Investigating the tonal system of Plastic Mandarin: A cross-varietal comparison

Chenzi Xu

St Hugh's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2022

This thesis is dedicated to my parents Shuguang Xu and Huayang Chen, and my granny Yanhui Liu.

致父母徐曙光、陈华扬及奶奶刘艳辉

Acknowledgements

As these are essentially the last lines I write as part of my doctorate thesis, I would like to take this opportunity to express my sincere gratitude to those who have supported me throughout this journey.

Words cannot express my appreciation and gratitude to my supervisor Professor John Coleman, who has been the point of contact whenever I encountered difficulties and challenges in research and teaching. I remembered an insightful conversation with him at the end of my first year in Oxford, which ignited my inspiration for this thesis. He is a remarkably patient and detail-oriented supervisor, who possesses extensive knowledge about Chinese languages. Apart from his wisdom and guidance in acoustic signal processing, scripting, argument presentation, and tutorial teaching, he also picked out the missing spaces and wrongly-used punctuation in my writing. My innumerable thanks also go to Professor Aditi Lahiri, who generously included me in the family of the language and brain lab where I participated in a few neurolinguistic research projects. Her kind encouragement after our encounter at the ICPhS 2019 conference has endowed me some sustained strength in this DPhil journey. I have benefited a lot from her extremely well-organised conference of historical linguistics (ICHL25). I enjoyed analysing linguistic data with her and I always left a meeting with a handful of juicy references.

I have been very fortunate to receive helpful and constructive feedback from Professor Elinor Payne and Professor Janet Pierrehumbert, who acted as my assessors for my Confirmation of Status, and whose comments have guided me in refining the scope of my research. Professor Elinor Payne later served as one of the examiners for my *viva voce* examination, together with Professor Brechtje Post. I appreciate our fruitful discussion during the viva and their suggestions that strengthened this thesis. I also thank Dr Rosalind Temple for her earlier comments on my MPhil thesis.

I am incredibly grateful to my former teachers in Changsha Nanya Middle School, Mr Shuguang Deng, Ms Bo Lü, and Mr Xuqi Gao. My fieldwork, an indispensable part of this research project, would not have been possible without their help. They graciously offered their lecture time in the midst of the hectic boarding school schedule to welcome me into their classrooms to meet their students and assisted me in distributing the recruitment materials. Mr Shuguang Deng kindly helped arrange the piano room in the Arts Centre for speech recordings. Most of all, I am deeply indebted to all of my participants, especially the final-year high school students volunteered their treasured time considering their stress and pressure from the college-entrance exam. It was a pity that I did not manage to deliver them the certificate of participation due to my hasty leave before the Covid travel restriction took effect.

Thanks also go to the Phonetics Laboratory, for the the recording equipment, workspace and lovely tea break chats, the Faculty of Linguistics, Philology, and Phonetics, for the Birla Graduate studentship and travel grants, and the University of Oxford and the Kaifeng Foundation, for the Oxford-Kaifeng Graduate Scholarship. Without these fully-endowed scholarships, my academic journey at Oxford would have never been possible. I thank Professor Wolfgang de Melo for his administrative support. I am also grateful for the internship with Dr Yishu Miao in my first year of DPhil studies, and financial support from my aunt Ms Ke Pan.

I would like to thank Dr Cathryn Yang with whom I had lovely chats at ICHL25 and whose research in tone change in a wide range of Asian languages gave me fresh insights in the case of Plastic Mandarin. I am thankful for Dr Hannah Davidson's proofreading of the early version of chapter 4, Professor Henning Reetz's Praat script of inserting text labels, and Mr Zhan Huang for transcribing some of my data.

Special thanks go to my friends and colleagues in Oxford, since it is them who make my Oxford journey colourful and memorable:

In the Phonetics Laboratory, Toby Hudson, for our teaching discussion and lab socials in Wetherspoons and Wolfson college; Yu-xian (Claire) Huang, for her homemade cakes and great conversations; Dr Mary Baltazani, for her invitation and guidance to conducting interviews; Dr Joanna Przedlacka for her Polish sweets in the lab and a needed hug; Jon Edwards, for his IT support in the lab; Dr José Elias-Ulloa, for the opportunity to deliver the Lab Skills Workshops and suggestions for job interviews.

My former West Cottage housemates, Jheng-Cyuan Lin, Leo Rio Kim, for their wonderful cooking and care; Merlin Wehrs, James Wallace, for their company and support during the pandemic lockdown; my yoga group led by Kamila Nigmatulina, Dr Maria Lucey, Ricardo De Luca e Tuma, for all the shared practices and energy; Jim Hicks, for the guitar and the music jam sessions.

My central support and sources of happiness: Huinan Zeng, Dr Kay Yi Chong, Siyu Tang, Hazal Gