and a positive, convex curve (similar to a U shape). The falling Tone 4 is characterized by a negative f_0 slope and a negative f_0 curve (Tupper et al., 2020). Note that Tone 4's negative, concave curve comes from the fact that there is often a high steady pitch before the falling portion of the tone (see for example Figure 2 in Xu, 1997). This steady portion before the fall mimics the shape of the second half of an upside-down U, leading to Tone 4's negative curve in a quadratic fit.

There are mixed findings on whether these characteristic features of the Mandarin tones are enhanced in clear speech. To the author's knowledge, only two studies have examined mean, slope, and curve in Mandarin clear speech. Tupper et al. (2021) found that mean of Tone 2 and Tone 3 increased in clear speech, while Zhang et al. (2024) found that mean of Tone 1 and Tone 2 were higher in clear speech, while mean of Tone 3 was lower in clear speech. Using raw f_0 values to calculate f_0 mean instead of mean f_0 from a polynomial fit, Rittenberry and Shport (2024) found that f_0 mean increased for all tones in clear speech, while both Zhang et al. (2024) and Chien and Brajot (2025) found that f_0 mean increased for Tone 1 and Tone 2. As for slope, Tupper et al. (2021) found that Tone 4 became steeper in clear speech, while Zhang et al. (2024) found that both Tone 2 and Tone 4's slopes became steeper in clear speech. Chien and Bradot (2025) calculated f_0 slope as the f_0 range divided by the duration between time points of maximum and minimum f_0 , and found that f_0 slope became steeper for both T3 and T4. Rittenberry and Shport (2024) calculated f_0 slope as the slope of f_0 in the second half of the syllable (instead of from a polynomial), and found no effect of clear speech on f_0 slope thus calculated. Neither Zhang et al. (2024) nor Tupper et al. (2021) found any difference in curve for any of the Mandarin tones.

Overall, for L1 Mandarin speakers, there is evidence for some effects of tone enhancement in clear speech, with increases in mean for Tone 1 and Tone 2 and steeper slopes for Tone 2, Tone 3, and Tone 4, as well as changes in Tone 3 mean. Yet, it is unknown what tone enhancement strategies, if any, may be used by HL and L2 speakers.

L2 and heritage speakers make acoustic adjustments between clear and casual speech (Smiljanić & Bradlow, 2007; Bradlow & Smiljanić, 2011; Granlund et al., 2012; Chang, 2024; Chang & Weiss-Cowie, 2021). The size of these clear speech adjustments may vary based on age of immersion (Rogers et al., 2010) and proficiency (Kato & Baese-Berk, 2021; 2022; 2023). Clear speech adjustments produced by HL and L2 speakers may show influence from the dominant language for HL speakers, or from the L1 for L2 speakers. For example, Chang (2024) found that HL Korean-L1 English speakers produced the Korean central vowel /a/ in a central, L1 Korean-like way in clear speech, but in a further back, more English-like way in reduced Korean speech. As for L2 speakers, Kato and Baese-Berk (2021) found that L1 Mandarin-L2 English speakers were just as effective as L1 English speakers at enhancing a phonetic contrast present in both English and Mandarin (word-initial /p/ versus /b/), but that lower proficiency L2 speakers did not enhance an L2-only phonetic contrast (word-final /p/ versus /b/) in the same ways or to the same extent as L1 and higher proficiency L2 speakers. This demonstrates the impact of a talker's L1 on the enhancement of language-specific contrasts in an L2.

For L2 Mandarin speakers, Mandarin tonal contrasts are challenging to acquire (Pelzl et al., 2019). Both L1 English, L1 Cantonese, and HL Cantonese speakers show difficulties with L2 Mandarin tone acquisition, and struggle to distinguish Tone 2 and Tone 3 perceptually (Hao, 2012; Tsukada et al., 2015); Tone 2 and Tone 3 also show lower levels of accuracy in production for L2 talkers (Hao, 2012). Even at very high levels of proficiency, though overall very accurate

in tone production, L1 English-L2 Mandarin speakers continue to produce tone errors, with the most production errors in Tone 2 (Song, 2021). Overall, L2 speakers show a smaller pitch range and tone space than L1 speakers, while HL Mandarin-L1 English speakers pattern in between L1 and L2 Mandarin speakers in regards to pitch range and tone space (Yang, 2015). HL speakers pattern with L1 speakers in producing more distinct Tone 2 and Tone 3 contours than L2 speakers (Chang & Yao, 2016), although HL Mandarin speakers also exhibit more variation in tone contour production than L1 and L2 speakers (Yang, 2015; Chang & Yao, 2016). Just as HL and L2 Mandarin speakers pattern differently from L1 speakers in many aspects of tone production, it is possible that HL and L2 speakers may pattern differently from L1 speakers in terms of tone enhancement.

The present study investigates the extent to which L1 Mandarin-L2 English, HL Mandarin-L1 English, and L2 Mandarin-L1 English speakers enhance contrasts between Mandarin tones, as characterized by polynomial mean, slope, and curve. If these speakers are enhancing contrasts between tones, then we expect the mean to be higher for tones with overall high pitch (Tone 1), and lower for tones with overall low pitch (Tone 3). We expect slope to be steeper for tones with large positive slope (Tone 2) and large negative slope (Tone 4). Correspondingly, we expect the curve to be curvier for tones with a large positive curve (Tone 3) and large negative curve (Tone 4). On the other hand, if speakers are not enhancing contrasts between tones, then we would expect to see no change between clear and casual speech, or a unidirectional change such as increased mean for all tones.

As for the effect of language background, earlier age of acquisition with Mandarin could lead to larger size of enhancements for both L1 and HL speakers as compared to L2 speakers (Rogers et al., 2010). HL speakers may even have larger enhancements than L1 speakers (Chang

& Weiss-Cowie, 2021; Chang, 2017). We may also see enhancement of different acoustic characteristics for HL and L2 speakers as compared to L1 speakers (Kato & Baese-Berk, 2021).

3.2 Methods

3.2.1 Participants

Data from the same participants as in Chapter 2 was used in this experiment. 8 L1 Mandarin-L2 English speakers, 11 HL Mandarin-L2 English speakers, and 10 L2 Mandarin-L1 English speakers were recruited to participate in this study. See Section 2.2.1 for a description of the participants.

3.2.2 Materials

The same word list was used for this experiment as for the experiment in Chapter 2. The word list for the experiment is made up of 12 Mandarin monosyllabic words - the three vowels /i, ɤ, u/ crossed with the four Mandarin lexical tones (Tupper et al., 2020; 2021). English hVd words with the vowels /i, ʌ, u/ were also used: *heed*, *hud*, and *who'd*. The same word order lists, which each contained two instances of the target words, were also used. See Section 2.2.2 and Table 2.1 for a description of the Mandarin and English stimuli used for the study.

3.2.3 Methods

We used the same recordings as described in Chapter 2. See Section 2.2.3 for a more detailed description of the methods. Participants familiarized themselves with the English words prior to the study and completed informed consent. HL and L2 Mandarin participants completed a character familiarity task (Bent & Bradlow, 2003). Then, participants completed the recording

task. For the clear speech portion, participants were asked to “Speak clearly as if you are talking to someone who is having a hard time understanding you” (Zellou et al., 2022), and they were told that their recordings would be played to language learners to help them learn. For the casual speech trials, participants were asked to “Speak as if you are talking to a friend or family member you have known for a long time who has no trouble understanding you, and speak quickly” (Zellou et al., 2022). Participants were randomly assigned to one of two trial orders. All participants read each target word aloud twice in each speech style, for a total of 24 clear Mandarin productions, 24 casual Mandarin productions, 6 clear English productions, and 6 casual English productions. After the recording task, participants completed the LEAP-Q questionnaire with information about their language backgrounds (Marian et al, 2007; Kaushanskaya et al, 2020).

3.2.4 Analysis

The same segmentation and pulse correction criteria as in Chapter 2 were used; see Section 2.3.4 for a more detailed description. Productions were segmented and pulses were corrected in Prosody Pro (Xu, 2013). F0 measurements were taken at 10 evenly-spaced timepoints across the vowel (0% to 90%). To normalize f0 across talkers, we took the log of all the f0 measurements, then subtracted a talker-specific mean from each log f0 measurement.

To get f0 mean, slope, and curve, we follow Tupper et al. (2021; 2020) who fit second-degree polynomials to f0 contours to get the coefficients c_0 , c_1 , and c_2 in Equation 3.1, which correspond to f0 mean, slope, and curve (when the time interval is set to 0 to 1). When this equation is expanded, it becomes clear that these coefficients may be calculated by fitting a polynomial to the f0 contours, then taking the 2nd-degree coefficient as the f0 curve, the slope of

a line from start and endpoint of the polynomial as the f0 slope, and the area under the curve as the f0 mean (since duration is 1). Tupper et al. (2020) found that these measures are broadly effective to characterize the Mandarin tones. Modeling these curves as polynomials also allows us to later graph our statistical models' estimates to show tone contour estimates for tones produced by different language groups in different speech styles.

$$f(t) \approx c_0 + c_1(t - 1/2) + c_2[(t - 1/2)^2 - 1/12] \quad (3.1)$$

For the analysis of the f0 contour, we ran three separate Bayesian multilevel models to investigate f0 mean, f0 slope, and f0 curve, each including fixed effects of speaking style (Clear or Casual speech), tone (the four Mandarin tones plus English), and Language Group (Heritage Mandarin speakers, L1 Mandarin speakers, and L2 Mandarin speakers). All possible interactions between these factors were also included. The models also include random effects of the Style and Tone factors for each individual talker. The model equations are shown in Equation 3.2.

$$Cue \sim Style * Tone * LanguageGroup + (Style * Tone|Talker) \quad (3.2)$$

We use t-distributions to model the data, including the priors. Since mean, slope, and curve can be both positive and negative but are expected to be small (less than 1), we used a prior of 0 with a scaling of 1. For all other priors, we used t-distributions centered on 0 with a scaling of 0.5.

3.3 Results

3.3.1 f0 mean

The model for mean resulted in an intercept estimate of 0 log Hz (s.d. = 0.002, 95% C.I. = [-0.003, 0.003]), with an estimated residual error of 0.034 (s.d. = 0.001, 95% C.I. = [0.031,

0.037]). The intercept of 0 is expected due to the method of normalization across talkers, where each talker's f0 values were centered by subtracting the talker's mean f0 from each individual value, resulting in a mean f0 of zero for each talker.

Figure 3.1 shows the credible intervals for the group means of each combination of Style, Tone, and Language Group for the polynomial estimate of f0 mean. Overall, we see a very small main effect for Style, with a very small increase in f0 mean in clear speech (mean = 0.009, s.d. = 0.004, 95% C.I. = [0.001, 0.017]). We can see that Tone 1 and Tone 4 have an overall higher pitch (T1 mean = 0.129, s.d. = 0.011, 95% C.I. = [0.107, 0.15]; T4 mean = 0.099, s.d. = 0.012, 95% C.I. = [0.075, 0.121]), while Tone 2 and Tone 3 have overall lower pitches (T2 mean = -0.052, s.d. = 0.008, 95% C.I. = [-0.068, -0.036]; T3 mean = -0.189, s.d. = 0.012, 95% C.I. = [-0.215, -0.166]).

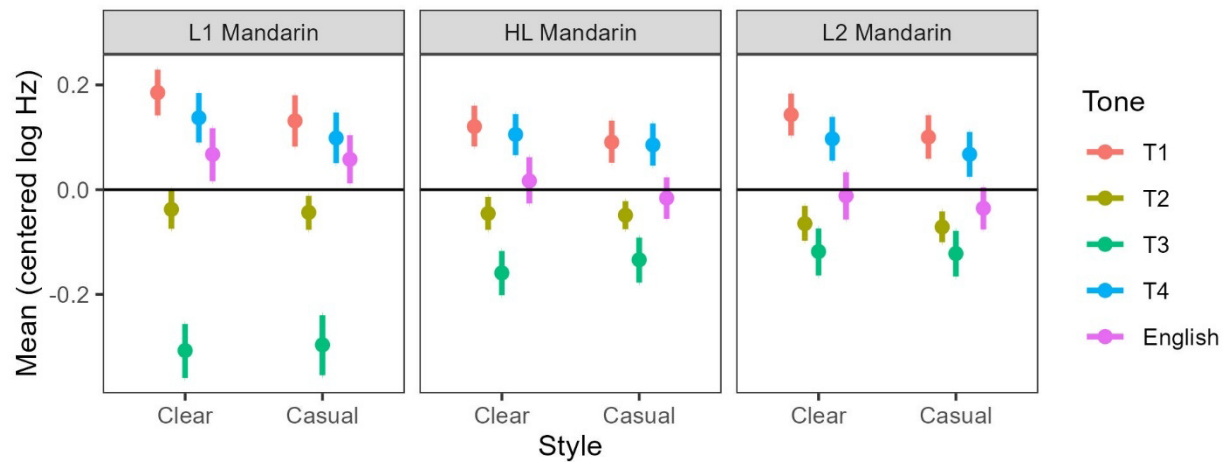


Figure 3.1. 95% credible intervals for the group means of each combination of Speech Style, Tone, and Language Group on f_0 mean. An overall increase in f_0 mean (global clear speech strategy) would appear as Clear productions being overall higher than Casual productions. An expansion in f_0 mean differences (tone enhancement strategy) would appear as an expansion in the Clear space means as compared to the Casual speech means (high means become higher, low means become lower).

Figure 3.2 shows the simple effect of Style on f_0 mean for each Tone and Language Group combination. We can see that mean f_0 increases for L1 speakers for Tone 1 and Tone 4. For L2 speakers, Tone 1 in clear speech has a higher mean f_0 than in casual speech, but there is no effect of clear speech on the mean f_0 for the other tones or for English. For heritage speakers, we also do not see an effect of clear speech on the mean f_0 of any of the tones or English. As for interactions, we see a small impact of Tone on the effect of Style on mean f_0 , with a slightly increased effect for Tone 1 (mean = 0.012, s.d. = 0.004, 95% C.I. = [0.005, 0.02]) and for Tone 4 (mean = 0.006, s.d. = 0.003, 95% C.I. = [0, 0.012]), and slightly decreased effects for Tone 2 (mean = -0.006, s.d. = 0.003, 95% C.I. = [-0.012, -0.001]) and Tone 3 (mean = -0.014, s.d. = 0.004, 95% C.I. = [-0.022, -0.006]). We also see interactions between Tone and Language Group, with Tone 3 produced with a lower mean for L1 Mandarin speakers (mean = -0.112, s.d. = 0.018, 95% C.I. = [-0.149, -0.075]).

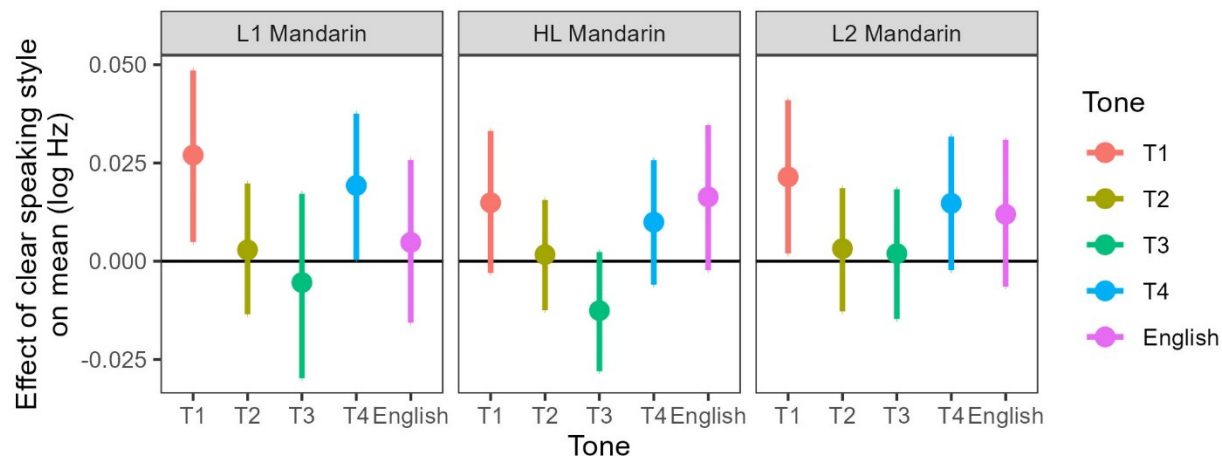


Figure 3.2. 95% credible intervals for the effect of clear speaking style on the polynomial estimate of mean, for each tone and for each language group.

Talkers varied very little in f_0 mean, with systematic between-talker variation in mean f_0 of 0.004 log Hz (s.d. = 0.002, 95% C.I. = [0, 0.009]). Again, this is expected due to the normalization method. Talkers did vary in the mean f_0 of each tone (T1 mean = 0.056, s.d. = 0.008, 95% C.I. = [0.042, 0.075]; T2 mean = 0.041, s.d. = 0.007, 95% C.I. = [0.03, 0.057]; T3 mean = 0.063, s.d. = 0.009, 95% C.I. = [0.048, 0.083]; T4 mean = 0.06, s.d. = 0.009, 95% C.I. = [0.046, 0.079]). Although talkers did vary in the effect of Style, as well as differing effects of Style for each tone, the size of this variance was smaller than that of variance in mean f_0 for each tone (Style mean = 0.02, s.d. = 0.003, 95% C.I. = [0.015, 0.027]; T1:Style1 mean = 0.015, s.d. = 0.004, 95% C.I. = [0.009, 0.023]; T2:Style1 mean = 0.009, s.d. = 0.004, 95% C.I. = [0.001, 0.017]; T3:Style1 mean = 0.014, s.d. = 0.004, 95% C.I. = [0.006, 0.022]; T4:Style1 mean = 0.008, s.d. = 0.004, 95% C.I. = [0.001, 0.016]).

Figure 3.3 shows credible intervals for the simple effect of clear speech on mean for each individual talker. A few talkers enhance T1 and T4's high positive mean in clear speech. As for T2 and T3, one L1 talker and two L2 talkers enhance T2's negative mean, and one HL and one

L2 talker enhance T3's negative mean. Some talkers increase the mean of clear speech in English, and one HL Mandarin talker actually decreases f0 mean of English clear speech. One L2 Mandarin talker increases f0 mean for all tones and English. Whenever there is an effect of speaking style on f0 mean, the effect is usually to increase f0 mean, and this effect usually occurs for tones with a high mean f0 (T1, T4, and sometimes English).

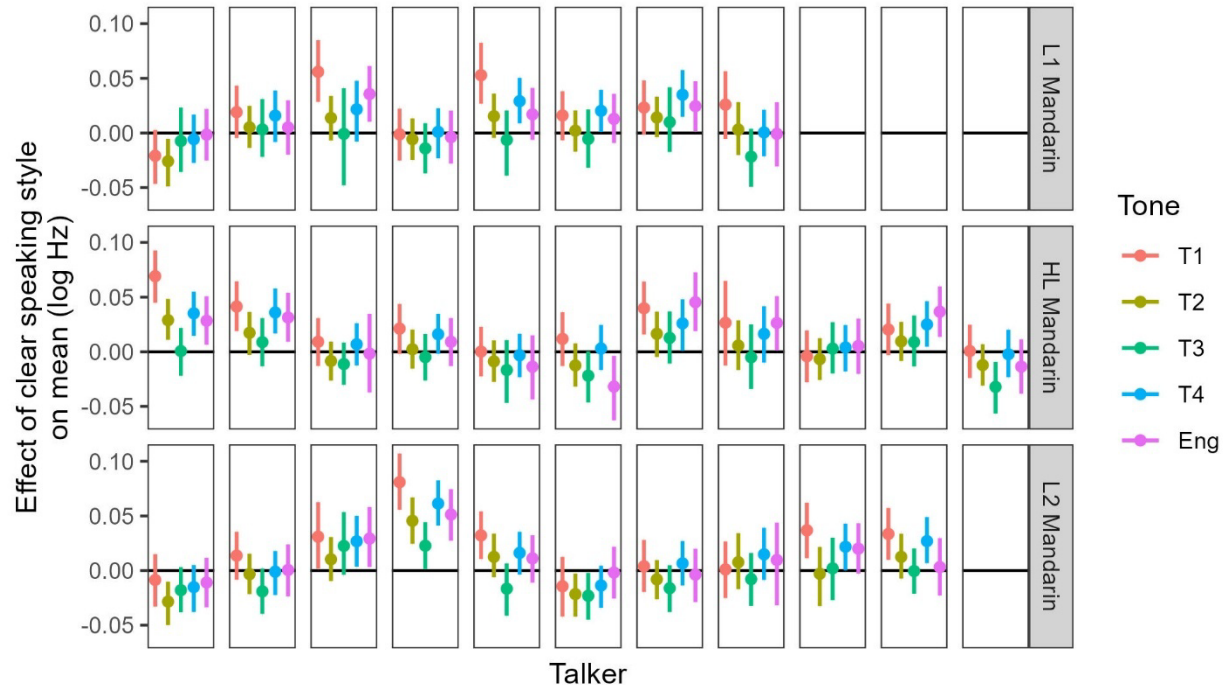


Figure 3.3. 95% credible intervals for the effect of clear speaking style on the polynomial estimate of f0 mean for each tone. Each box represents an individual talker, with L1 Mandarin speakers in the top row, HL Mandarin speakers in the middle row, and L2 Mandarin speakers in the bottom row.

3.3.2 f0 slope

The model for f0 slope resulted in an intercept estimate of -0.11 log Hz (s.d. = 0.01, 95% C.I. = [-0.13, -0.09]), with an estimated magnitude of residual error of 0.05 log Hz (s.d. = 0, 95%

C.I. = [0.04, 0.05]). Figure 3.4 shows group means for the polynomial estimates of f0 slope for each combination of Style, Tone, and Language Group. We can see that Tone 1, a level tone, has a slope close to zero (although this slope is slightly more positive for L1 Mandarin speakers: mean = 0.05, s.d. = 0.01, 95% C.I. = [0.03, 0.08]). Tone 2, a rising tone, has a positive slope (even more positive for L1 Mandarin speakers: mean = 0.07, s.d. = 0.03, 95% C.I. = [0.02, 0.12]). Tone 4 and English both have negative slopes, while the slope of Tone 3 is more negative for L1 Mandarin speakers (mean = -0.11, s.d. = 0.03, 95% C.I. = [-0.18, -0.05]) and close to zero for HL and L2 Mandarin speakers. Tone 3 and English both have negative slopes, while the slope of Tone 4 is more negative for L1 Mandarin speakers (mean = -0.11, s.d. = 0.03, 95% C.I. = [-0.18, -0.05]) and close to zero for HL and L2 Mandarin speakers.

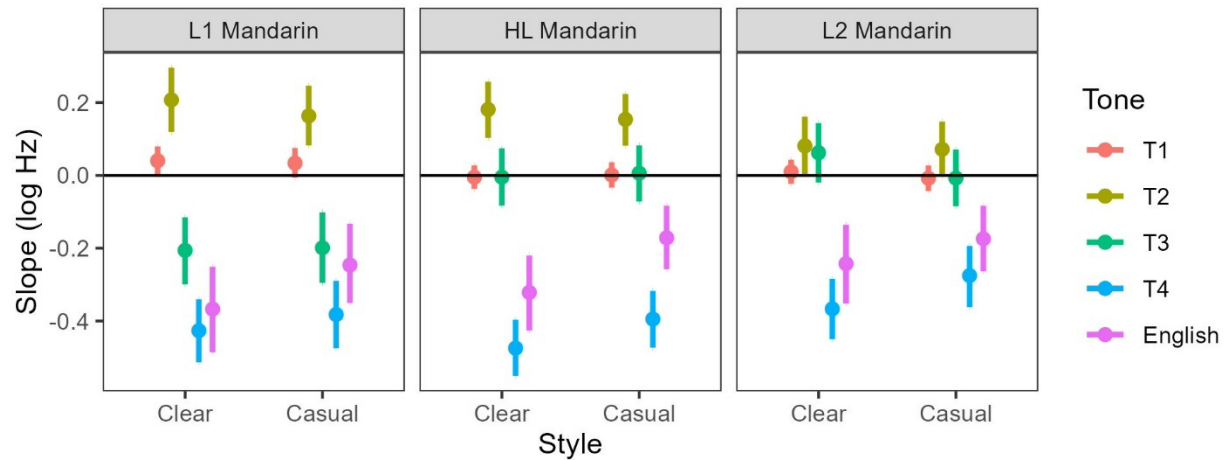


Figure 3.4. 95% credible intervals for group means for polynomial estimates of f0 slope in clear and casual speech, for L1 Mandarin-L2 English, HL Mandarin-L1 English, and L2 Mandarin-L1 English speakers.

Figure 3.5 shows the effect of a clear speaking style on f0 slope for each talker group and tone. For L1 Mandarin speakers, T2 shows a small increase in f0 slope in clear speech. For L2 Mandarin speakers, clear T3 shows a small increase in f0 slope. As for T4, HL and L2 Mandarin speakers enhance T4's negative slope. Both L1 Mandarin and HL Mandarin speakers enhance English's negative slope.

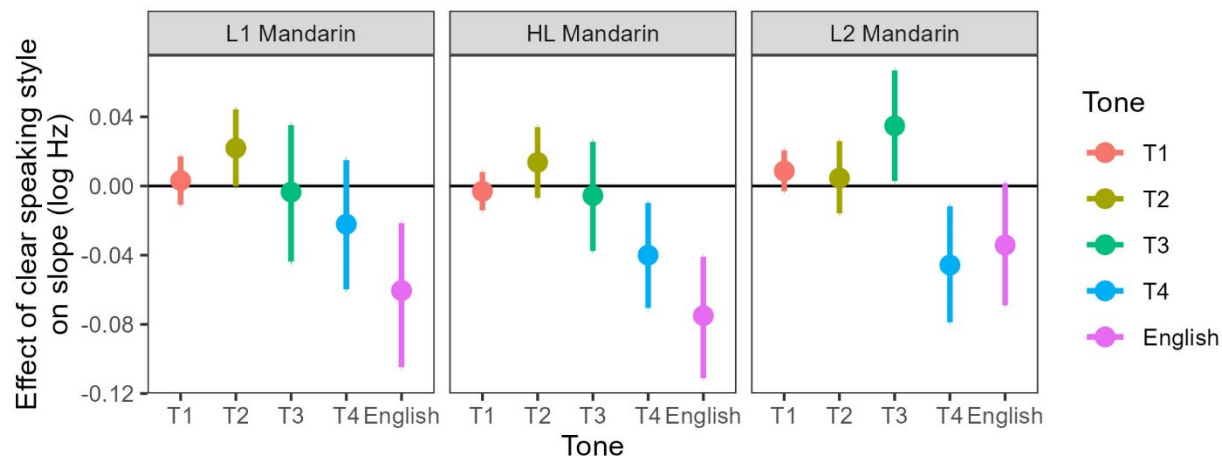


Figure 3.5. 95% credible intervals for the effect of clear speaking style on the polynomial estimate of slope, for each tone and for each language group.

Talkers showed some variation in slope, with systematic between-talker variation of 0.05 log Hz (s.d. = 0.01, 95% C.I. = [0.04, 0.06]). Talkers varied most widely in the slopes of Tone 2 (mean = 0.09, s.d. = 0.01, 95% C.I. = [0.07, 0.12]), Tone 3 (mean = 0.11, s.d. = 0.02, 95% C.I. = [0.08, 0.15]), and Tone 4 (mean = 0.10, s.d. = 0.01, 95% C.I. = [0.07, 0.13]). Figure 3.6 shows simple effects for clear speech for each individual talker. No talker changed the slope of level T1 in clear speech. A few talkers enhanced T2's positive slope. A few HL and L2 talkers increased the slope of T3 in clear speech. Several talkers enhanced T4's negative slope, and many talkers enhanced English's negative slope. A few L2 talkers do not change slope for Mandarin productions, but do for English productions. However, we also see some L2 talkers that enhance the slopes of all productions except for the level tone T1.

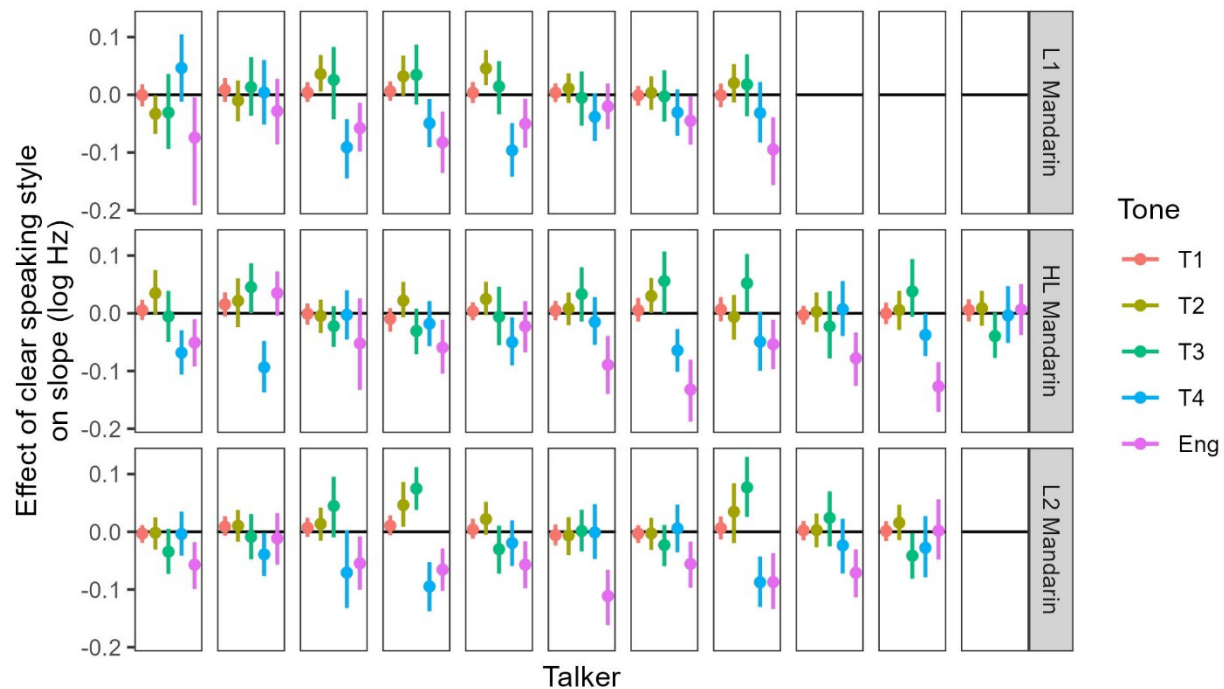


Figure 3.6. 95% credible intervals for the effect of clear speaking style on the polynomial estimate of f_0 slope for each tone. Each box represents an individual talker, with L1 Mandarin speakers in the top row, HL Mandarin speakers in the middle row, and L2 Mandarin speakers in the bottom row.

3.3.3 f_0 curve

The model for the polynomial estimate of f_0 curve resulted in an intercept estimate of 0.21 log Hz (s.d. = 0.03, 95% C.I. = [0.16, 0.27]), with a magnitude of residual error of 0.10 log Hz (s.d. = 0, 95% C.I. = [0.09, 0.11]). Figure 3.7 shows group means of f_0 slope for each combination of Speaking Style, Tone, and Language Group. Overall, Tone 1 and English had a curve close to zero (T1 effect mean = -0.18, s.d. = 0.02, 95% C.I. = [-0.22, -0.14]; English effect mean = -0.22, s.d. = 0.04, 95% C.I. = [-0.3, -0.14]). Tone 4 had a negative curve (T4 effect mean = -0.64, s.d. = 0.04, 95% C.I. = [-0.73, -0.56]), and Tone 2 and Tone 3 had a positive curve (T2

effect mean = 0.38, s.d. = 0.03, 95% C.I. = [0.32, 0.44]; T3 effect mean = 0.67, s.d. = 0.05, 95% C.I. = [0.56, 0.78]). L1 Mandarin speakers produced curvier Tone 3 and Tone 4 than the other speaker groups (T3 mean = 0.67, s.d. = 0.05, 95% C.I. = [0.56, 0.78]; T4 mean = -0.64, s.d. = 0.04, 95% C.I. = [-0.73, -0.56]).

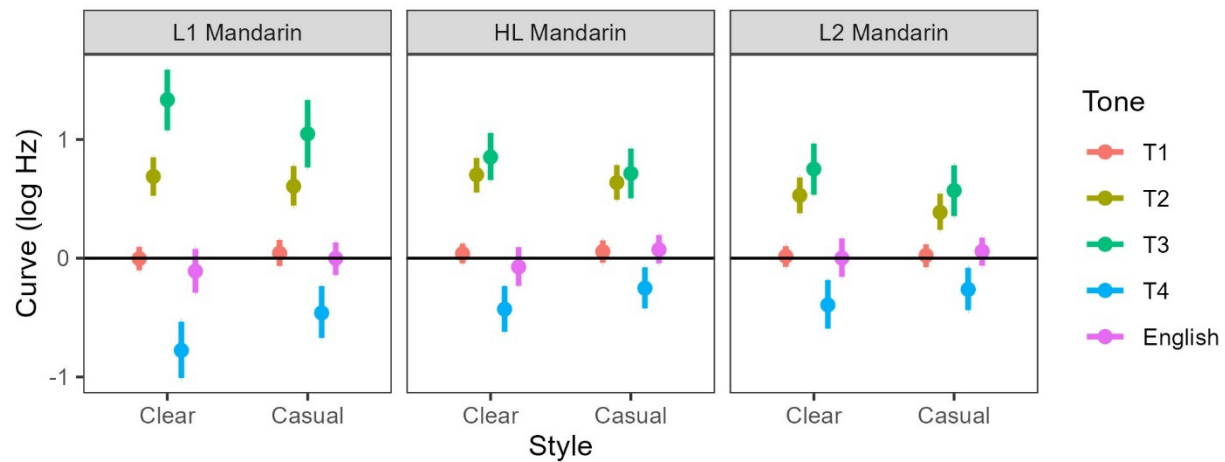


Figure 3.7. 95% credible intervals for group means of the polynomial estimates of f_0 curve, for each tone and for each speaker group for both clear and casual speaking styles.

There was no main effect of Style on f_0 curve (mean = 0, s.d. = 0.01, 95% C.I. = [-0.02, 0.01]); however, there were interactions for different tones. Tones with high positive curve (T2 and T3) were enhanced in clear speech (T2 mean = 0.05, s.d. = 0.01, 95% C.I. = [0.02, 0.08]; T3 mean = 0.10, s.d. = 0.02, 95% C.I. = [0.07, 0.15]). Tone 4, which has a high negative curve, was also enhanced in clear speech (mean = -0.10, s.d. = 0.02, 95% C.I. = [-0.14, -0.06]). Figure 3.8 shows the simple effect of clear speech on f_0 curve for each tone and speaker group. For all speaker groups, T3's positive curve was enhanced. For L1 and HL Mandarin speakers, T4's negative curve was enhanced. L2 speakers also enhanced T2's positive curve, and HL Mandarin speakers produced a more negatively curvy English contour in clear speech.

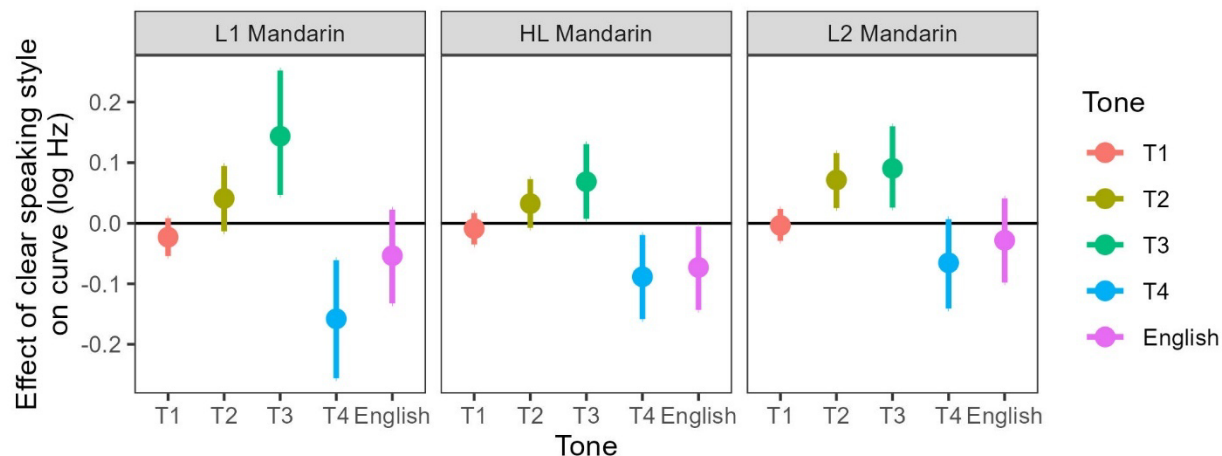


Figure 3.8. 95% credible intervals for the simple effect of a clear speaking style on the f0 curve of each tone, for each speaker group.

There was systematic between-talker variation in f0 curve, with an estimated magnitude of 0.14 log Hz (s.d. = 0.02, 95% C.I. = [0.11, 0.19]). The largest source of variation was in the curve of each tone, with variation in Tone 1 (mean = 0.1, s.d. = 0.02, 95% C.I. = [0.07, 0.14]), Tone 2 (mean = 0.16, s.d. = 0.03, 95% C.I. = [0.12, 0.22]), Tone 3 (mean = 0.26, s.d. = 0.04, 95% C.I. = [0.18, 0.36]), and Tone 4 (mean = 0.21, s.d. = 0.03, 95% C.I. = [0.16, 0.28]). Figure 3.9 shows credible intervals for the simple effect of clear speech on curve, for each individual talker. No talker produces a curvier T1 in clear speech. Some talkers enhance T2's positive curve. Several talkers enhance T3's positive curve. Several talkers also enhance T4's negative curve. A few talkers produce a more negatively curved f0 contour in English clear speech. Whenever there is an effect of clear speech on f0 curve for any of the Mandarin tones, the direction of the effect is to increase the curviness of an already curved contour (tones T2, T3, and T4).

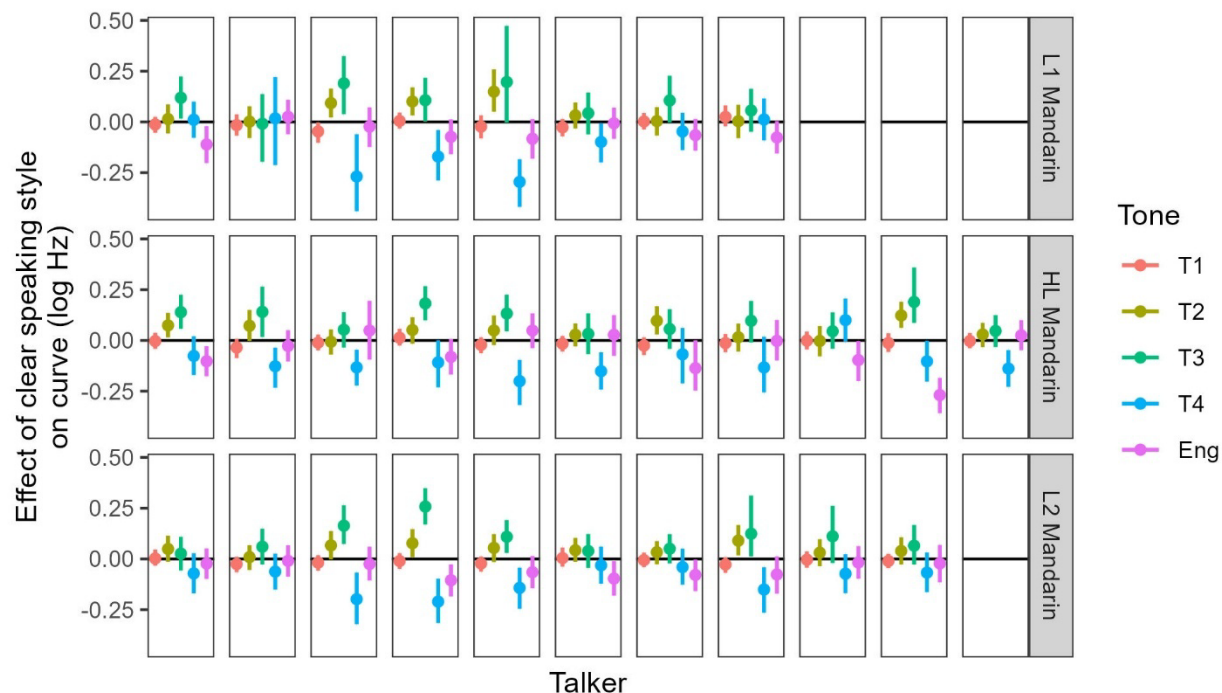


Figure 3.9. 95% credible intervals for the effect of clear speaking style on the polynomial estimate of f_0 curve for each tone. Each box represents an individual talker, with L1 Mandarin speakers in the top row, HL Mandarin speakers in the middle row, and L2 Mandarin speakers in the bottom row.

Figure 3.10 shows polynomials constructed from the models' group means of f_0 mean, f_0 slope, and f_0 curve, using Equation 3.1. This plot must be interpreted cautiously, as these polynomials are calculated solely from point estimates, and the plot does not take into account the fact that the estimated values of mean, slope, and curve could believably be any value within the calculated credible intervals. Furthermore, the three variables of mean, slope, and curve are calculated from separate models. However, the plot does give a visualization of what it means for (for example) Tone 1 to have a higher mean f_0 in clear speech for L1 and L2 Mandarin talkers. We can see the enhancement of HL and L2 Tone 4 slope, the enhancement of L1 Mandarin and HL Mandarin English slope, and the increase in slope for HL Tone 3. We can also

see the enhancement of Tone 3's curve for all talker groups, as well as the enhancement of Tone 4's curve for L1 and HL Mandarin speakers. We can also see that HL Mandarin speakers' English productions become more negatively curvier in clear speech, while L2 Mandarin's Tone 2 productions also become curvier in clear speech.

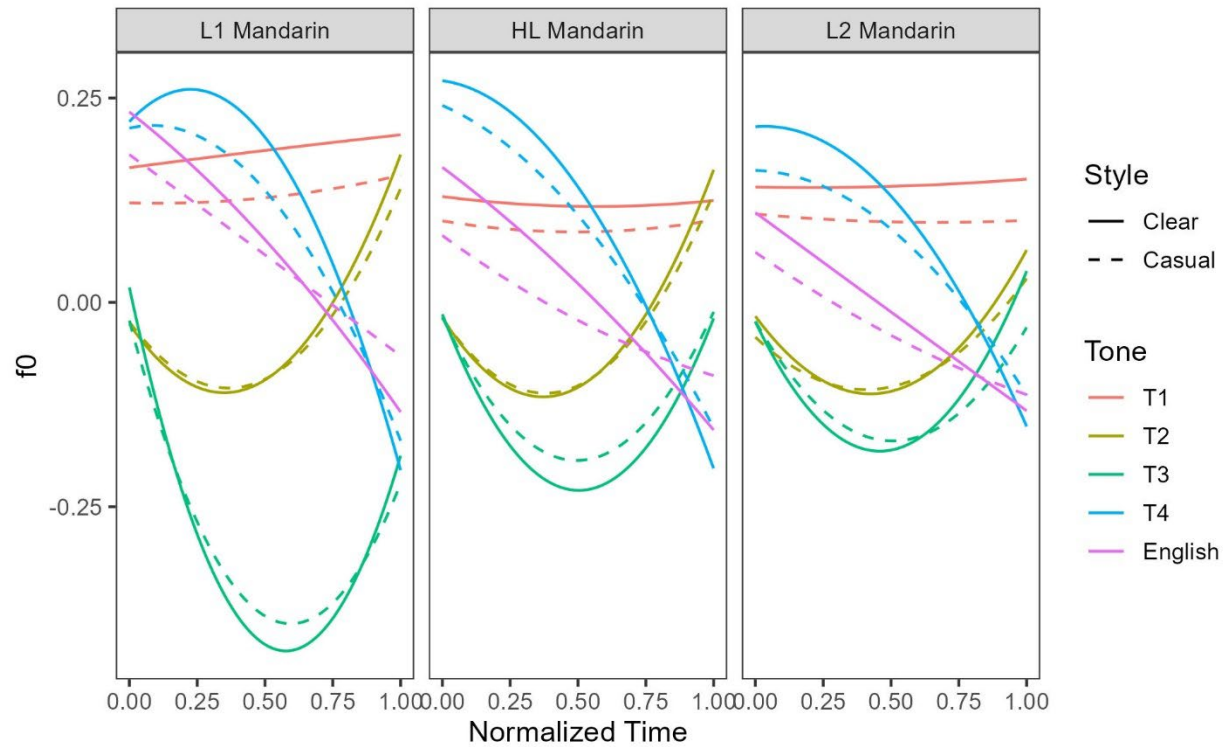


Figure 3.10. *Polynomials calculated from the models, representing the tone contours for each tone and language group, in clear and casual speech.*

Figure 3.11 shows polynomials calculated from point estimates of mean, slope, and curve for each individual talker. From this plot, we can visually see the extent of differences in the tone contour production produced by different talkers, as well as in the clear speech strategies used by each talker.

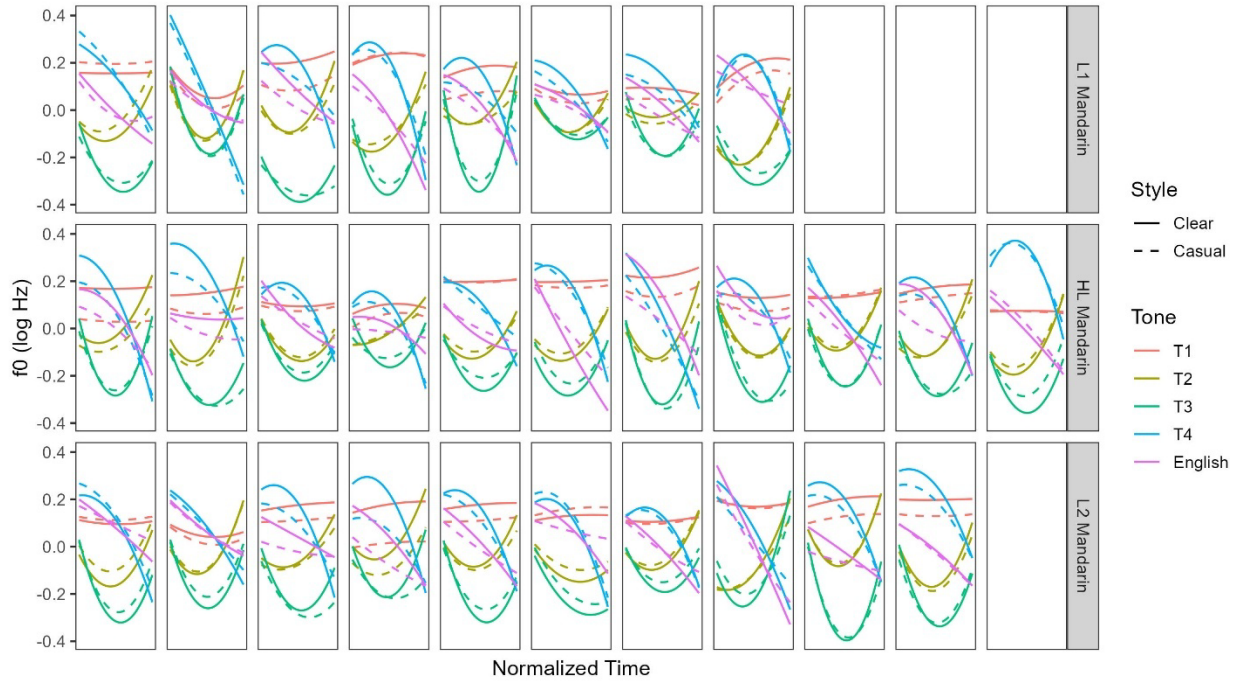


Figure 3.11. Polynomial estimates of the tone contours of each tone in clear and casual speech, for each individual talker. L1 Mandarin talkers are presented in the first row, HL Mandarin talkers in the middle row, and L2 Mandarin talkers in the bottom row.

3.4 Discussion

Overall, the results of the present study indicate that L1, HL, and L2 Mandarin speakers all enhance aspects of f_0 contours in clear speech, albeit in different ways and for different tones. A summary of the results is shown in Table 3.1. L1 Mandarin speakers tend to enhance the high f_0 mean of Tone 1 and Tone 4, enhance the negative slope of English, and enhance the curves of Tone 3 and Tone 4. HL Mandarin speakers tend to enhance the negative slope of Tone 4 and English, as well as enhancing the curves of Tone 3 and Tone 4. HL Mandarin speakers also produce a curvier English f_0 contour in clear speech. L2 Mandarin speakers tend to enhance the

high f0 mean of Tone 1, enhance the slopes of Tone 2 and Tone 3, and produce curvier Tone 2 and Tone 3 contours.

Table 3.1: Overview of the direction of clear speech effects for each talker group, language, and acoustic cue. A plus sign + indicates a positive effect in clear speech and a minus sign - indicates a negative effect in clear speech. Blank cells indicate no effect.

Cue	L1 Mandarin, L2 English					HL Mandarin, L1 English					L2 Mandarin, L1 English				
	T1	T2	T3	T4	En	T1	T2	T3	T4	En	T1	T2	T3	T4	En
F0 mean	+			+	+	+		-	+	+	+			+	+
F0 slope		+			-		+		-	-			+	-	-
F0 curve		+	+	-	-			+	-	-		+	+	-	-

In the present study, we found f0 mean enhancement of Tone 1 and Tone 4 by L1 and L2 speakers, and Tones 1, 3, and 4 by HL speakers. Some individual talkers also show a lowering effect for Tone 2 and Tone 3. The direction of these effects is consistent with tone enhancement, as the tones with high mean (Tone 1 and Tone 4; see Figure 3.1) are becoming higher, and a few talkers show tones with low mean (Tone 2 and Tone 3) becoming lower. These findings are counter to previous studies that have found Tone 2 and/or Tone 3 raising (Tupper et al., 2021; Zhang et al., 2024). These differences may result from differences in elicitation methodologies or stimuli structure; Tupper et al. (2021) used a simulated misunderstanding task to elicit clear speech, while Zhang et al. (2024) used bisyllabic words. Alternatively, these differences may

result from differences in the participants in these studies. The present study found a great deal of variation in the strategies used by participants with respect to mean in clear speech; some individual talkers did indeed raise Tone 2 and Tone 3 in clear speech. However, whenever this raising effect occurred for low tones, Tone 1 and Tone 4 were still raised to a greater degree than Tone 2 and Tone 3, showing tone-specific patterns in the enhancement of f_0 mean in clear speech.

The present study's finding of enhancement of Tone 4's slope only in HL and L2 speech contradicts with previous findings that L1 Mandarin speakers enhance Tone 4's slope in clear speech (Tupper et al., 2021; Zhang et al., 2024). A few individual L1 talkers in the present study did also enhance Tone 4 slope, so perhaps this difference in findings could be due to variation in the strategies used by different speakers. A small effect was found for enhancement of Tone 2 slope by L1 speakers, with 3 of 8 individual L1 talkers showing an effect of Tone 2 slope enhancement in clear speech; this is in accordance with mixed findings on Tone 2 slope enhancement from previous studies (Tupper et al., 2021; Zhang et al., 2024). The present study also found that L2 Mandarin speakers increased the slope of Tone 3, perhaps indicating a higher ending point in the falling-rising Tone 3 contour produced by these speakers. It is possible that this indicates a tendency of some speakers to produce the full falling-rising contour of Tone 3 in clear speech, but produce the half-Tone 3 contour in casual speech.

The largest effects of clear speech on f_0 contour were found in f_0 curve, with curve of Tone 3 increasing for all speaker groups. We also see Tone 4 curve increasing for L1 and HL speakers, and Tone 2 curve increasing for L1 and L2 Mandarin speakers. These differences in curve in clear speech are counter to previous works on Mandarin clear speech, which found no difference in f_0 curve between clear and casual speech for any tone (Zhang et al., 2024; Tupper

et al., 2021). Why might this be? Evidence from Mandarin fast speech research suggests that, in fast speech, Mandarin tones have higher pitch and flatter contours, with reduced slope and curve for Tone 2, Tone 3, and Tone 4 (Tang & Li, 2020). The “speak quickly” instruction in the casual speech condition may result in speakers producing fast speech, instead of Tupper et al.’s (2021) “plain speech” or Zhang et al.’s (2024) adult-directed speech. Thus, the effect on curve may come from reduction in the casual condition, instead of enhancement of the clear condition. However, even if this is the case, we are still seeing differences in tone contour between two registers, one of which is more casual and one of which is more clear.

Overall, all speaker groups enhanced some aspect of the tone contour for all four Mandarin tones and the English pitch contour. The size of these enhancements was still relatively small, indicating that the tones, while enhanced, were not hyper-articulated as might be expected in IDS (Zhang et al., 2024). Interestingly, the pattern of HL enhancement of the English tone contour is similar to the pattern for Tone 4, suggesting some influence of heritage Mandarin on production of tone contours in English. Examining Figure 3.11, a few HL speakers produce an English contour in clear speech that is very similar to Mandarin Tone 4’s contour. However, the reduced English contour of these speakers is not as similar to Tone 4. This pattern is reminiscent of a vowel enhancement effect found by Chang (2024), where L1 English-heritage Korean speakers produced the Korean low central vowel /a/ in a more Korean-like way in clear speech, with a central pronunciation, but in casual speech, these speakers produced a reduced vowel more similar to the English /a/, which is further back.

There was a great deal of variation in the strategies used by talkers in all three language groups to enhance tones in Mandarin clear speech. Investigating the correlates of the different strategies is a ripe area for further research. Additionally, the L2 speakers in this study include

both talkers with no prior tone language experience, and talkers who are heritage Cantonese speakers. Just as childhood tone language experience influences the overall acquisition of Mandarin tones (e.g., Hao, 2012; Tsukada et al., 2015), it may be that having childhood tone language experience impacts the ways an L2 learner enhances tones in clear speech. Additionally, the present study's monosyllabic focus limits the generalizability of the findings. Further research is needed to understand the extent of tone contour enhancement in L2 and HL Mandarin in longer and more natural contexts. Future research can also examine the perceptual benefit of Mandarin clear speech. In Chapter 4, we investigate the impact of a clear speaking style on intelligibility of Mandarin tones produced by L1, HL, and L2 talkers.

In conclusion, all speaker groups enhanced the tonal contours of all four Mandarin tones and English in some way. Whenever there was a change in tonal contour between clear and casual speech, the effect was to enhance a characteristic aspect of the contour in clear speech (higher tones becoming higher, etc). The exception to this is L2 Mandarin T3 having a positive slope in clear speech. Effect sizes were relatively small, indicating the preservation of the tonal contour across speaking styles, so that tones were not over-articulated in clear speech. Speakers in all language groups varied in the extent to which they enhanced tonal contours, with some showing strong patterns of enhancement, and others making little or no difference in aspects of the tonal contours between clear and casual productions in Mandarin.

Chapter 4. Perception of L1, HL, and L2 Mandarin speakers' clear speech productions

4.1 Introduction

Clear speech is a speaking style used when talkers try to produce speech that is more intelligible to a listener (Bradlow, 2002). Clear speech produced by L1 talkers consistently results in a perceptual benefit for L1 listeners (Picheny et al., 1985; Bradlow & Bent, 2002; Smiljanić & Bradlow, 2005), although different talkers vary in the extent to which their clear speech benefits intelligibility (Ferguson, 2012).

The picture is more complex for clear speech produced by L2 talkers. Clear speech produced by L2 talkers often results in an intelligibility benefit for listeners (Jung & Dmitrieva 2023; Kato & Baese-Berk, 2022; Smiljanić & Bradlow, 2007; 2011), but some studies have found a smaller intelligibility benefit or even no intelligibility benefit for L2 clear speech, especially clear speech produced by lower proficiency talkers (Rogers et al., 2010; Kato & Baese-Berk, 2023; 2022). Kato and Baese-Berk (2022) found that high-proficiency L2 speakers showed a similar amount of clear speech intelligibility benefit as L1 talkers (although L1 stimuli was embedded in noise with a larger SNR than the L2 stimuli), while lower-proficiency L2 speakers did not show any clear speech intelligibility benefit. Additionally, the presence of acoustic modifications between clear and casual speaking styles in L2 speech does not necessarily lead to intelligibility benefits for listeners (Kato & Baese-Berk, 2023).

No study that the author is aware of has investigated the intelligibility of clear speech produced by HL talkers. In terms of baseline intelligibility of heritage speech, Daland and Oh

(2014) found that L1 English-heritage Korean speakers were more intelligible in Korean than monolingual Korean speakers. Chang and Yao (2016) found that tones produced by heritage Mandarin speakers were less intelligible than L1 speakers, but more intelligible than L2 speakers.

Mandarin Chinese is a tonal language with four lexical tones (see Table 1.1), and L1 Mandarin listeners are highly accurate at perceiving and identifying these tones (e.g., Wang & Xu, 2020). When speech is embedded in high levels of noise, tone identification accuracy is reduced for L1 listeners (Lee et al., 2013; Wang & Xu, 2020). When perception is made more difficult by embedding noise or altering tonal context, accuracy of Tone 3 perception is the lowest (Hao, 2012), and listeners commonly confuse Tone 2 and Tone 3 productions. Some studies have found that listeners often identify Tone 2 productions as Tone 3 (Lee et al., 2013; Lee et al., 2008; Gottfried & Suiter, 1997; Shen & Lin, 1991), while others have found that listeners identify Tone 3 productions as Tone 2 productions (Chuang et al., 1972; Shen & Lin, 1991). When it comes to L1 listeners perceiving L2 and HL Mandarin tones, the tones with lowest identification accuracy have likewise been Tone 3 (Chang & Yao, 2016; Hao, 2012) and Tone 2 (Hao, 2012).

A limited number of studies have investigated the intelligibility of Mandarin clear speech, mostly studies investigating the use of visual cues in Mandarin tone perception. Zeng et al. (2023) found that L1 Mandarin clear speech benefited intelligibility for L1 Mandarin listeners in an audio-only condition, with increases in tone identification accuracy for Tone 1 and Tone 4, but reduced tone identification accuracy for Tone 2. A few studies have also found that L1 Mandarin clear speech (a “teaching style”) increases tone identification accuracy for listeners with no tone language experience (Han et al., 2018; Zeng et al., 2023), although Han et al.

(2019) found a clear speech benefit only for specific talkers. The existence of a clear speech benefit for speech produced by HL and L2 Mandarin talkers has not yet been investigated.

The present study investigates the existence of a perceptual benefit for clear tones produced by L1 Mandarin-L2 English, HL Mandarin-L1 English, and L2 Mandarin-L1 English speakers, as perceived by L1 Mandarin listeners. We expect to see that clear speech is more intelligible for all speakers, though there may be differences between L1 and HL or L2 talkers (Rogers et al., 2010). We might also see some tones enhanced more than others: Tone 1 and Tone 4 were found to be enhanced in a previous study, while Tone 2 had reduced intelligibility in clear speech (Zeng et al., 2023). Likewise, Tone 3 may also show a unique pattern in terms of intelligibility of clear speech given that this tone is the most difficult to perceive (Hao, 2012; Chuang et al., 1972) and is perceptually confusable with Tone 2 (Blicher et al., 1990; Chuang et al., 1972).

4.2 Methods

4.2.1 Talkers

Recordings from 3 L1 Mandarin, 3 HL Mandarin, and 3 L2 Mandarin talkers were selected from the talkers recorded in the previous experiments in Chapters 2 and 3. Talkers were selected whose clear speech enhancements were roughly representative of the overall clear speech patterns in f0 mean, f0 slope, and f0 curve for their respective language groups. During selection of talkers, we did not consider any talkers for whom more than 3 tokens were excluded from analysis in the production studies (for presence of background noise or an obvious tone mispronunciation), as well as one additional L2 talker who consistently pronounced /ɾ/ as /e/.

All 9 talkers were undergraduate or graduate students at an American university. Most talkers were recruited through a joint psychology and linguistics subject pool where participants received course credit for participation in research. 2 L2 and 1 HL talkers were recruited through flyers shared by Chinese instructors or through word of mouth; these participants received a \$7.50 gift card for their participation.

The three L1 Mandarin talkers were all born and lived in China until adulthood (all female, ages 18-19, mean 18.7). Mandarin Chinese is their most dominant language and their most dominant Chinese dialect. Two speakers spoke other Chinese dialects at home in addition to Mandarin when they were children (Wu and Gan). These talkers are all second language speakers of English who began acquiring English in school (ages 5-7, mean 6). At the time of the study, they attended an American university and had lived in an English speaking country for 5-6 months.

Of the HL Mandarin talkers, one was nonbinary and two were female (ages 18-23, mean 20.3 years). These speakers were born and lived in the United States until adulthood, and English is their current most dominant language. They spoke Mandarin at home with their family members when they were children, reporting starting to learn Mandarin from the age of 0 or 1. Two reported speaking Mandarin and Taiwanese Mandarin at home with their families, and one reported speaking only Taiwanese Mandarin with their family. Table 4.1 shows a summary of the HL talkers' exposure to Mandarin Chinese. They studied Mandarin for an average of 4.6 years, including an average of 2.3 years in elementary or middle school, 2 years in high school, and 1.3 years in college. One participant also participated in 8 years of a weekend or after school program. On a scale from 1 to 10 where 10 is most proficient, the participants reported themselves as having a mean speaking proficiency of 5 (range 3-6), listening proficiency of 6

(range 4-8), and reading proficiency of 2 (range 1-3). They reported perceiving themselves as having a “moderate” (5 out of 10; two participants) or “considerable” (6 out of 10; one participant) foreign accent in Mandarin. Two of the HL participants were also L2 speakers of ASL, Japanese, and/or Spanish.

Table 4.1. *Heritage talkers’ exposure to Mandarin Chinese*

ID	Age of onset of acquisition of Mandarin	Chinese dialects spoken at home	Years weekend or after school program	Years elementary or middle school classes	Years high school classes	Years college courses	Top factor contributing to Mandarin learning
E02	0	Taiwanese Mandarin	0	1	0	1.7	Reading
E18	0	Mandarin, Taiwanese Mandarin	0	3	2	0	Interacting with friends
E29	1	Mandarin, Taiwanese Mandarin	8	3	4	1.3	Interacting with family

Of the L2 Mandarin talkers, one was female and two were male (ages 21-26, mean 23.3). These speakers were born and lived in the United States until adulthood, and English is their current most dominant language. None of the talkers spoke any tonal language at home. Table 4.2 shows a summary of the L2 talkers’ exposure to Mandarin Chinese. The talkers started acquiring Mandarin from the ages of 16-17 (mean 16.7). One participant took 9 months of Mandarin in high school. All talkers studied Mandarin in college, with an average of 1.6 years of instruction. Two participants also studied abroad in China, one for four months and one for 1.5 years. On a scale from 1 to 10 where 10 is most proficient, the talkers rated themselves as having a mean speaking proficiency of 6 (range 5-7), listening proficiency of 5 (range 3-8) and reading

proficiency of 5.3 (range 4-7). The participants rated themselves as having “some” (4 out of 10; two participants) or “moderate” (5 out of 10; 1 participant) foreign accent in Mandarin. The talkers were also L2 speakers of Cantonese, Dutch, and/or Spanish.

Table 4.2. *L2 talkers’ exposure to Mandarin Chinese*

ID	Age of onset of acquisition of Mandarin	Chinese dialects spoken at home	Years weekend or after school program	Years high school classes	Years college courses	Years study abroad	Top factor contributing to Mandarin learning
E07	17	None	0	0.8	2.3	0	Reading
E08	16	None	0	0	0.5	0.3	Listening to radio/music
E30	17	None	0	0	2	1.5	Interacting with friends

4.2.2 Listeners

38 L1 Mandarin listeners (ages 18-24, mean 20 years; 1 nonbinary, 23 female, 14 male) participated in the experiment. Listeners were undergraduates at an American university recruited from a linguistics and psychology study pool and received course credit for their participation in the study. Listeners reported that Mandarin (n=37) or Taiwanese Mandarin (n=1) was their most dominant dialect of Chinese and their most dominant language in general. They began learning Mandarin between the ages of 0-2 (mean 0.1 years old) and lived in a Chinese-speaking country until adolescence. Apart from Standard Mandarin, 20 of the participants reported also knowing Wu, Sichuanese, Cantonese, Hakka, Gan Chinese, Taiwanese Mandarin, or Yunnan, Shandong, Shanxi, Dongbei, Nanchang, Henan, Kunming, or Wenzhou dialects. For

most participants, English was their second-most dominant language, but participants also reported knowledge of Japanese (n=11), Korean (n=3), Spanish (n=3), French (n=2), German (n=1), and Uyghur (n=1).

An additional 14 listeners completed the study, but were excluded as Mandarin Chinese was not their dominant language (n=10). One listener was also excluded for hearing loss, and three listeners were excluded for failing to pass a comprehension check question at the end of the experiment.

4.2.3 Materials

12 Mandarin syllables were used in this study: the vowels /i, ʌ, u/ crossed with the Mandarin tones T1, T2, T3, and T4 (Tupper et al., 2020; 2021). 3 English monosyllabic words were also used: the vowels /i, ʌ, u/ with a /hVd/ syllable structure (“heed,” “hud,” and “who’d”). Each syllable was produced twice in clear speech and twice in casual speech by each talker, for a total of 48 Mandarin and 12 English productions. For clear speech, talkers were instructed to speak clearly as if they were talking to someone who was having a hard time understanding them (Zellou et al, 2022). For the clear speech condition, talkers were also told that the recordings would be played to language learners to help them learn. For casual speech, talkers were instructed to speak as if they were talking to a friend or family member they had known for a long time who had no trouble understanding them, and to speak quickly (Zellou et al, 2022).

Any tokens with background noise (sniffs, coughs, loud rustling, etc) were excluded from use in the study, as well as any tokens where there was an obvious pronunciation of a different tone or vowel than was intended. For L1 talkers, this was determined auditorily by a phonetically trained L1 Mandarin speaker. For HL and L2 talkers, this was determined by graphing the tone

contours for each talker; productions where a particular tone contour clearly and obviously followed that speaker's pattern of production for a different tone were excluded. For five of the talkers, one Mandarin token was excluded, and no tokens were excluded for the remaining four talkers, resulting in a total of 59 or 60 productions by each talker.

Each token was amplitude-normalized to 65 dB. 500 ms of silence was added before and after each word, and then files were mixed with speech-shaped noise (SSN) with -9 dB SNR (Zeng et al., 2023; Bradlow & Alexander, 2007; Wang & Xu, 2020). The SNR of -9 dB was chosen based on Wang and Xu's (2020) investigation of L1 Mandarin tone recognition under different noise conditions; recognition under -9 dB was found to be about 75% for the L1 talkers investigated in the study. This high (but not ceiling) performance for tone recognition of L1 productions was desirable for this study because HL and/or L2 talkers might be expected to have lower intelligibility than L1 talkers. The SSN was generated from white noise using a filter created from the spectrum of all of the tokens used in the experiment (Wang & Xu, 2020). Finally, 400ms of leading silence was added to the beginning of each token. This resulted in a stimulus format of 400ms of silence, 500ms of SSN, the target word mixed with SSN, and 500ms of SSN (as in Jung & Dmitrieva, 2023).

4.2.4 Methods

Listeners completed the study online on the Qualtrics survey platform (Qualtrics, 2025). In the first section of the experiment, participants completed a Mandarin tone recognition task with a four-alternative forced choice procedure. Listeners heard a Mandarin syllable in noise over headphones and selected the word they heard from four multiple choice options, each with the same vowel as the stimulus and with one of the four Mandarin tones. Each option was

displayed both as a character and with Pinyin with tone marks (Wang & Xu, 2020). Listeners heard tokens from three different talkers, one selected randomly from each of the three talker groups. Presentation of stimuli was blocked by talker and the order of the three talker blocks was randomized. Within each talker block, the order of presentation of syllables was randomized, such that clear and casual tokens both appeared mixed in together. From each of the three talkers, listeners heard 47-48 tokens.

The second section of the experiment was a three-alternative forced choice procedure with the English tokens. Listeners heard an English syllable in noise and selected which of three words (“heed,” “hud,” or “who’d”) that they heard. Each option was presented with a rhyme (for example, “hud (rhymes with mud)”). Again, tokens were blocked by talker but randomized within each talker, and the order of talker blocks was randomized. Listeners heard 12 English tokens from each talker.

Before the experiment began, participants completed informed consent, confirmed that they were using headphones in a quiet environment, and completed a sound calibration procedure to ensure that they could hear at a comfortable volume. Participants completed a practice block before the experiment using syllables in noise recorded by an L1 Mandarin talker. At the end of the experimental trials, listeners completed a comprehension question where they listened to a sentence and indicated what the final word in the sentence was. This question was intended to ensure that listeners were paying attention to the experiment. After the experiment, participants completed the LEAP-Q language background questionnaire (Marian et al, 2007; Kaushanskaya et al, 2020).

4.2.5 Analysis

For the analysis of tone identification accuracy, we ran a Bayesian multilevel logistic regression model, including fixed effects of speaking style (Clear or Casual speech), tone (the four Mandarin tones plus English), and Talker Language Group (Heritage Mandarin speakers, L1 Mandarin speakers, and L2 Mandarin speakers). All possible interactions between these factors were also included. The model also included random effects of the Style and Tone factors for each individual talker and each individual listener, as well as random effects of the interaction between Style and Tone for each listener and talker. The equation for the model is shown in Equation 4.1. Since divergent transitions were found with this model structure, the model was simplified by removing the random effects structure with the least amount of variation, until the model converged. The final model structure is shown in Equation 4.2.

$$\text{Accuracy} \sim \text{Tone} * \text{Style} * \text{TalkerLanguageGroup} + ((\text{Tone} * \text{Style}) | \text{Talker}) + ((\text{Tone} * \text{Style}) | \text{Listener}) \quad (4.1)$$

$$\text{Accuracy} \sim \text{Tone} * \text{Style} * \text{TalkerLanguageGroup} + ((\text{Tone} + \text{Style}) | \text{Talker}) + (\text{Tone} | \text{Listener}) \quad (4.2)$$

We use t-distributions to model the data, including the priors. Since Wang and Xu (2020) found a tone identification accuracy of about 75% for L1 tone productions with SSN at a SNR of -9 dB, we use the logit of 0.75 as the prior for the intercept (1.099). We use t-distributions centered on 0 for all other priors, with a scaling of 3.

4.3 Results

4.3.1 Intelligibility

The model resulted in an intercept accuracy of 2.54 logits (s.d. = 0.25, 95% C.I. = [2.04, 3.04]), which corresponds to an overall probability of accuracy of 92.7% (95% C.I. = [88.5, 95.4]). Figure 4.1 shows group means of accuracy in logits for each combination of Speaking Style, Tone, and Language Group. Overall, accuracy was lower for Tone 3 (mean = -1.65, s.d. = 0.31, 95% C.I. = [-2.22, -1]), and higher for Tone 4 (mean = 1.33, s.d. = 0.44, 95% C.I. = [0.49, 2.23]). Accuracy was overall lower for L2 Mandarin speakers (mean = -0.64, s.d. = 0.32, 95% C.I. = [-1.27, 0.02]).

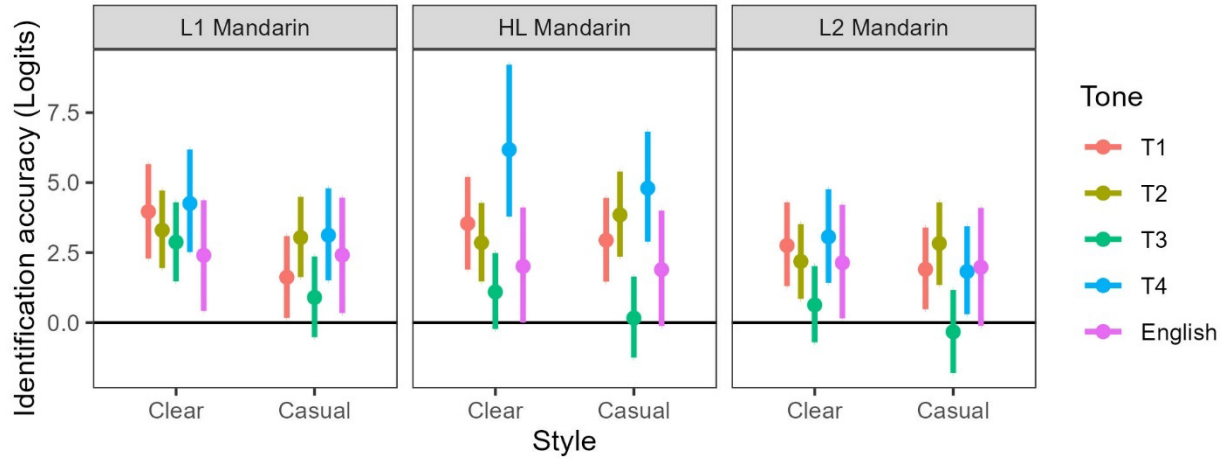


Figure 4.1. 95% credible intervals of group means of accuracy of tone identification, for each Talker Language Group (L1 Mandarin-L2 English, HL Mandarin-L1 English, and L2 Mandarin-L1 English), Tone, and Speaking Style (Casual and Clear).

There was a main effect of clear speech (mean = 0.34, s.d. = 0.12, 95% C.I. = [0.11, 0.59]), an effect which was larger for Tone 1 (mean = 0.29, s.d. = 0.11, 95% C.I. = [0.07, 0.52])

and Tone 3 (mean = 0.3, s.d. = 0.08, 95% C.I. = [0.14, 0.46]) and not present for Tone 2 (mean = -0.57, s.d. = 0.11, 95% C.I. = [-0.79, -0.35]). Figure 4.2 shows credible intervals for the simple effect of a clear speaking style on accuracy for each talker language group and tone. Listeners showed increased accuracy at identification of Tone 1 and Tone 3 produced by L1 Mandarin speakers in clear speech, of Tone 3 produced by HL Mandarin speakers in clear speech, and of Tone 3 and Tone 4 produced by L2 Mandarin speakers in clear speech. There is also a small effect of increased accuracy of identification of clear L1 Tone 4 and clear L2 Tone 1, as well as decreased accuracy of HL and L2 Tone 2 in clear speech.

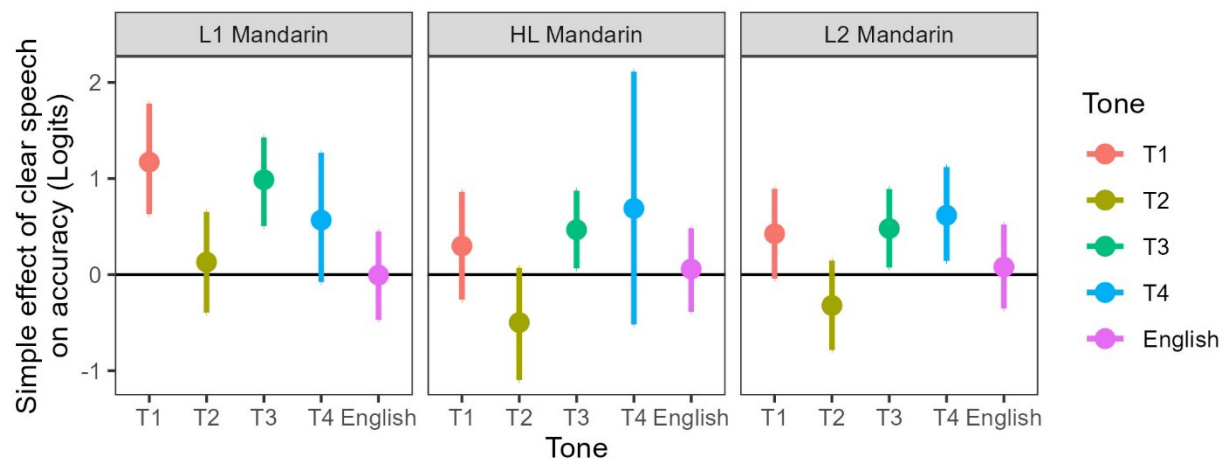


Figure 4.2. 95% credible intervals of the simple effect of a clear speaking style on tone identification accuracy, for each Talker Language Group and Tone.

There was some between-listener variation in accuracy (mean = 0.5, s.d. = 0.08, 95% C.I. = [0.37, 0.67]), with the greatest variation in accuracy of Tone 2 (mean = 0.71, s.d. = 0.15, 95% C.I. = [0.44, 1.02]). There was also between-talker variation in accuracy (mean = 0.63, s.d. = 0.27, 95% C.I. = [0.29, 1.27]), including variation in the effect of speaking style on accuracy (mean = 0.27, s.d. = 0.15, 95% C.I. = [0.07, 0.64]). Figure 4.3 shows credible intervals for the effect of clear speaking style on accuracy, for each individual talker. Listeners show an increase

in accuracy for one L1, one HL, and two L2 talkers when those talkers produce clear speech, as compared to casual speech.

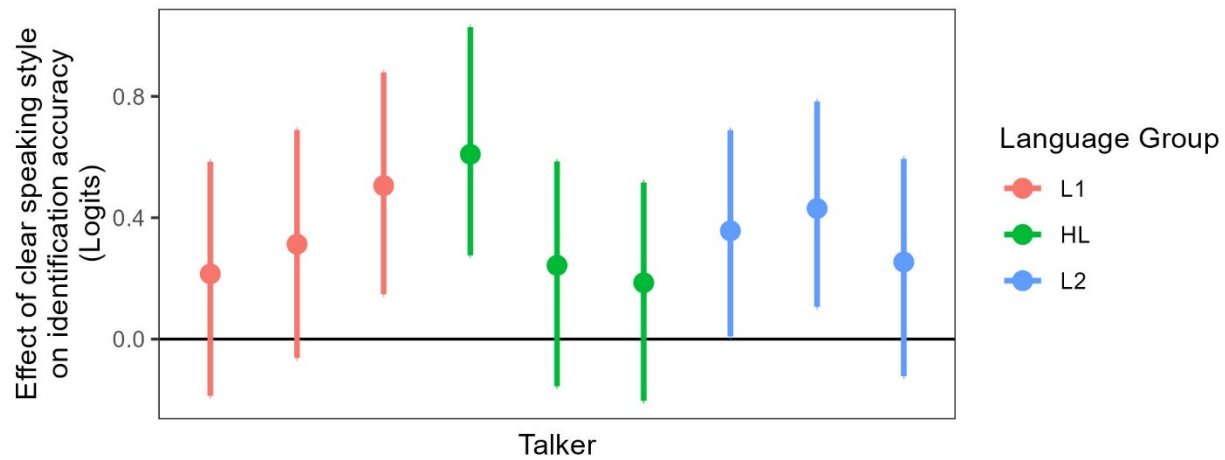


Figure 4.3. 95% credible intervals for the effect of a clear speaking style on identification accuracy, for each individual talker.

4.3.2 Error analysis and accuracy of Tone 3 identification

Figure 4.4 shows a heat map plot of the Mandarin data, indicating how tones were perceived by listeners, broken down by which tones were produced by talkers in each of the three talker groups. Overall, accuracy was high and tones were usually perceived as they were intended. The most prominent error pattern is Tone 3 productions often being perceived as Tone 2.

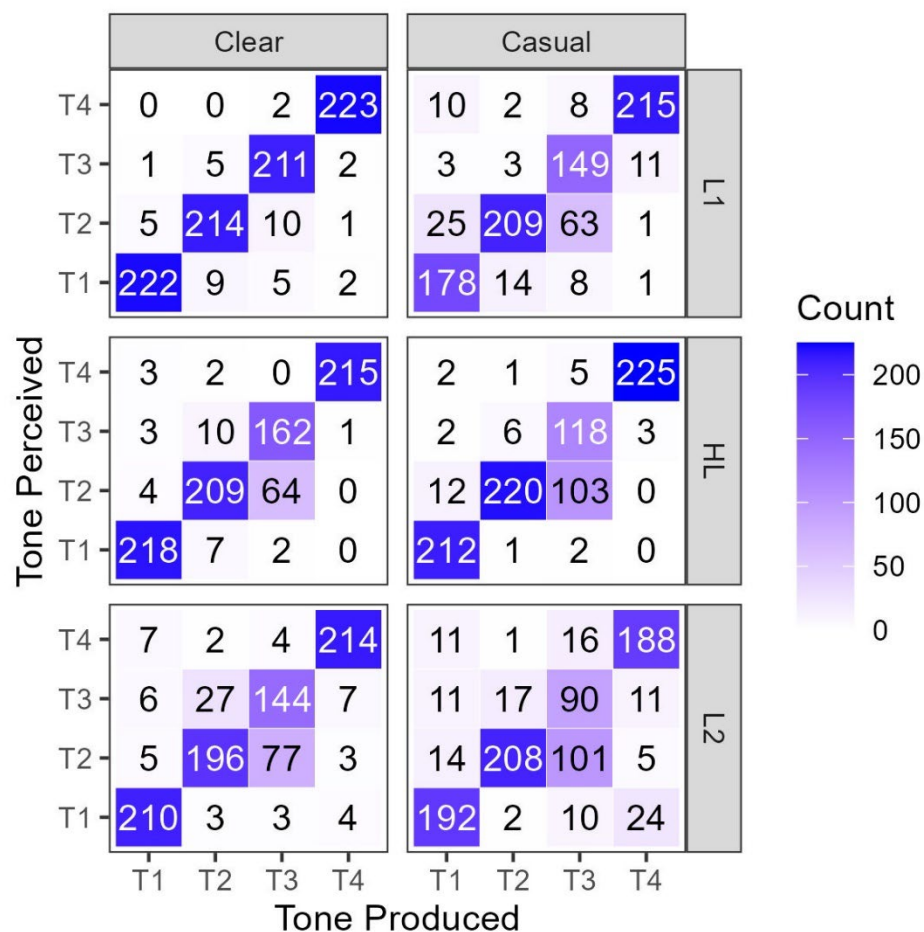


Figure 4.4. Heat map of tone confusions, representing how the talkers' productions were perceived. The left column shows clear speech, while the right shows casual speech. Each row represents the talkers' language background, with L1 Mandarin speakers in the top row, HL Mandarin speakers in the middle, and L2 Mandarin speakers in the bottom row.

To test whether clear speech improved accuracy of Tone 3 identification, we ran a post-hoc Bayesian multilevel logistic regression model on a subset of the data: Tone 3 productions that were perceived as either Tone 2 or Tone 3. Fixed effects of Speaking Style (Clear or Casual) and Language Group (L1 Mandarin, HL Mandarin, or L2 Mandarin) were included, as well as

interactions between these. Random slopes of Talker Language Group and Style were included for each Listener, as well as the slope of Style for each Talker. Equation 4.3 shows the model structure. We use t-distributions to model the data, including the priors. We use t-distributions centered on 0 for priors for the intercept and fixed effects, with a scaling of 3.

$$\text{Accuracy} \sim \text{Style} * \text{TalkerLanguageGroup} + (\text{Style} | \text{Talker}) + ((\text{Style} * \text{TalkerLanguageGroup}) | \text{Listener}) \quad (4.3)$$

The model resulted in an intercept of 1.12 logits (s.d. = 0.53, 95% C.I. = [0.08, 2.19]), which corresponds to a probability of correct Tone 3 identification of 75.4% (95% C.I. = [52.0, 89.9]). Figure 4.5 shows credible intervals for the group means of Tone 3 identification accuracy, for clear speech, casual speech, and for each talker language group. Tone 3 identification accuracy was high for L1 Mandarin clear speech productions.

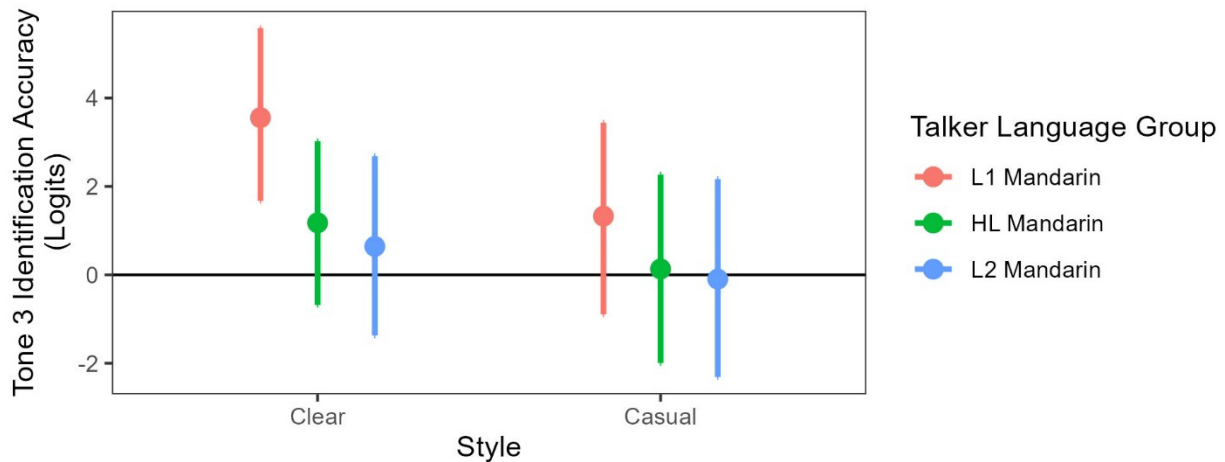


Figure 4.5. 95% credible intervals for group means of Tone 3 identification accuracy, for clear and casual speech and for each talker language group.

Figure 4.6 shows credible intervals for the simple effect of clear speech on Tone 3 identification accuracy, for each talker language group. There was a main effect of speaking style on Tone 3 identification accuracy (mean = 0.67, s.d. = 0.32, 95% C.I. = [0.02, 1.3]). However,

only L1 Mandarin speakers show a clear increase in Tone 3 identification accuracy in clear speech.

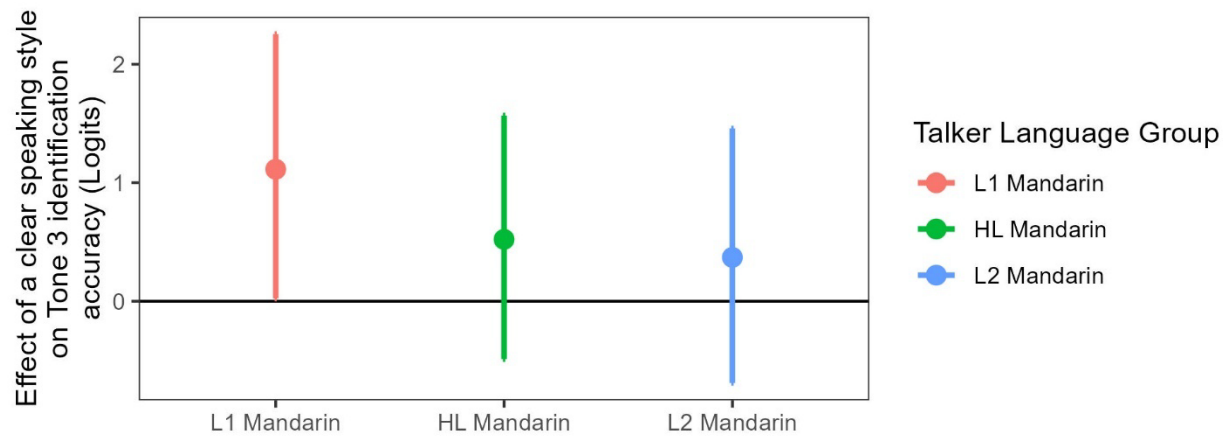


Figure 4.6. 95% credible intervals for the simple effect of a clear speaking style on Tone 3 identification accuracy, for each talker language group.

There was between-talker variation in overall accuracy (mean = 1.45, s.d. = 0.56, 95% C.I. = [0.72, 2.88]), as well as variation between talkers in the size of any clear speech benefit for accuracy (mean = 0.84, s.d. = 0.34, 95% C.I. = [0.39, 1.71]). Figure 4.7 shows the effect of a clear speaking style on identification accuracy for each individual talker. One L1 talker, one HL talker, and two L2 talkers produce Tone 3 in such a way that listeners identify Tone 3 with increased accuracy in clear speech than in casual speech. The four talkers for whom listeners show increased tone 3 identification accuracy are the same four talkers that showed a clear speech benefit to intelligibility in the previous analysis (Section 4.3.1).

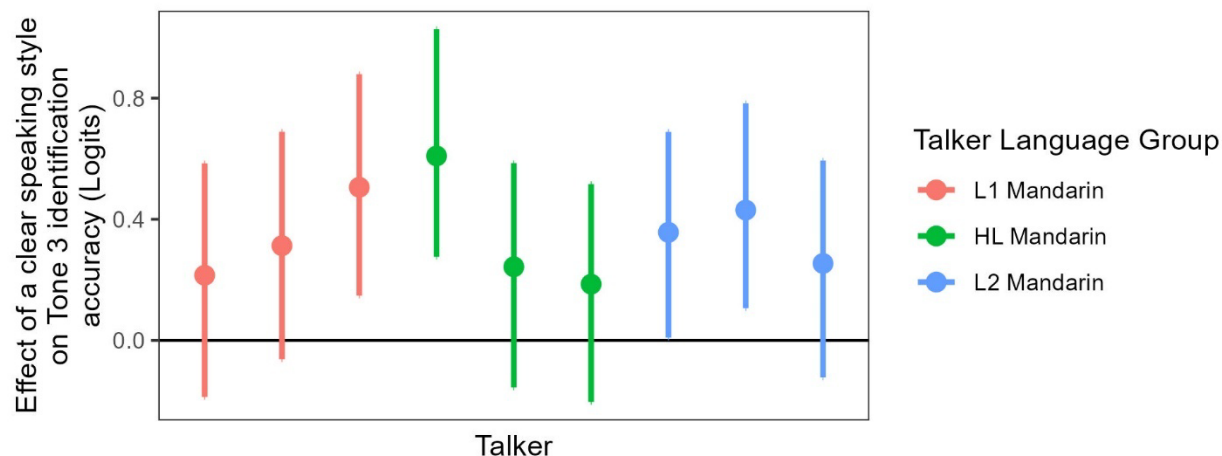


Figure 4.7. 95% credible intervals for the simple effect of a clear speaking style on Tone 3 identification accuracy, for each individual talker.

Compared to the amount of variability between talkers, there was very little between-listener variation in Tone 3 identification accuracy (mean = 0.27, s.d. = 0.13, 95% C.I. = [0.03, 0.53]), including variation in the effect of a clear speaking style on accurate identification of Tone 3 (mean = 0.13, s.d. = 0.09, 95% C.I. = [0.01, 0.35]). Figure 4.8 shows the simple effect of a clear speaking style on Tone 3 identification accuracy for each individual listener, broken down by the language group of each talker. There is little variation in the patterns shown for each listener.

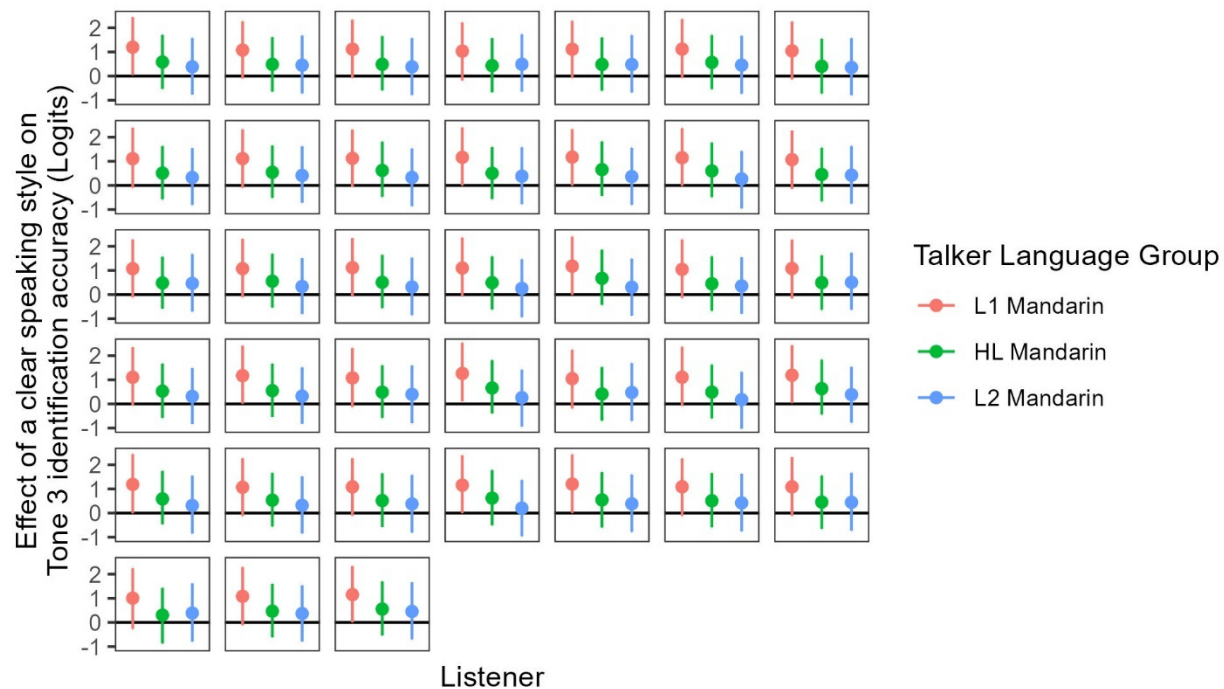


Figure 4.8. 95% credible intervals for the simple effect of a clear speaking style for Tone 3 identification accuracy. Each box represents an individual listener, with estimates for the effect of clear speech for each talker language group.

4.4 Discussion

The results from the present study show that L1, HL, and L2 Mandarin talkers are all able to increase the intelligibility of some tones in clear speech. A summary of the clear speech effects for accuracy are shown in Table 4.3. L1 Mandarin speakers increase the intelligibility of Tone 1, Tone 3, and Tone 4 in clear speech. HL Mandarin speakers increase intelligibility of Tone 3 in clear speech, but show a marginal effect for a reduction in Tone 2 intelligibility in clear speech. L2 Mandarin speakers show an increase in intelligibility of Tone 1, Tone 3 and Tone 4 in clear speech. All effect sizes are small, perhaps due to the overall high accuracy in

tone identification. No clear speech differences were found for the English data for any language group.

Table 4.3: Overview of the directionality of clear speech intelligibility effects for each talker group, language, and acoustic cue. A plus sign + indicates a positive effect in clear speech, a minus sign - indicates a negative effect in clear speech, and blank cells indicate no effect.

L1 Mandarin, L2 English					HL Mandarin, L1 English					L2 Mandarin, L1 English				
T1	T2	T3	T4	En	T1	T2	T3	T4	En	T1	T2	T3	T4	En
+		+	+			-	+			+	-	+	+	

Overall, identification accuracy was very high: highest for Tone 4 and lowest for Tone 3. The low Tone 3 identification accuracy is consistent with previous research on perception of L1, HL, and L2 Mandarin tones (Hao, 2012; Chuang et al., 1972). Yet, listeners received an intelligibility benefit for clear Tone 3 produced by all three talker groups. This may stem from the fact that accuracy was low for Tone 3, so there was more room for variation. However, other tones also showed an intelligibility benefit, with Tone 1 and Tone 4 receiving a clear speech intelligibility benefit for L1 and L2 talkers, just as Zeng et al. (2023) found a benefit for L1 talkers' T1 and T4 productions. The finding that HL and L2 Mandarin talkers' clear T2 actually received a small decrease in intelligibility in clear speech is also in accordance with previous findings that L1 Mandarin talkers' T2 productions have a decrease in intelligibility in clear speech (Zeng et al., 2023). Blicher et al. (1990) found that a longer syllable duration made listeners more likely to hear a Tone 3 over a Tone 2; since clear speech is longer than plain speech, this could be a driving factor in reduced intelligibility of Tone 2 in clear speech.

The present study found that Tone 3 productions were often perceived as Tone 2, but that Tone 2 productions were not as commonly perceived as Tone 3 (in accordance with Chuang et al., 1972 and counter to Lee et al., 2013; Lee et al., 2008; and Gottfried & Suiter, 1997). This error pattern was reduced in L1 clear speech, but not in HL and L2 clear speech. These findings suggest an area of focus for teaching L2 and HL Mandarin speakers to produce clearer tones. These speakers are able to increase the intelligibility of Tone 3 in general, but were not able to decrease the Tone 2-Tone 3 confusion pattern in clear speech to the same extent as L1 speakers; targeting this Tone 2-Tone 3 confusion directly could be useful for learners. Similarly, an important direction for future work is HL and L2 perception of clear speech in Mandarin: are HL and L2 listeners able to benefit from the acoustic modifications of a clear speaking style?

In the present study, there was variability in the extent to which different talkers' clear speech productions resulted in an intelligibility benefit for listeners, similar to previous research (Ferguson, 2012). Future work can investigate which aspects of a talker's speech lead to greater intelligibility benefits in Mandarin. Furthermore, only 3 talkers from each language group were investigated in the present study, and talkers who made acoustic modifications similar to the overall group patterns found in Chapter 3 were selected. Would other talkers who made smaller acoustic modifications show any clear speech benefit? Additionally, the acoustic correlates of a clear speech benefit in Mandarin Chinese is an important question for future work.

One limitation of this study was the high accuracy for tone identification, which is fairly easy for L1 listeners even in noise levels that are higher than those used for speech intelligibility studies on non-tonal language (Wang & Xu, 2020; see for example Smiljanić & Bradlow, 2011). To get a more accurate picture of relative difficulties of identification of different tones, a more

difficult task would be needed, for example transcribing full sentences, or longer strings of semantically anomalous words.

In conclusion, the present study found that L1 Mandarin, HL Mandarin, and L2 Mandarin talkers' clear speech enhancements resulted in intelligibility benefits for L1 Mandarin listeners. These intelligibility benefits were present for some tones but not others: L1 Mandarin talkers' enhancements resulted in an intelligibility benefit for Tone 1, Tone 3, and Tone 4. HL Mandarin talkers' enhancements resulted in an intelligibility benefit for Tone 3, and a reduction in intelligibility for Tone 2. L2 Mandarin talkers' enhancements resulted in an intelligibility benefit for Tone 1, Tone 3 and Tone 4, with a reduction in intelligibility for Tone 1. Talkers varied in the extent to which their clear speaking style benefited intelligibility. The size of the clear speech intelligibility benefit was small, perhaps due to the overall high intelligibility. In Chapter 5, we discuss these results in conjunction with the results from Chapter 2 and Chapter 3.

Chapter 5. Discussion

The present dissertation investigated the acoustic correlates of L1, HL, and L2 Mandarin clear speech, as well as the perceptual benefit associated with a clear speaking style produced by these talkers. We asked whether the acoustic adjustments made in Mandarin clear speech were the same adjustments used for English clear speech, and whether L1, HL, and L2 Mandarin speakers enhanced Mandarin tonal contours in clear speech.

In this chapter, we first lay out a summary of the main findings in Section 5.1. Then, we discuss the findings through various lenses, relating the findings to theoretical positions and questions in prior work in Section 5.2. Finally, we detail limitations of the current study and directions for future work in Section 5.3, and finish with concluding remarks in Section 5.4.

5.1 Summary of Findings

Chapter 2 compared clear speech enhancements in English and Mandarin. All language groups produced clear speech and casual speech as two acoustically distinct registers, in both English and Mandarin. Slower duration and expanded pitch range were the two most prominent strategies for enhancing speech in both English and Mandarin, although HL Mandarin speakers did not expand pitch range in Mandarin clear speech. Small increases in mean intensity were found for clear speech, and some individual talkers slightly increased mean pitch in clear speech, but the size of these enhancements in mean pitch and intensity were often small enough that it would not be possible for a human listener to perceive a difference.

Chapter 3 investigated tone contour enhancements in Mandarin. L1, HL, and L2 Mandarin speakers enhanced the tonal contours of Mandarin tones in tone-specific ways. Tone 1

and Tone 4, as tones with an overall high mean pitch, were the tones that saw an increase in pitch in clear speech. L1 Mandarin speakers produced Tone 2 with a steeper pitch slope, and HL and L2 Mandarin speakers enhanced Tone 4's negative pitch slope. L2 Mandarin speakers produced T3 with a positive slope in clear speech, and all talker groups enhanced English's negative pitch contour. Curve was also enhanced in clear speech, with all groups enhancing the convex curve of T3 and the concave curve of T4. L2 Mandarin speakers also slightly enhanced the convex curve of T2. Different talkers showed variability in the aspects of the tone contour that they enhanced, and in the extent to which they enhanced the tonal contour. Whenever there was an effect of a clear speaking style on the tone contour, the effect was almost always enhancement of some property of the canonical tone, i.e., a high tone getting higher, a sloped tone getting more steeply sloped, a curvy tone getting curvier. Yet, the size of these differences was relatively small, indicating that while the tone contours were being enhanced, the contours were not being over-articulated in a way that would distort the canonical tone contour and might inhibit speech perception.

Chapter 4 investigated the clear speech intelligibility benefit for clear speech in Mandarin. The clear speech enhancements for Tone 1, Tone 3, and Tone 4 produced by L1, HL, and L2 Mandarin speakers did result in increased intelligibility for L1 Mandarin listeners. The exception to this is HL Mandarin Tone 4, which did not show increased intelligibility in clear speech, likely due to the fact that HL Mandarin Tone 4 was overall highly intelligible to listeners. Tone 2 did not show an increase in intelligibility for any speaker group, and in fact, HL and L2 Tone 2 *decreased* in intelligibility in clear speech. Tone 2 and Tone 3 were confusable for listeners, with many listeners identifying Tone 3 productions as Tone 2. As a group, L1

Mandarin speakers were able to increase Tone 2-Tone 3 identification accuracy in clear speech, while HL and L2 speakers were not.

5.2 Overall Discussion

Each experiment chapter from this dissertation (Chapters 2-4) has discussed individual findings from that experiment. In this section, we will discuss results across chapters, synthesizing the results and relating them to different theories and topics in clear speech research and L2 and HL acquisition. In particular, we are interested in comparing the results of global clear speech strategies, tone contour enhancements, and clear speech intelligibility benefits from a variety of viewpoints, including cross-linguistic clear speech strategies, L2 and HL acquisition, and speech production theories.

5.2.1 Cross-linguistic clear speech strategies

A primary research question of this dissertation was to identify clear speech strategies that are used in both English, a non-tonal language, and Mandarin, a tonal language, in a direct comparison. We found that longer duration and expanded pitch range were characteristic of clear speech in both languages. This is in accordance with previous research finding longer duration and expanded pitch range in clear speech across languages, including English (Smiljanić & Bradlow, 2005; Kato & Baese-Berk, 2023), Mandarin (Zhang et al., 2024; Rittenberry & Shport, 2024), and Croatian (Smiljanić & Bradlow, 2005). There are some exceptions to expansions in pitch range in clear speech; in our study, HL Mandarin speakers did not expand pitch range in Mandarin clear speech. Kato and Baese-Berk (2023) found that only high-proficiency L1

Mandarin-L2 English talkers expanded pitch range in English clear speech, while lower proficiency L2 talkers did not expand pitch range.

Intensity was sometimes used as a cue in clear speech in both English and Mandarin, although this seems to vary depending on talker, with only some talkers using intensity as a clear speech strategy (as in Rittenberry & Shport, 2024). Long (2014) characterizes changes in intensity of 1 dB as “noticeable” to humans, changes of 3 dB as “very noticeable,” and changes of 6 dB as “substantial.” Between-talker variation in intensity was 5.6 dB, reflecting large differences in baseline intensity for different talkers. Yet, while differences in intensity between speaking styles were statistically detectable, they were small, usually around 1 dB. Some individual talkers produced differences of 2 dB between speaking styles. The size of the clear speech adjustments in the present study is comparable to those found by Tupper et al. (2021), who found increases of about 0.5 dB to 1 dB for each Mandarin tone in clear speech.

In the present study, mean f_0 increased only in L1 Mandarin clear speech, with little to no increase in English clear speech for any language background. This differs from previous studies finding an increase in f_0 mean in English clear speech (e.g., Kato & Baese-Berk, 2023). The L1 Mandarin increase in mean f_0 found in Chapter 2 was likely driven by tone-specific f_0 mean increases in the higher-pitch tones T1 and T4, which we found in Chapter 3 (a finding that conflicts with Tupper et al. (2021)’s finding that mean pitch increased only for T2 and T3 in clear speech, and Zhang et al. (2024) and Chien and Brajot (2025) that mean f_0 increases for T1 and T2). Additionally, in the present study, different talkers used different strategies for changes in mean f_0 in clear speech, with some increasing f_0 for all tones, and some increasing f_0 only for higher-pitch tones. Overall, it seems that raising of f_0 mean is one strategy for clear speech enhancement. However, lack of consistency across talkers in changes in f_0 mean in clear speech

in the present study, as well as in prior work (Rittenberry & Shport, 2024), indicates that this strategy is not one that is used by all talkers, unlike more widely-used strategies of increasing duration and widening f_0 range. Instead, only some speakers use a strategy of increasing mean f_0 .

The clear speech strategy of increasing f_0 may also differ across talkers, as we see some speakers using a strategy for increasing mean f_0 in a tone-specific way: increasing the mean f_0 of higher pitch only for higher-pitch tones in Mandarin (as in Papoušek and Hwang (1991), where maximum f_0 increased in Mandarin clear speech, but minimum f_0 stayed the same). On the other hand, a few speakers also use this strategy in a tone-general way, increasing f_0 for all tones in Mandarin (as in Rittenberry and Shport (2024), where maximum and minimum f_0 increased in clear speech for all tones, though maximum f_0 increased more than minimum f_0). It is unclear whether the “tone-specific” strategy for f_0 raising observed in some talkers in the present study is truly associated with specific tones, or whether this could be the result of a more general strategy of “raising the highs.” Such a strategy could be applied for both tonal and non-tonal languages. For example, if we examined pitch of clear speech in English in longer speech, perhaps we would see some talkers increasing pitch across the board (“raising all”), and other talkers increasing the pitch of high-pitch targets (“raising the highs,” e.g., only increasing the pitch of words in the sentence with more emphasis). Either way, from the current study and from inconsistent results in f_0 raising in clear speech in previous studies, it seems like f_0 raising is a strategy used by only some talkers in clear speech, unlike the more widely-used strategies of increased duration and increased pitch range.

As for f_0 contour, this study was the first to compare the pitch contour of English clear and casual speech directly to the pitch contours of a tonal language. English monosyllables

produced in isolation had a pitch contour characterized by a negative slope and a curve close to zero, indicating a falling contour with a steady decrease in pitch over time. The shape of the English pitch contour was distinct from the shape of Mandarin T4, which also has a negative slope but is characterized by a concave curve. The English pitch contour was enhanced in clear speech through increased slope; this corresponds to the increase in pitch range found in previous studies (Kato & Baese-Berk, 2023). From the perspective of comparing the shape of the pitch contour in clear and casual speech, the English pitch contour was enhanced in the same ways as the Mandarin pitch contours. In other words, characteristic aspects of each contour were enhanced in clear speech. In L1 English and L1 Mandarin, these enhancements included steeper slope for English, higher mean pitch for T1, lower mean pitch and curvier curve for T3, and steeper slope and curvier curve for T4.

When it comes to intelligibility, there were differences for intelligibility gains in English and Mandarin clear speech, though those differences reflect the nature of the tasks for each language. For the Mandarin perception task, listeners were asked to identify only the tone, while for the English perception task, listeners were asked to identify only the vowel. In high levels of noise (SNR -9 dB), vowel identification is challenging, and any clear speech modifications made by talkers were not enough to overcome the high noise level and increase intelligibility of the vowel for listeners. On the other hand, clear speech adjustments in f_0 (or other cues) were enough to increase intelligibility for certain tones in L1, HL, and L2 Mandarin clear speech.

5.2.2 HL and L2 acquisition of Mandarin tones

The results from this dissertation research can give insight into HL and L2 speakers' production and perception of Mandarin tones. As in previous work, the present study found

differences in the acoustic properties and relative intelligibility of L1, HL, and L2 tones. In Chapter 2 we found that L2 Mandarin talkers produced a narrower f_0 range in Mandarin than L1 and HL talkers. This differs from Chang and Yao's (2016) finding that L1, HL, and L2 speakers did not differ in overall f_0 range, but matches Yang's (2015) finding that L1 speakers produce a wider pitch range and L2 speakers produce a narrower pitch range, with HL speakers patterning intermediately to the other two groups.

As for the shape of the tonal contour, in Chapter 3, HL and L2 speakers largely produced tonal contours with the same characteristic features as L1 speakers. The greatest differences were found in the contours of T2 and T3, tones which have previously been found to be the most challenging for HL and L2 speakers (Song, 2021; Chang & Yao, 2016; Hao, 2012; 2014). All speaker groups produced T1 and T4 with positive (higher) mean f_0 and T2 and T3 with negative (lower) mean f_0 , although L1 T3 had a lower mean than L2 and HL speakers' T3, perhaps contributing to the finding in Chapter 2 of a wider f_0 range for L1 Mandarin speakers as compared to L2 Mandarin speakers. As for f_0 slope, all speakers produced T1 with a slope close to zero and T4 with a negative slope. L1 and HL speakers produced T2 with a positive slope, but L2 speakers produced T2 with a slope very close to zero. L1 speakers produced T3 with a negative slope, while HL and L2 speakers produced T3 with a slope close to zero. As for f_0 curve, all speaker groups patterned similarly, with no curve for T1 and a small convex curve for T2. The curviest tone contours were T3 and T4; all speaker groups produced convex curves for T3 and concave curves for T4, though L1 speakers produced curvier tone contours for T3 and T4 than L2 and HL speakers. The curve of T3 may correspond to the turning point investigated by Chang and Yao (2016), who found that L2 speakers had an earlier turning point for T3 than L1 and HL speakers, perhaps reflecting a shallower curve for T3 produced by L2 speakers.

It should be noted that many HL Mandarin talkers in this study spoke Taiwanese Mandarin, while the L1 Mandarin talkers and listeners tended to be from mainland China. Overall, the general shape of tone contours in Taiwanese Mandarin are similar to Standard Mandarin. In Taiwanese Mandarin, like in Standard Mandarin, both T2 and T3 show convex curves, with a rising slope being a cue for T2 and a falling slope being a cue for T3, and T3 being curvier than T2 (Fon et al., 2004). Taiwanese Mandarin and Standard Mandarin do show differences in tonal contour, especially for T3, which is more commonly produced as with a falling contour than a dipping contour in Taiwanese Mandarin (Chang, 2010). We might expect the differences between these two dialects to be reflected in the tone contour production of HL and L1 Mandarin speakers in this study. Yet, the tone contour patterns of Taiwanese Mandarin do not explain either the HL pattern of zero slope for T3, nor the similar amount of convex curve for T2 and T3 produced by HL speakers in the current study.

Overall, intelligibility was very high for both HL and L2 talkers, though slightly lower for L2 speakers than for HL and L1 speakers. For both HL and L2 talkers, as in previous research, intelligibility was lowest for T3 (Chang and Yao, 2016). This resulted from frequent misidentification of T3 as T2, an effect that was more common for HL and L2 talkers than for L1 talkers. This could result from differences in the tonal contour of T3 as produced by HL and L2 talkers compared to L1 talkers. In particular, HL and L2 talkers produced T3 with a slope close to zero, and with reduced curve as compared to L1 talkers. Yet, HL talkers' productions of T4 were more intelligible than both L1 and L2 talkers' T4 productions. Increased intelligibility for HL talkers as compared to L1 talkers has been found before (Dahland & Oh, 2014). Overall, effect sizes between different tones and language groups were small (except for the reduced accuracy of T3), likely because overall accuracy was very high for most tones.

As in previous research, the present dissertation found inter-talker variability in Mandarin tone production and perception. Talkers varied in the overall duration and f0 range of their productions, and also showed large differences in overall intensity. Talkers also differed in the mean and curve of each tone contour, as well as in the slopes of T2, T3, and T4. The present study did not investigate the relative amount of variation between different talker groups, though previous research has found a great deal of variation in HL Mandarin tone production (Chang & Yao, 2016; Yang, 2015).

5.2.3 HL and L2 acquisition of a clear speech register

Speakers in all language groups showed acoustic differences between clear and casual registers in Mandarin, and intelligibility increased overall in the clear register, demonstrating the ability of HL and L2 talkers to enhance speech to increase intelligibility to a listener. As found in previous studies of L1 Mandarin clear speech (Zhang et al., 2024; Tupper et al., 2021; Rittenberry & Shport, 2024), duration was the most widely-used cue for clear speech, with all speaker groups speaking slower in clear speech. Increase in pitch range was another widely-used cue, with talkers producing an expanding pitch range in clear speech, as in prior work (Rittenberry & Shport, 2024; Zhang et al., 2024; Papoušek & Hwang, 1991). Like the L1 talkers in this study, HL and L2 talkers also changed the tonal contour in clear speech, enhancing the characteristic features of each tone. This is different from previous studies which have found few tone-specific contour enhancements between clear and casual registers in L1 Mandarin (Rittenberry & Shport, 2024; Tupper et al., 2021), although other studies have found some tone-specific enhancements in mean pitch and slope of a few tones in clear speech (higher mean pitch

of T1 and T2, lower mean pitch for T3, and steeper slope for T2 and T4 (Zhang et al., 2024); higher mean pitch of T1 and T2, and steeper slope for T3 and T4 (Chien & Brajot, 2025)).

Although L2 and HL Mandarin speakers produced acoustically distinct clear and casual speaking styles, L2 and HL speakers did not always make the same acoustic changes as L1 speakers. Tone contour changes made by L1 speakers always enhanced a characteristic feature of the tone (higher mean pitch for T1 and T4, steeper slope for rising T2, curvier T2, T3, and T4), as did tone contour changes made by HL speakers (higher mean pitch for T1 and T4, lower mean pitch for T3, steeper slope for falling T4, curvier T3 and T4). On the other hand, L2 speakers produced T3 with more of a rising slope in clear speech, which is not characteristic of L1 T3 and makes the tone contour more similar to the rising T2. This could reflect an overemphasis by L2 speakers on the rising portion of the falling-rising T3 contour. The other tone contour changes made by L2 speakers did enhance the characteristics of the tones (higher mean pitch for T1 and T4, steeper slope for falling T4, curvier T2, T3, and T4). The steeper slope of T4 in clear speech produced by HL and L2 speakers in this study matches with previous work that has found steeper slope of T4 for L1 speakers (Zhang et al., 2024; Tupper et al., 2021), although the L1 speakers in this study did not as a group enhance T4's negative slope in clear speech.

At times, the acoustic adjustments produced by HL and L2 speakers were smaller in size than changes made by L1 speakers, as in some previous work (Kato & Baese-Berk, 2022); L1 speakers made larger enhancements for curve of T3 and T4 in clear speech than HL and L2 speakers. The present study is the first to find differences in the curve of the tonal contour between clear and casual speaking styles, as neither Tupper et al. (2021) nor Zhang et al. (2024) found a difference in curve of tones in L1 Mandarin clear speech.

In clear speech, intelligibility increased for T1, T3, and T4 in clear speech for all speaker groups, demonstrating the effectiveness of the clear speaking style in increasing intelligibility for a listener for L1, HL, and L2 talkers. The exception to this is that intelligibility did not increase for HL T4, which had very high baseline intelligibility. Intelligibility decreased for T2 in clear speech for HL and L2 talkers. HL talkers enhanced the slope of T2 in clear speech, while L2 speakers increased the curve of the T2 contour in clear speech. On the other hand, L1 talkers increased both the steepness and curviness of the T2 contour in clear speech, and while the intelligibility of L1 T2 did not increase in clear speech, it did not decrease either. The decrease in intelligibility for HL and L2 talkers may reflect the inefficacy of the L2 strategy of simply increasing the curviness of T2, and the HL strategy of simply increasing the slope of T2. One reason for this change in accuracy for HL and L2 T2 in clear speech could be that the longer duration of T2 in clear speech leads to perception of T2 as T3. Shen and Lin (1990) found that one of the primary perceptual differences between T2 and T3 is the turning point of the convex curve: earlier in the syllable for T2, and later in the syllable for T3. When T2 is produced in clear speech, the syllable is produced with a longer duration. Without sufficient changes to the tonal contour, this results in a later turning point in the syllable, which leads to an increase in perception of T2 as T3.

The most common error in tone perception was perception of a T3 as a T2, for T3 productions from all talker groups. This aligns with previous research finding acoustic similarity between T2 and T3, and difficulty in distinguishing these tones perceptually (e.g., Hao, 2012). In L1 speech, this tone confusion occurred only in casual speech, and accuracy was very high in clear speech, demonstrating the effect of the clear speaking style on discriminability of T2 and

T3. On the other hand, this tone confusion persisted in the clear speech productions of L2 and HL speakers.

Talkers in all language groups varied in the extent to which they enhanced different acoustic features of clear speech, and in the extent to which talkers produced an intelligibility benefit in clear speech, although variation in the size of clear speech effects was smaller than overall variation across tones. As with previous studies, talkers also varied in which features they enhanced the most (Rittenberry & Shport, 2024). Overall, longer duration was the most common strategy, followed by increased pitch range, but some individual talkers also increased intensity and mean f_0 in clear speech, though effect sizes were small.

5.2.4 Implications for language teaching

The results from this study have several implications for language teaching. Firstly, the present study identifies T2 and T3 as the most challenging tones for L2 and HL Mandarin speakers, in accordance with many previous studies (Hao et al., 2012; Chang & Yao, 2016; Song et al., 2021). Previous research on perception of Mandarin tones has found that lengthening T2 makes the turning point later in the syllable, resulting in increased perception of T3 (Shen & Lin, 1991). The present study did see some examples of this. In particular, the intelligibility of T2 decreased slightly in clear speech for HL and L2 speakers, while intelligibility of T2 did not change for L1 listeners, suggesting that acoustic adjustments between clear and casual T2 are not effective. Targeted practice in producing a T2 contour in clear speech may help learners produce a T2 contour that is distinct from T3.

Despite the decrease in intelligibility of T2 in clear speech, overall intelligibility of T2 was high. Instead, the tone that was least intelligible for L2 and HL speakers was T3, especially

in a casual speaking style. T3 was often perceived as T2 by L1 listeners, perhaps due to the earlier turning point of T3 in the shorter syllables of the casual speaking style. While this pattern of misidentification also occurred in L1 Mandarin casual speech, L1 Mandarin speakers were successful at keeping these tones distinct in clear speech, as this pattern of misidentification disappeared in L1 Mandarin clear speech. This suggests a possible area of focus for supporting L2 and HL Mandarin speakers in the ability to increase intelligibility of their speech in a clear speaking style when needed. Behrman (2017) describes a teaching methodology that was successful at helping L2 speakers develop a clear speaking register that they could use to increase the intelligibility of their speech when needed. The production of T3 in clear speech would be an appropriate target for such a training program in Mandarin.

Another contributor to the misidentification of T3 in this study may be the fact that L2 speakers increased the slope of T3 in clear speech. A positive slope is not one of the tonal characteristics of L1 T3, which has a negative slope. Producing T3 with an overall positive slope may result from the unique learning experience of L2 Mandarin speakers, who are often explicitly taught T3's full rising-falling contour before the "half-T3" pronunciation of T3, which has a negative slope and no rising portion. Zhang (2017) argues that teaching full-T3 before half-T3 can result in overuse of the full-T3 pronunciation by L2 speakers, as L2 speakers extend that pronunciation to contexts outside of the monosyllabic and phrase-final positions where the full-T3 contour is appropriate. The monosyllabic stimuli in this study were appropriate for a full-T3 pronunciation, and indeed many talkers in all language groups pronounced T3 with a falling-rising contour. Yet, the fact that the L2 T3 contour had an overall positive slope in clear speech may reflect an over-emphasis on the rising portion of the full-T3 contour, perhaps stemming from an emphasis in instruction on the rise in the T3 contour. Thus, this study adds evidence to a

body of work which promotes a pedagogical practice for L2 tone instruction in which half-T3 is introduced before full-T3.

5.2.5 Speech production theories

Lindblom (1990) proposed that speakers balance different functional pressures during communication, placing different emphasis on factors such as ensuring listener understanding and decreasing articulatory effort in production. The results of the present study are in accordance with this model, as speakers produced acoustic adjustments in their speech under different conditions: a clear speech condition where talkers were told that their productions would be used to help language learners, and a casual speech condition, where talkers were asked to speak to someone familiar with their voice. The clear speech adjustments, which include longer duration and expanded f_0 range, as well as pitch contour enhancements and smaller modifications to intensity and mean f_0 for some talkers, require greater articulatory effort from talkers, showing evidence that talkers prioritize listener intelligibility over minimizing articulatory effort in a clear speaking style.

The size of the acoustic enhancements in a clear speaking style can also inform speech production theories. In the present study, the size of duration and f_0 range enhancements was relatively large and would be noticeable to a human listener. On the other hand, the size of intensity and f_0 mean adjustments were very close to the just noticeable difference in human perception, especially for perception outside of a laboratory setting (Long, 2014). These acoustic differences clearly reflect greater underlying effort on the part of the talker in a clear speaking style. Yet, since these adjustments are so small, it is unclear whether these differences in

intensity and f_0 mean could be meaningful to speech perception, despite being statistically detectable.

This could reflect a lack of efficacy of a clear speaking style, in which some of the adjustments that speakers make are not the correct adjustments (or adjustments of the correct size) to increase intelligibility. On the other hand, these modifications could simply be associated with greater speech effort in a clear speaking style. This raises the question of the conscious or unconscious nature of speech production and differences in speaking style. Do these small changes in mean f_0 and intensity reflect the speaker's idea of what would be helpful to a listener's speech perception? Or do these changes come from an unconscious increase in effort, without being necessarily intentional on the speaker's part?

A similar debate occurs on the unconscious or conscious nature of Lombard speech. The speech enhancements in Lombard speech are often thought to be unconscious and automatic, and Lombard speech is accordingly often described as the Lombard reflex (Junqua, 1996; Zhao and Jurafsky, 2009). Yet, Lau (2008) argues that there is a communicative element to Lombard speech beyond an unconscious physiological reflex, as the size of acoustic adjustments in Lombard speech increases when there is a communicative element to elicited speech, as opposed to simply reading a list of words aloud in a noisy environment (Garnier et al., 2018). On the other hand, a noisy environment and the communicative nature of a task may be two separate functional pressures that can both affect the way a talker produces speech.

The results of the current study touch on the mechanisms and degree of automaticity of clear speech adjustments; are all acoustic adjustments of clear speech intentional and oriented towards aiding listener perception, or are some automatic or associated with higher effort? Future research can investigate these questions in more detail.

5.3 Future Directions

It is well-established that Mandarin tones have different tonal contours and acoustic characteristics depending on context, with influence from neighboring tones in connected speech (Xu, 1994; Yang et al., 2017). The present study focused on tones in monosyllables, but investigation of clear speaking styles in multisyllabic speech is vital for thoroughly understanding clear speech strategies. In particular, while L2 Mandarin learners often produce individual syllables as pronunciation practice in the classroom, production of isolated monosyllables is not a frequent or common linguistic task in day-to-day Mandarin Chinese, as commented by one L1 Mandarin participant in the study. Rittenberry and Shport (2024) and Zhang et al. (2024) both investigated Mandarin clear speech in longer contexts. Following these studies, investigation of HL and L2 Mandarin clear speech in longer utterances is an important future direction to enhance our understanding of the clear speech strategies used by L2 and HL Mandarin speakers. In particular, the present study found enhancements of the tonal contour in monosyllabic L2 and HL Mandarin clear speech; it remains to be seen whether these enhancements would also be found in longer clear and casual utterances by these speakers.

The exact methodology used to elicit clear and casual speech may also influence the results of this dissertation. The wording of instructions given to participants can result in different acoustic modifications (Lam et al., 2012), and acoustic modifications are often larger in read speech than in interaction with real interlocutors (Scarborough & Zellou, 2013; Scarborough et al., 2007; Hazan & Baker, 2011). In particular, the instruction to “speak quickly” may have induced greater tone reduction due to fast speech in the casual speech condition. The present study also only investigated read laboratory speech; future research can also examine the

speech enhancements produced by L2 and HL Mandarin speakers in communicative tasks or more naturalistic environments.

The differing syllable structures of the English and Mandarin stimuli also limits some of the conclusions we can draw from this study. The English words were hVd monosyllables, while the Mandarin words were V monosyllables. This allowed investigation of clear speech characteristics of vowels in a monosyllable across languages. However, these different syllable properties can also have an influence on the acoustic properties of speech in these languages, and a more robust cross-linguistic comparison could use syllable types that exist in both languages (e.g., CV), or longer utterances that contain many different syllable types. This would allow for investigation of clear speech enhancements in many different phonological contexts.

A further methodological consideration in the present study is the handling of creaky voice in f_0 measurement. Creaky voice is a well-established feature of T3, and has also been found to occur in T4 (Zhang et al., 2024). In the present study, irregular pulses at the end of a syllable were excluded and the syllable boundary was established at the last regular cycle in the waveform. This meant that some irregular pulses at the end of a syllable were often excluded for T4 and sometimes T3. On the other hand, creaky voice in the middle of a syllable, as often occurred in T3, was included the analysis by marking each pulse in ProsodyPro (Xu et al., 2013), including the irregular creaky pulses. Many studies remove tokens with creaky voice when analyzing f_0 (e.g., Zhang et al., 2024). It is possible that the present study's methodological handling of T4 and creak at the end of a syllable could obscure some differences in T4 between clear and casual speech.

The statistical models in Chapter 3 estimate pitch contours in two steps: first fitting a polynomial, and then modeling those estimates based on factors such as speaking style and talker

language group. This follows Tupper et al. (2021), who also use a two-step model approach to model tone contours in Mandarin. However, treating polynomial estimates as dependent variables in a second statistical model is not best statistical practice. Future analyses can follow Zhang et al. (2024) in using growth curve analysis to directly model the tonal contour as changes in fundamental frequency over time, with linear and quadratic components.

The present study provided a cross-linguistic examination of clear speech strategies in English and Mandarin, comparing a tonal language and a non-tonal language. Yet, just as cross-linguistic clear speech studies of non-tonal languages were needed to investigate proposed global clear speech strategies such as longer duration and expanded vowel space (Smiljanić & Bradlow, 2005), more cross-linguistic clear speech studies of tonal languages are important to verify the findings of the current study that pitch contours are enhanced in clear speech according to the contour's characteristics.

In general, more research is also needed investigating HL clear speech enhancements, as current research on this topic is very limited (e.g., Chang, 2017; 2024; Chang & Weiss-Cowie, 2021). In particular, there is a great deal of variation in HL speech stemming from the immense range of linguistic experiences, language backgrounds, family situations, proficiency, and attitudes among HL speakers. In the present study, we included a broad range of HL speakers with various levels of proficiency and exposure to Mandarin, and correspondingly found variation in the clear speech strategies of these speakers. Likewise, we included a broad range of L2 speakers with various levels of proficiency, as well as speakers who were HL Cantonese speakers. These are important factors which could influence the clear speech strategies of speakers. Future research can investigate this variation in clear speech enhancements, perhaps by controlling for speaker characteristics (Chang & Weiss-Cowie, 2021).

The present study investigated L1 Mandarin perception of HL and L2 Mandarin clear speech and found that L1 listeners benefited from acoustic adjustments made in clear speech by HL and L2 Mandarin speakers. However, future research should investigate HL and L2 perception of clear speech as well. This is a particularly interesting question for a tonal language such as Mandarin, as tone perception is vital to language learning, and lexical tone may be a novel phonological dimension for L2 learners without experience with a tonal language. If HL and L2 listeners benefit from certain tonal enhancements in clear speech, that could inform language teaching and guide the creation of educational materials.

It is also possible that, even in cases when clear speech does not benefit intelligibility (which, as it involves word identification, may operate at a lexical level), that clear speech could benefit speech perception in other ways, including discrimination of speech sounds. Investigating the benefit of L1, HL, and L2 Mandarin clear speech on tone discrimination would be an interesting direction for future work.

5.4 Concluding Remarks

Overall, the present dissertation found that L1, HL, and L2 Mandarin speakers acoustically enhance Mandarin clear speech. In both English and Mandarin, these speakers produce clear speech with longer duration, an expanded pitch range, and an enhanced tonal contours, though HL Mandarin speakers did not expand pitch range in Mandarin clear speech. Speakers varied in which acoustic properties were enhanced in clear speech, and on which aspects of the tonal contours were enhanced.

These speech enhancements resulted in intelligibility benefits for L1 Mandarin listeners, as intelligibility increased for T1, T2, and T4 produced in clear speech by L1, HL, and L2

Mandarin speakers (with the exception of no intelligibility benefit for HL T4, which was already very high). There was no difference in intelligibility for L1 T2, but intelligibility of T2 decreased slightly for HL and L2 speakers. Overall accuracy was high except for T3 productions, an effect driven by misidentification of T3 productions as T2. L1 speakers were successful in disambiguating these tones in clear speech, but these tones produced by HL and L2 talkers continued to be misidentified in clear speech. As with production of clear speech, there were individual differences in the intelligibility benefit of clear speech produced by different talkers.

Taken together, the results of this dissertation demonstrate that, despite the increased effort involved in L2 speech production, HL and L2 speakers are able to balance functional pressures in communication and produce speech in a clear speaking style that is more intelligible to a listener. This demonstrates the remarkable flexibility and listener-oriented sensitivity of human speech production.

References

- Androsova, S., Karavaeva, V., & Zhang, Z. (2022). Pitch Variation for Chinese Syllables of Different Information Load (Based on Commercial and Social Radio Ads). *Chuzhdoezikovo Obuchenie-Foreign Language Teaching*, 49(5), 440–450. <https://doi.org/10.53656/for22.512pitc>
- Aoki, N., & Zellou, G. (2023). When speaking clearly does not enhance comprehension: Comparing intelligibility of hard-of-hearing-and non-native-directed speech for native and non-native listeners. *The Journal of the Acoustical Society of America*, 154(4_supplement), A157–A157.
- Au, T. K., Knightly, L. M., Jun, S.-A., & Oh, J. S. (2002). Overhearing a Language During Childhood. *Psychological Science*, 13(3), 238–243. <https://doi.org/10.1111/1467-9280.00444>
- Badh, G., & Knowles, T. (2023). Acoustic and perceptual impact of face masks on speech: A scoping review. *PLOS ONE*, 18(8), e0285009. <https://doi.org/10.1371/journal.pone.0285009>
- Behrman, A. (2017). A Clear Speech Approach to Accent Management. *American Journal of Speech-Language Pathology*, 26(4), 1178–1192. https://doi.org/10.1044/2017_AJSLP-16-0177
- Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, 114(3), 1600–1610. <https://doi.org/10.1121/1.1603234>
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In *The Development of speech perception: The transition from speech sounds to spoken words* / (pp. 167–224). MIT Press,.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language Learning & Language Teaching* (Vol. 17, pp. 13–34). John Benjamins Publishing Company. <https://doi.org/10.1075/llt.17.07bes>
- Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18(1), 37–49. [https://doi.org/10.1016/S0095-4470\(19\)30357-2](https://doi.org/10.1016/S0095-4470(19)30357-2)

- Bradlow, A. R. (2002). Confluent talker-and listener-oriented forces in clear speech production. In *Laboratory Phonology* (Vol. 7, pp. 241–273).
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121(4), 2339–2349. <https://doi.org/10.1121/1.2642103>
- Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *J. Acoust. Soc. Am.*, 112(1).
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking Clearly for Children With Learning Disabilities: Sentence Perception in Noise. *Journal of Speech, Language, and Hearing Research*, 46(1), 80–97. [https://doi.org/10.1044/1092-4388\(2003/007\)](https://doi.org/10.1044/1092-4388(2003/007))
- Burnham, D., Joeffry, S., & Rice, L. (2010). Computer- and Human-Directed Speech Before and After Correction.
- Carreira, M., & Kagan, O. (2011). The Results of the National Heritage Language Survey: Implications for Teaching, Curriculum Design, and Professional Development. *Foreign Language Annals*, 44(1), 40–64. <https://doi.org/10.1111/j.1944-9720.2010.01118.x>
- Chang, C. B., & Yao, Y. (2024). 10 An Individual-Differences Perspective on Variation in Heritage Mandarin Speakers. In *The Phonetics and Phonology of Heritage Languages* (pp. 208–236). Cambridge University Press.
- Chang, C., & Yao, Y. (2016). Toward an Understanding of Heritage Prosody: Acoustic and Perceptual Properties of Tone Produced by Heritage, Native, and Second Language Speakers of Mandarin. *Heritage Language Journal*, 13(2), 134–160. <https://doi.org/10.46538/hlj.13.2.4>
- Chang, C.-Y. (2010). Dialect Differences in the Production and Perception of Mandarin Chinese Tones [PhD Thesis]. Ohio State University.
- Chang, S.-E. (2017). Enhancement effects of clear speech and word-initial position in Korean glides. *The Journal of the Acoustical Society of America*, 141(6), 4188–4199. <https://doi.org/10.1121/1.4984061>
- Chang, S.-E. (2024). Acoustic adjustments of clear speech in Korean heritage vowel production. *International Journal of Bilingualism*, 28(5), 1033–1052. <https://doi.org/10.1177/13670069231209814>

- Chang, S.-E., & Weiss-Cowie, S. (2021). Hyper-articulation effects in Korean glides by heritage language learners. *International Journal of Bilingualism*, 25(1), 3–20.
<https://doi.org/10.1177/1367006920935512>
- Chen, A. C.-H., & Tseng, S.-C. (2019). Prosodic encoding in Mandarin spontaneous speech: Evidence for clause-based advanced planning in language production. *Journal of Phonetics*, 76, 100912. <https://doi.org/10.1016/j.wocn.2019.100912>
- Chen, F., Pan, C., Hu, H., Hochmuth, S., Kollmeier, B., & Warzybok, A. (2025). Understanding the Lombard Effect for Mandarin: Relation Between Speech Recognition Thresholds and Acoustic Parameters. *Trends in Hearing*, 29, 23312165251324266.
<https://doi.org/10.1177/23312165251324266>
- Cheng, C. C. (2011). A synchronic phonology of Mandarin Chinese. De Gruyter Mouton.
- Cheng, M. -c., & Chang, K. -c. (2014). Tones in Hakka Infant-Directed Speech: An Acoustic Perspective. *Language and Linguistics*, 15(3), 341–390.
<https://doi.org/10.1177/1606822X14520662>
- Chien, C. T., & Brajot, F.-X. (2025). Changes to Mandarin vowel and lexical tone acoustics in clear speech production. *The Journal of the Acoustical Society of America*, 157(5), 3561–3571. <https://doi.org/10.1121/10.0036645>
- Chuang, C. K., Hiki, S., Sone, T., & Nimura, T. (1972). The acoustical features and perceptual cues of the four tones of standard colloquial Chinese. *Proceedings of the 7th International Congress of Acoustics*, 3, 297–300.
- Cohn, M., Pycha, A., & Zellou, G. (2021). Intelligibility of face-masked speech depends on speaking style: Comparing casual, clear, and emotional speech. *Cognition*, 210, 104570.
<https://doi.org/10.1016/j.cognition.2020.104570>
- Cohn, M., Segedin, B. F., & Zellou, G. (2022). Acoustic-phonetic properties of Siri- and human-directed speech. *Journal of Phonetics*, 90, 101123.
<https://doi.org/10.1016/j.wocn.2021.101123>
- Cooper, W. E., & Sorensen, J. M. (2012). *Fundamental frequency in sentence production*. Springer Science & Business Media.
- Cychosz, M., Edwards, J. R., Bernstein Ratner, N., Torrington Eaton, C., & Newman, R. S. (2021). Acoustic-Lexical Characteristics of Child-Directed Speech Between 7 and 24 Months and Their Impact on Toddlers' Phonological Processing. *Frontiers in Psychology*, 12, 712647. <https://doi.org/10.3389/fpsyg.2021.712647>

- Daland, R., & Oh, M. (2014). Heritage bilingual talkers do not exhibit an intelligibility penalty in either language. *Linguistic Research*, 31(3), 403–440.
<https://doi.org/10.17250/KHISLI.31.3.201412.001>
- Dietrich, S., & Hernandez, E. (2022). *Language Use in the United States: 2019* (ACS 50; American Community Survey Reports). U.S. Census Bureau.
<https://www.census.gov/content/dam/Census/library/publications/2022/acs/acs-50.pdf>
- Domingo, Y., Holmes, E., & Johnsrude, I. S. (2020). The benefit to speech intelligibility of hearing a familiar voice. *Journal of Experimental Psychology: Applied*, 26(2), 236.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (eds.). (2023). Mandarin Chinese. In *Ethnologue: Languages of the World* (26th ed.). SIL International.
<http://www.ethnologue.com>
- Ferguson, S. H. (2012). Talker Differences in Clear and Conversational Speech: Vowel Intelligibility for Older Adults With Hearing Loss. *Journal of Speech, Language, and Hearing Research*, 55(3), 779–790. [https://doi.org/10.1044/1092-4388\(2011/10-0342\)](https://doi.org/10.1044/1092-4388(2011/10-0342))
- Ferguson, S. H., & Kewley-Port, D. (2007). Talker Differences in Clear and Conversational Speech: Acoustic Characteristics of Vowels. *Journal of Speech, Language, and Hearing Research*, 50(5), 1241–1255. [https://doi.org/10.1044/1092-4388\(2007/087\)](https://doi.org/10.1044/1092-4388(2007/087))
- Fernald, A., Taeschner, T., Dunn, J., Papoušek, M., De Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, 16(3), 477–501.
<https://doi.org/10.1017/S0305000900010679>
- Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. York Press.
- Fon, J., Chiang, W.-Y., Cheung, H., 冯怡蓁, 江文瑜, & 张显达. (2004). PRODUCTION AND PERCEPTION OF THE TWO DIPPING TONES (TONE 2 AND TONE 3) IN TAIWAN MANDARIN / 台湾地区国语抑扬调 (二声与三声) 之发声与听辨. *Journal of Chinese Linguistics*, 32(2), 249–281. JSTOR. <http://www.jstor.org/stable/23756118>
- Garnier, M., Ménard, L., & Alexandre, B. (2018). Hyper-articulation in Lombard speech: An active communicative strategy to enhance visible speech cues? *The Journal of the Acoustical Society of America*, 144(2), 1059–1074. <https://doi.org/10.1121/1.5051321>

- Godson, L. (2004). Vowel Production in the Speech of Western Armenian Heritage Speakers. *Heritage Language Journal*, 2(1), 45–70. <https://doi.org/10.46538/hlj.2.1.3>
- Gonçalves, A. R., & Silveira, R. (2015). FREQUENCY EFFECTS ON THE INTELLIGIBILITY OF ENGLISH WORDS WITH HIGH FRONT VOWELS. *Organon*, 30(58). <https://doi.org/10.22456/2238-8915.52584>
- Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, 25(2), 207–231. <https://doi.org/10.1006/jpho.1997.0042>
- Granlund, S., Hazan, V., & Baker, R. (2012). An acoustic–phonetic comparison of the clear speaking styles of Finnish–English late bilinguals. *Journal of Phonetics*, 40(3), 509–520. <https://doi.org/10.1016/j.wocn.2012.02.006>
- Han, H. J., Munson, B., & Schlauch, R. S. (2021). Fundamental frequency range and other acoustic factors that might contribute to the clear-speech benefit(s). *The Journal of the Acoustical Society of America*, 149(3), 1685–1698. <https://doi.org/10.1121/10.0003564>
- Han, Y., Goudbeek, M., Mos, M., & Swerts, M. (2018). Mandarin Tone Identification in Musicians and Non-musicians: Effects of Modality and Speaking Style. *6th International Symposium on Tonal Aspects of Languages (TAL 2018)*, 119–123. <https://doi.org/10.21437/TAL.2018-24>
- Han, Y., Goudbeek, M., Mos, M., & Swerts, M. (2019). Effects of Modality and Speaking Style on Mandarin Tone Identification by Non-Native Listeners. *Phonetica*, 76(4), 263–286. <https://doi.org/10.1159/000489174>
- Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269–279.
- Hao, Y.-C. (2014). The Application of the Speech Learning Model to the L2 Acquisition of Mandarin Tones. *Proceedings of the 4th International Symposium on Tonal Aspects of Languages (TAL 2014)*. 4th International Symposium on Tonal Aspects of Languages (TAL-2014), Nijmegen.
- Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America*, 130(4), 2139–2152. <https://doi.org/10.1121/1.3623753>
- Hazan, V., Uther, M., & Granlund, S. (2015). How does foreigner-directed speech differ from other forms of listener-directed clear speaking styles? *Proceedings of the 18th International Congress of Phonetic Sciences*. ICPhS, Glasgow.

- He, A. W. (2008). An identity-based model for the development of Chinese as a heritage language. In *Chinese as a heritage language: Fostering rooted world citizenry* (pp. 109–121). University of Hawai'i, National Foreign Language Resource Center.
- Ji, J. X., Li, Y., & Yang, X. (2022). 噪音背景下韩国留学生普通话声调产出调整方式研究 [Adjustment strategies in Mandarin tone production in noise among Korean learners]. *汉语学习*, 2022(4), 92–101.
- Ji, J., Hu, Y., Peng, G., & Yang, X. (2025, June 4). Acoustic Features of Mandarin Tone Production in Noise: A Comparison Between Chinese Native Speakers and Korean L2 Learners. <https://doi.org/10.5281/zenodo.15592362>
- Jung, Y.-J., & Dmitrieva, O. (2023). Non-native talkers and listeners and the perceptual benefits of clear speech. *The Journal of the Acoustical Society of America*, 153(1), 137–148. <https://doi.org/10.1121/10.0016820>
- Junqua, J.-C. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication*, 20(1–2), 13–22. [https://doi.org/10.1016/S0167-6393\(96\)00041-6](https://doi.org/10.1016/S0167-6393(96)00041-6)
- Kasisopa, B., Attina, V., & Burnham, D. (2014). The Lombard effect with Thai lexical tones: An acoustic analysis of articulatory modifications in noise. *Fifteenth Annual Conference of the International Speech Communication Association*.
- Kato, M., & Baese-Berk, M. M. (2021). Contextually-relevant enhancement of non-native phonetic contrasts. *Journal of Phonetics*, 88, 101099. <https://doi.org/10.1016/j.wocn.2021.101099>
- Kato, M., & Baese-Berk, M. M. (2022). Perceptual consequences of native and non-native clear speech. *The Journal of the Acoustical Society of America*, 151(2), 1246–1258. <https://doi.org/10.1121/10.0009403>
- Kato, M., & Baese-Berk, M. M. (2023). The Effects of Acoustic and Semantic Enhancements on Perception of Native and Non-Native Speech. *Language and Speech*, 1–32. <https://doi.org/10.1177/00238309231156615>
- Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2020). The language experience and proficiency questionnaire (LEAP-Q): Ten years later. *Bilingualism: Language and Cognition*, 23(5), 945–950.

- Kennedy, S., & Trofimovich, P. (2008). Intelligibility, Comprehensibility, and Accentedness of L2 Speech: The Role of Listener Experience and Semantic Context. *The Canadian Modern Language Review*, 64(3), 459–489. <https://doi.org/10.3138/cmlr.64.3.459>
- Kissling, E. (2018). An Exploratory Study of Heritage Spanish Rhotics: Addressing Methodological Challenges of Heritage Language Phonetics Research. *Heritage Language Journal*, 15(1), 25–70. <https://doi.org/10.46538/hlj.15.1.3>
- Knoll, M. A., & Costall, A. (2015). Characterising F(0) contour shape in infant- and foreigner-directed speech. *Speech Communication*, 66, 231–243. <https://doi.org/10.1016/j.specom.2014.10.007>
- Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, 115(1), 362–378. <https://doi.org/10.1121/1.1635842>
- Lam, J., & Tjaden, K. (2016). Clear Speech Variants: An Acoustic Study in Parkinson’s Disease. *Journal of Speech, Language, and Hearing Research*, 59(4), 631–646. https://doi.org/10.1044/2015_JSLHR-S-15-0216
- Lam, J., Tjaden, K., & Wilding, G. (2012). Acoustics of Clear Speech: Effect of Instruction. *Journal of Speech, Language, and Hearing Research*, 55(6), 1807–1821. [https://doi.org/10.1044/1092-4388\(2012/11-0154\)](https://doi.org/10.1044/1092-4388(2012/11-0154))
- Lau, P. (2008). The Lombard Effect as a Communicative Phenomenon. *UC Berkeley Phonology Lab Annual Reports*, 4. <https://doi.org/10.5070/P719J8J0B6>
- Lee, C.-Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics*, 36(4), 537–563. <https://doi.org/10.1016/j.wocn.2008.01.002>
- Lee, C.-Y., Tao, L., & Bond, Z. S. (2013). Effects of speaker variability and noise on Mandarin tone identification by native and non-native listeners. *Speech, Language and Hearing*, 16(1), 46–54. <https://doi.org/10.1179/2050571X12Z.00000000003>
- Leung, K. K. W., Jongman, A., Wang, Y., & Sereno, J. A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America*, 140(1), 45–58. <https://doi.org/10.1121/1.4954737>
- Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. https://doi.org/10.1007/978-94-009-2037-8_16

- Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43(4), 912.
- Liu, Y., Yao, T.-C., Bi, N.-P., Ge, L., & Shi, Y. (2017). Integrated Chinese: Simplified characters textbook, level 1, part 1 (English and Chinese edition) (4th ed.). Cheng & Tsui Co, Inc.
- Liu, Y., Yao, T.-C., Bi, N.-P., Ge, L., & Shi, Y. (2018a). Integrated Chinese: Simplified characters textbook, level 1, part 2 (English and Chinese edition) (4th ed.). Cheng & Tsui Co, Inc.
- Liu, Y., Yao, T.-C., Bi, N.-P., Ge, L., & Shi, Y. (2018b). Integrated Chinese: Simplified characters textbook, level 2, part 1 (English and Chinese edition) (4th ed.). Cheng & Tsui Co, Inc.
- Lleó, C. (2018). Aspects of the Phonology of Spanish as a Heritage Language: From Incomplete Acquisition to Transfer. *Bilingualism: Language and Cognition*, 21(4), 732–747. <https://doi.org/10.1017/S1366728917000165>
- Lohndal, T., Rothman, J., Kupisch, T., & Westergaard, M. (2019). Heritage language acquisition: What it reveals and why it is important for formal linguistic theories. *Language and Linguistics Compass*, 13(12), e12357. <https://doi.org/10.1111/Inc3.12357>
- Lombard, E. (1911). Le signe de televation de la voix. *Annu. Maladies Oreille Larynx Nez Pharynx*, 27, 101–119.
- Long, M. (2014). Human Perception and Reaction to Sound. In *Architectural Acoustics* (2nd ed., pp. 81–127). Elsevier. <https://doi.org/10.1016/B978-0-12-398258-2.00003-9>
- Lu, X., & Li, G. (2008). Motivation and achievement in Chinese language learning: A comparative analysis. In *Chinese as a heritage language: Fostering rooted world citizenry* (pp. 89–108). University of Hawai'i, National Foreign Language Resource Center.
- Maddieson, I. (2013). Tone (v2020.4). In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Zenodo. <https://doi.org/10.5281/zenodo.13950591>
- Marcoux, K., Cooke, M., Tucker, B. V., & Ernestus, M. (2022). The Lombard intelligibility benefit of native and non-native speech for native and non-native listeners. *Speech Communication*, 136, 53–62. <https://doi.org/10.1016/j.specom.2021.11.007>
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and

- Multilinguals. *Journal of Speech, Language, and Hearing Research*, 50(4), 940–967.
[https://doi.org/10.1044/1092-4388\(2007/067\)](https://doi.org/10.1044/1092-4388(2007/067))
- Mirman, D. (2017). Growth curve analysis and visualization using R. Chapman and Hall/CRC.
- Montrul, S. (2010). Current Issues in Heritage Language Acquisition. *Annual Review of Applied Linguistics*, 30, 3–23. <https://doi.org/10.1017/S0267190510000103>
- Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge University Press.
- Papoušek, M., & Hwang, S.-F. C. (1991). Tone and intonation in Mandarin babytalk to presyllabic infants: Comparison with registers of adult conversation and foreign language instruction. *Applied Psycholinguistics*, 12(4), 481–504.
- Paradis, J., & Genesee, F. (1996). Syntactic Acquisition in Bilingual Children: Autonomous or Interdependent? *Studies in Second Language Acquisition*, 18(1), 1–25.
<https://doi.org/10.1017/S0272263100014662>
- Pelzl, E. (2019). What makes second language perception of Mandarin tones hard?: A non-technical review of evidence from psycholinguistic research. *Chinese as a Second Language (漢語教學研究—美國中文教師學會學報)*. *The Journal of the Chinese Language Teachers Association, USA*, 54(1), 51–78.
<https://doi.org/10.1075/csl.18009.pel>
- Perkell, J. S., Zandipour, M., Matthies, M. L., & Lane, H. (2002). Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues. *The Journal of the Acoustical Society of America*, 112(4), 1627–1641.
<https://doi.org/10.1121/1.1506369>
- Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 28(1), 96–103.
- Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research*, 29(4), 434–446.
- Polinsky, M., & Kagan, O. (2007). Heritage Languages: In the ‘Wild’ and in the Classroom. *Language and Linguistics Compass*, 1(5), 368–395. <https://doi.org/10.1111/j.1749-818X.2007.00022.x>
- Qualtrics. (2025). [Computer software]. Qualtrics. <https://www.qualtrics.com/>

- Raveh, E., Steiner, I., Siegert, I., Gessinger, I., & Möbius, B. (2019). *Comparing Phonetic Changes in Computer-Directed and Human-Directed Speech*.
- Rittenberry, J., & Shport, I. A. (2024). Clear speech effects in production of sentence-medial Mandarin lexical tones. *JASA Express Letters*, 4(5).
- Rogers, C. L., DeMasi, T. M., & Krause, J. C. (2010). Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers. *The Journal of the Acoustical Society of America*, 128(1), 410–423.
<https://doi.org/10.1121/1.3436523>
- Röhr, C. T., Baumann, S., & Grice, M. (2022). The influence of expectations on tonal cues to prominence. *Journal of Phonetics*, 94, 101174.
<https://doi.org/10.1016/j.wocn.2022.101174>
- Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *The Journal of the Acoustical Society of America*, 121(5_Supplement), 3044–3044. <https://doi.org/10.1121/1.4781735>
- Scarborough, R., & Zellou, G. (2013). Clarity in communication: “Clear” speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, 134(5), 3793–3807.
<https://doi.org/10.1121/1.4824120>
- Schaefer, V., & Darcy, I. (2020). Applying a Newly Learned Second Language Dimension to the Unknown: The Influence of Second Language Mandarin Tones on the Naïve Perception of Thai Tones. *Psychology of Language and Communication*, 24(1), 90–123.
<https://doi.org/10.2478/plc-2020-0007>
- Shen, X. S., & Lin, M. (1991). A Perceptual Study of Mandarin Tones 2 and 3. *Language and Speech*, 34(2), 145–156. <https://doi.org/10.1177/002383099103400202>
- Smiljanić, R. (2013). Can older adults enhance the intelligibility of their speech? *The Journal of the Acoustical Society of America*, 133, EL129–EL135.
<https://doi.org/10.1121/1.4776774>
- Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America*, 118(3), 1677–1688.
<https://doi.org/10.1121/1.2000788>
- Smiljanić, R., & Bradlow, A. R. (2007). Clear speech intelligibility: Listener and talker effects. *Proceedings of the 16th International Congress of Phonetic Sciences*. ICPhS XVI, Saarbrücken, Germany. icphs2007.de

- Smiljanić, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics*, 36(1), 91–113.
<https://doi.org/10.1016/j.wocn.2007.02.002>
- Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America*, 130(6), 4020–4031.
<https://doi.org/10.1121/1.3652882>
- So, C. K., & Best, C. T. (2008). Do English Speakers Assimilate Mandarin Tones to English Prosodic Categories? *Interspeech 2008: Proceedings of the 9th Annual Conference of the International Speech Communication Association*, 1120.
- Song, C. (2021). What is in the Final Stage of Inter-Language? Tone Errors and Phonological Constraints in Spontaneous Speech in Very Advanced Learners of Mandarin. In C. Yang (Ed.), *The Acquisition of Chinese as a Second Language Pronunciation* (pp. 21–54). Springer Singapore. https://doi.org/10.1007/978-981-15-3809-4_2
- Tang, P., & Li, S. (2020). The Acoustic Realization of Mandarin Tones in Fast Speech. *Interspeech 2020*, 1938–1941. <https://doi.org/10.21437/Interspeech.2020-1274>
- Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Phonetic enhancement of Mandarin vowels and tones: Infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, 142(2), 493–503. <https://doi.org/10.1121/1.4995998>
- Tang, P., Xu Rattanasone, N., Yuen, I., Demuth, K., & Benders, T. (2025). Due to increased variability, the expanded vowel and tone space in Mandarin IDS does not lead to enhanced contrasts. *Journal of Child Language*, 1–13.
<https://doi.org/10.1017/S0305000925000133>
- Tsukada, K., Xu, H. L., & Rattanasone, N. X. (2015). The perception of Mandarin lexical tones by listeners from different linguistic backgrounds. *Chinese as a Second Language Research*, 4(2), 141–161. <https://doi.org/10.1515/caslar-2015-0009>
- Tupper, P., Leung, K. W., Wang, Y., Jongman, A., & Sereno, J. A. (2021). The contrast between clear and plain speaking style for Mandarin tones. *The Journal of the Acoustical Society of America*, 150(6), 4464–4473. <https://doi.org/10.1121/10.0009142>
- 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. <https://doi.org/10.1121/10.0001024>

- Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, 49(1), 2–7.
<https://doi.org/10.1016/j.specom.2006.10.003>
- Vonessen, J., Aoki, N. B., Cohn, M., & Zellou, G. (2024). Comparing perception of L1 and L2 English by human listeners and machines: Effect of interlocutor adaptations. *The Journal of the Acoustical Society of America*, 155(5), 3060–3070.
<https://doi.org/10.1121/10.0025930>
- Wang, X., & Xu, L. (2020). Mandarin tone perception in multiple-talker babbles and speech-shaped noise. *The Journal of the Acoustical Society of America*, 147(4), EL307–EL313.
- Wu, X. (2023). Use of the “ba” construction in Heritage Chinese speakers and L2 speakers of Chinese [MA Thesis]. University of Saskatchewan.
- Wu, Y., Rough, D., Bleakley, A., Edwards, J., Cooney, O., Doyle, P. R., Clark, L., & Cowan, B. R. (2020). See what I’m saying? Comparing intelligent personal assistant use for native and non-native language speakers. *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–9.
- Xu, N., & Burnham, D. (2010). Tone hyperarticulation and intonation in Cantonese infant directed speech. *Speech Prosody 2010-Fifth International Conference*.
- Xu, Y. (1994). Production and Perception of Coarticulated Tones. *J. Acoust. Soc. Am.*, 95(4), 2240–2253.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25(1), 61–83.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, 27(1), 55–105.
- Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis.
- Yang, B. (2015). *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-44645-4>
- Yang, J., & Fox, R. A. (2017). L1–L2 interactions of vowel systems in young bilingual Mandarin-English children. *Journal of Phonetics*, 65, 60–76.
<https://doi.org/10.1016/j.wocn.2017.06.002>
- Yang, J., Zhang, Y., Li, A., & Xu, L. (2017). On the Duration of Mandarin Tones. *Interspeech 2017*, 1407–1411. <https://doi.org/10.21437/Interspeech.2017-29>

- Yip, M. J. W. (2002). *Tone*. Cambridge University Press.
- Zellou, G., Lahrouchi, M., & Bensoukas, K. (2022). Clear speech in Tashlhiyt Berber: The perception of typologically uncommon word-initial contrasts by native and naive listeners. *The Journal of the Acoustical Society of America*, 152(6), 3429–3443. <https://doi.org/10.1121/10.0016579>
- Zeng, Y., Leung, K. K., Jongman, A., Sereno, J. A., & Wang, Y. (2023). Multi-modal cross-linguistic perception of Mandarin tones in clear speech. *Frontiers in Human Neuroscience*, 17, 1247811.
- Zhang, H. (2017). The effect of theoretical assumptions on pedagogical methods: A case study of second language Chinese tones. *International Journal of Applied Linguistics*, 27(2), 363–382. <https://doi.org/10.1111/ijal.12132>
- Zhang, Y., Chen, F., Xu, F., Guo, C., & Li, K. (2024). Acoustic characteristics of infant-and foreigner-directed speech with Mandarin as the target language. *The Journal of the Acoustical Society of America*, 155(6), 3877–3888. <https://doi.org/10.1121/10.0026359>
- Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37(2), 231–247.
- Zhu, X. (2004). 基频归一化——如何处理声调的随机差异? [F0 Normalization: How to Deal with Between speaker Tonal Variations?] 语言科学, 3(2), 3–19. http://journal15.magtechjournal.com/Jwk_yyqx/CN/Y2004/V3/I2/3

Appendix A: Model Outputs

Figure A.1: Model output for Duration (Chapter 2)

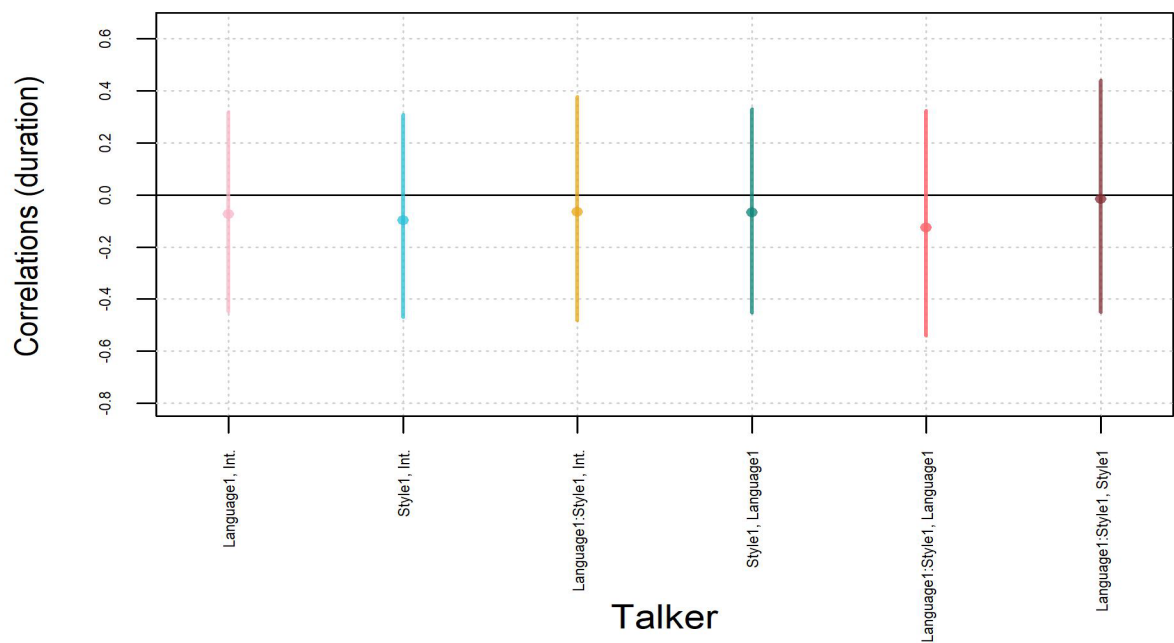
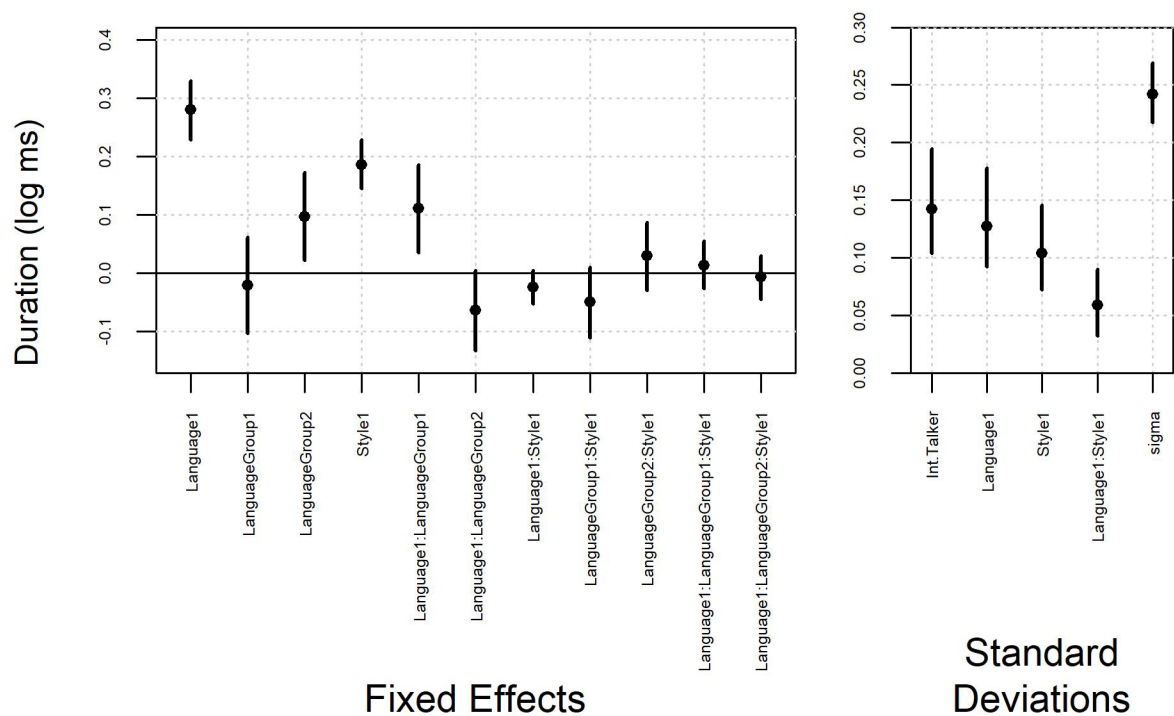


Figure A.2: Model output for Intensity (Chapter 2)

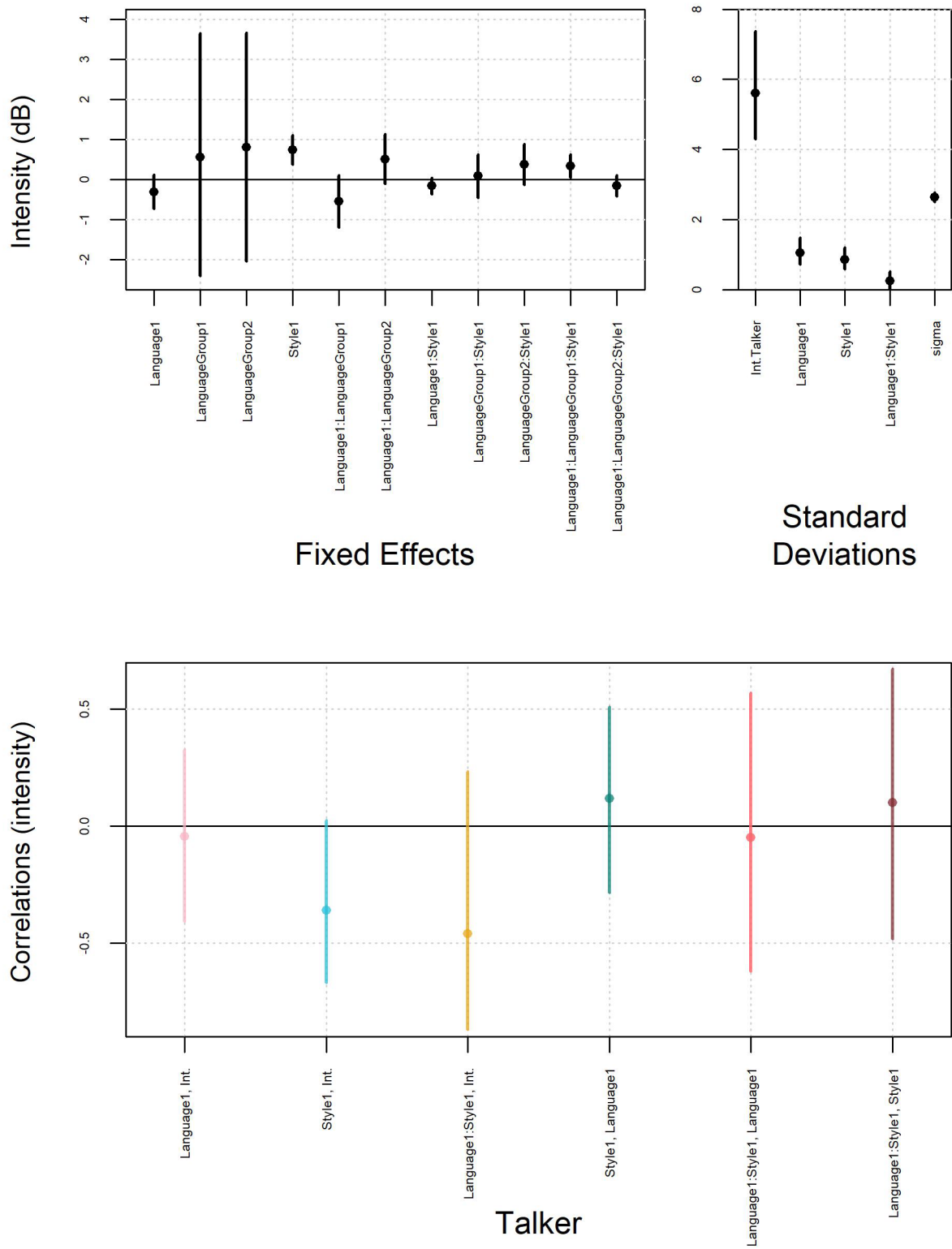


Figure A.3: Model output for Mean f_0 (Chapter 2)

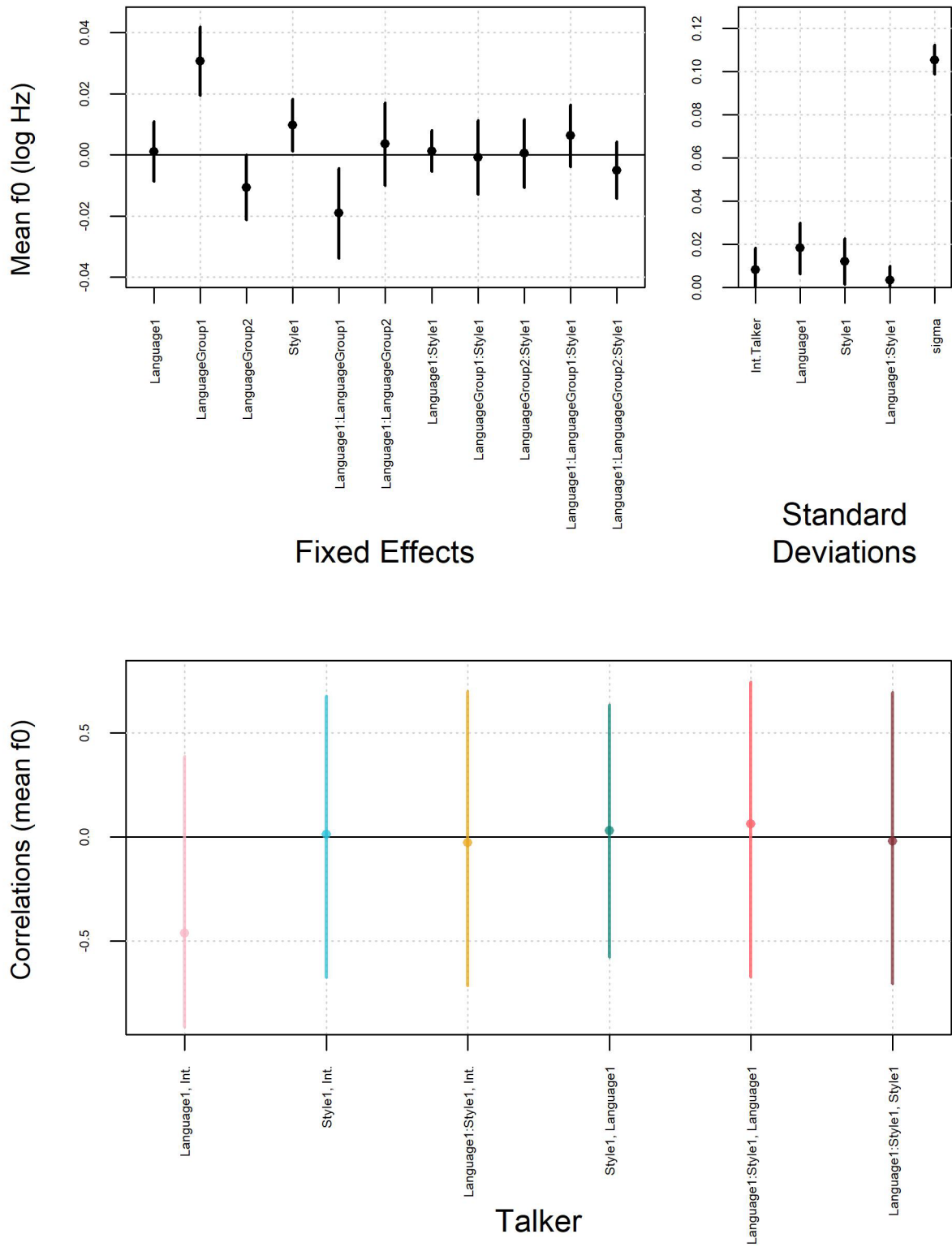


Figure A.4: Model output for F0 range (Chapter 2)

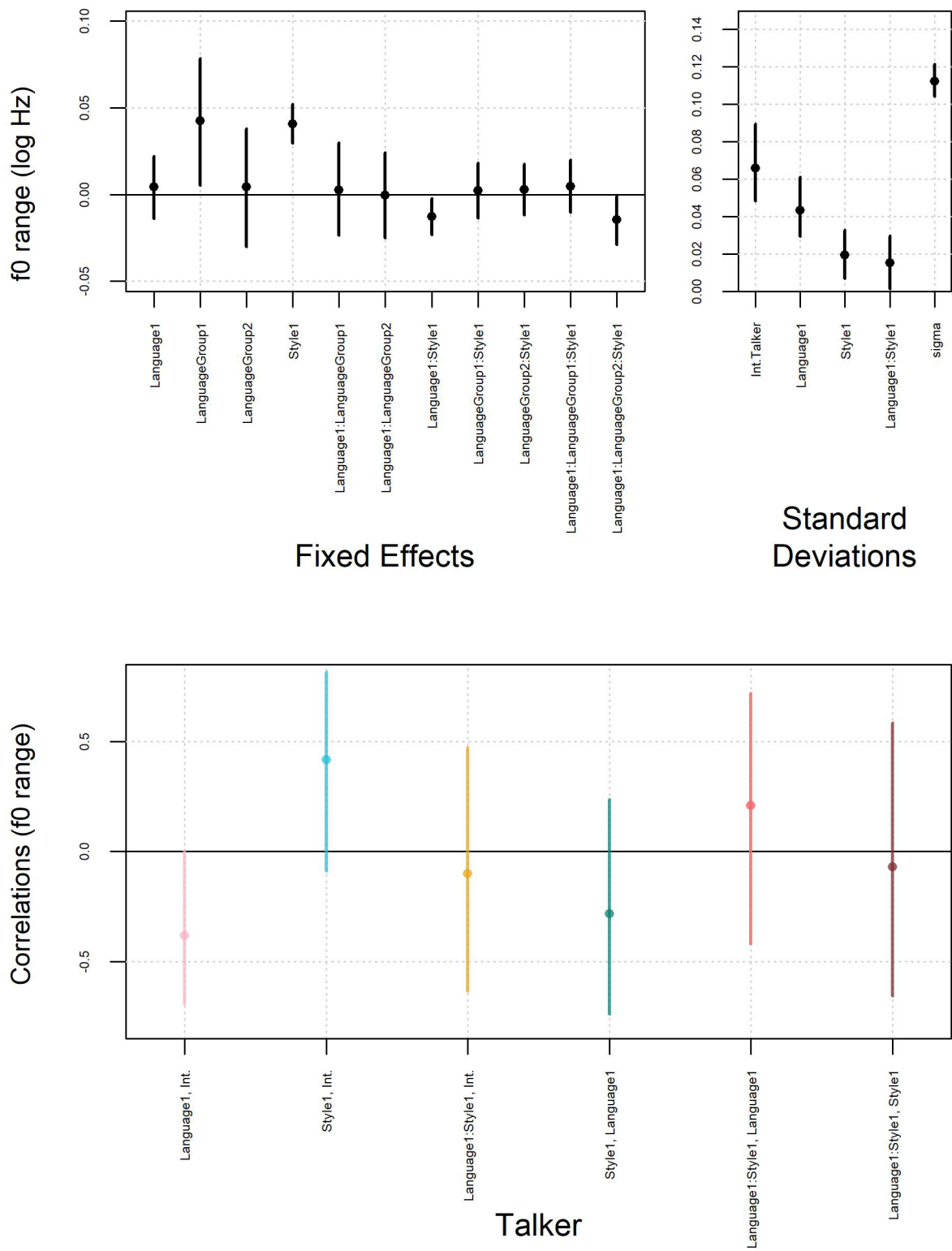


Figure A.5: Model output for f_0 mean (Chapter 3)

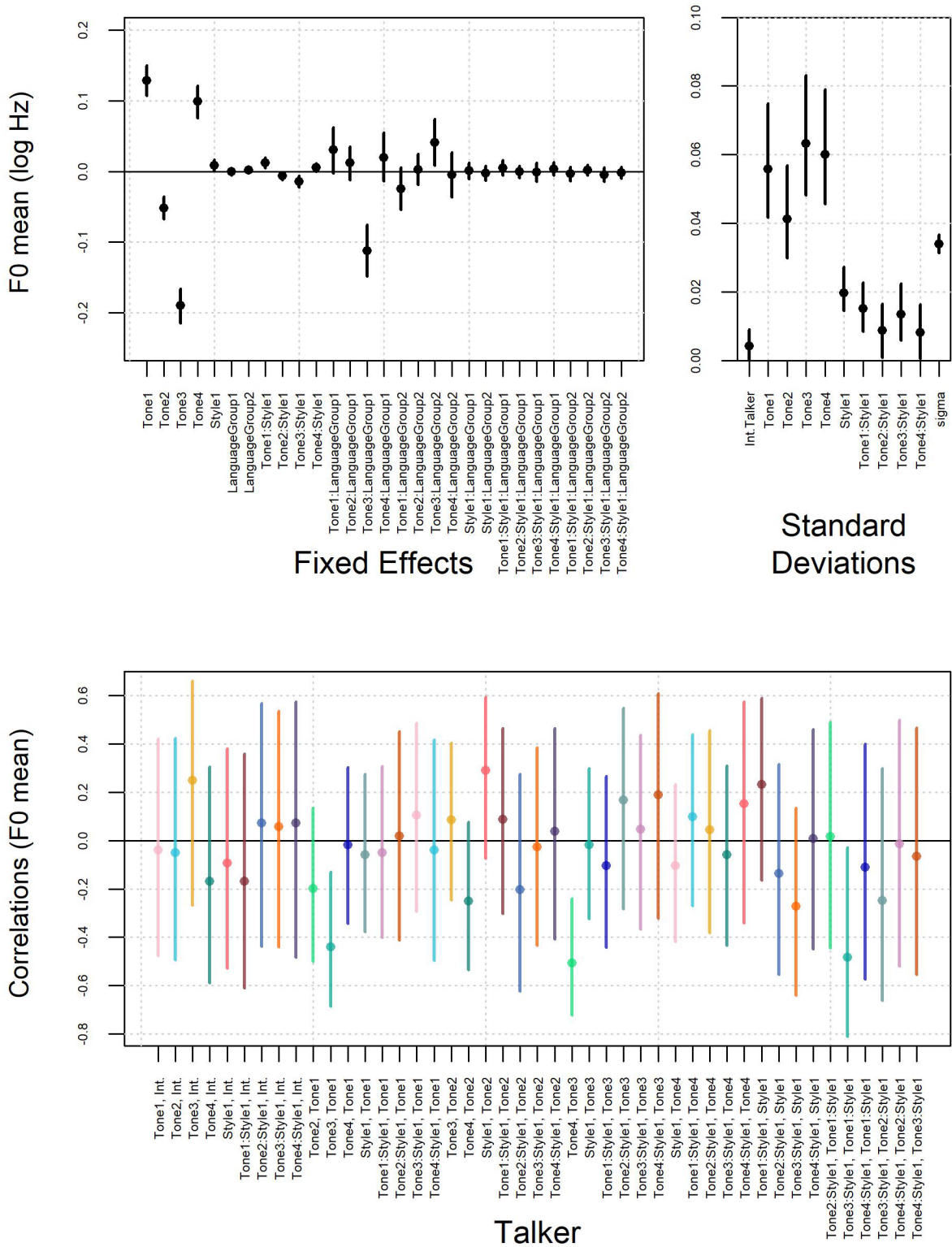


Figure A.6: Model output for f_0 slope (Chapter 3)

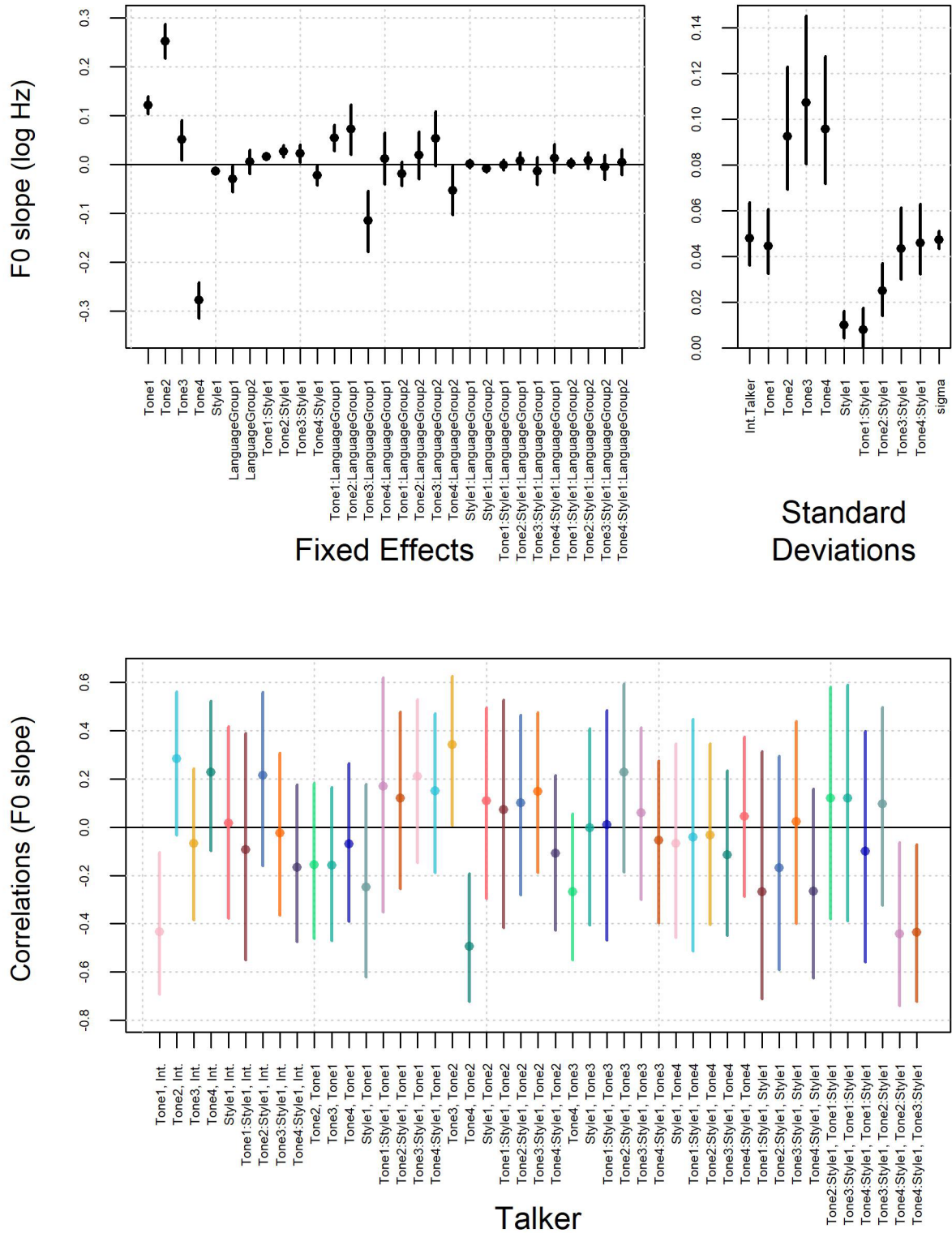


Figure A.7: Model output for f_0 curve (Chapter 3)

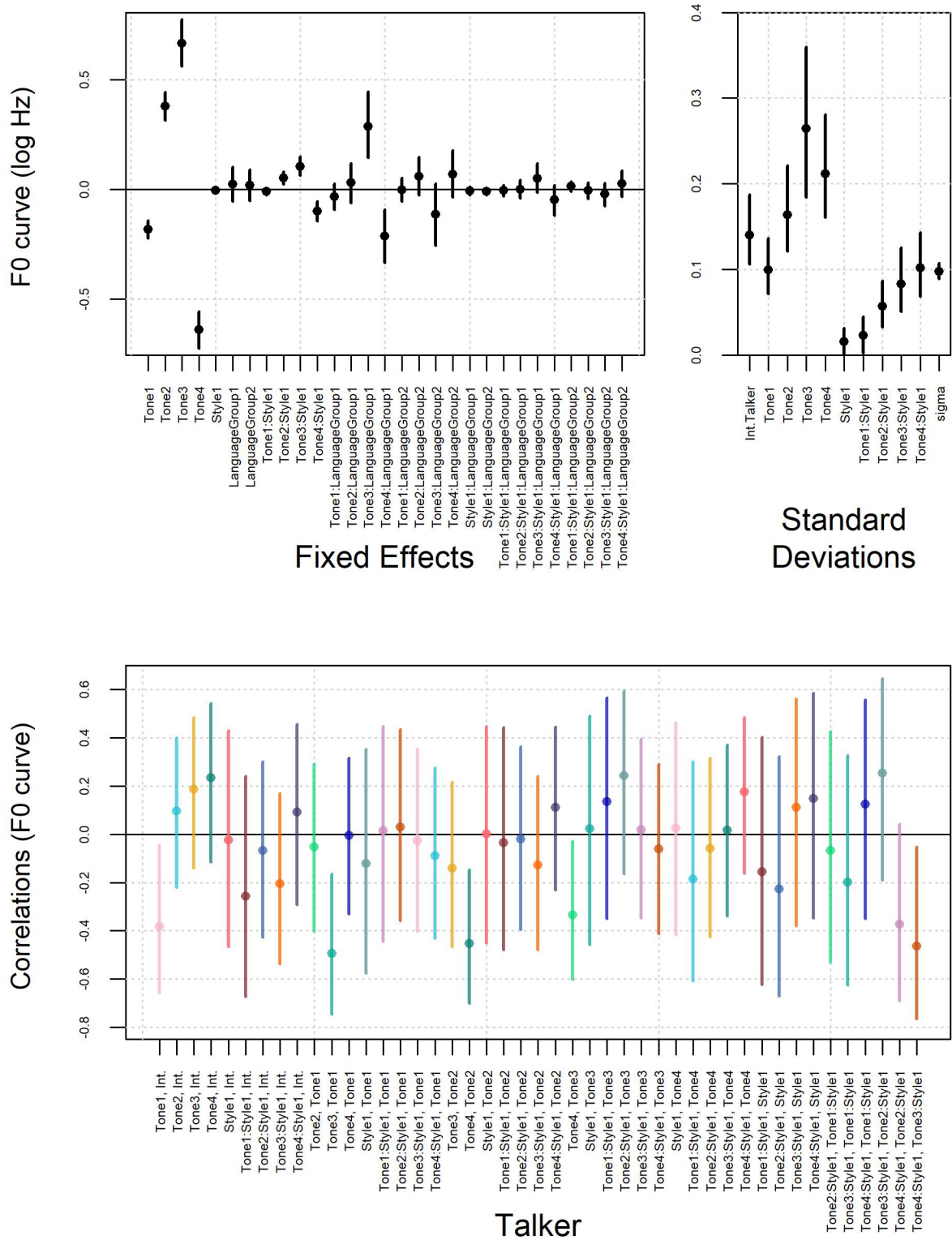


Figure A.8: Model output for intelligibility (Chapter 4)

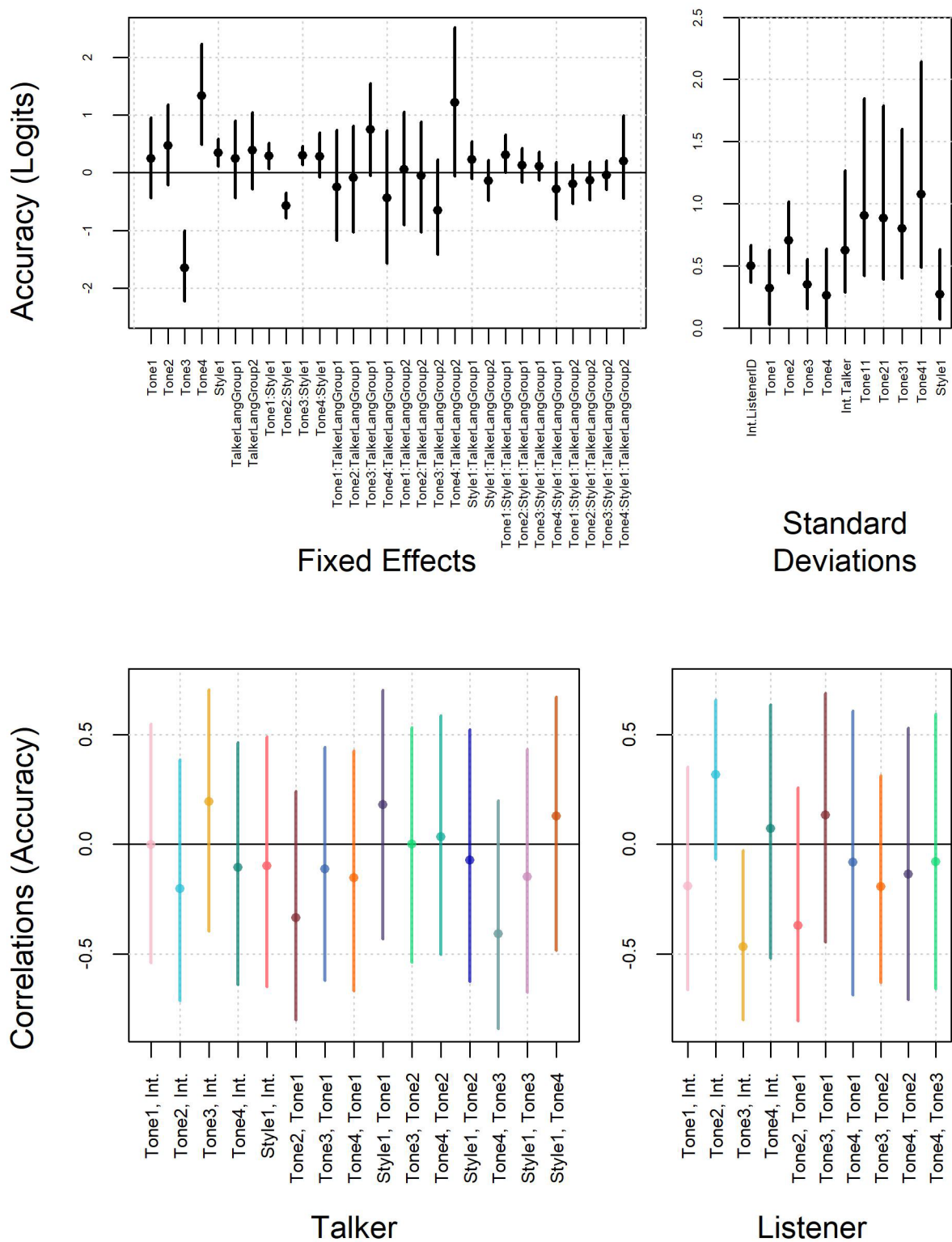


Figure A.9: Model output for Tone 3 data (Chapter 4)

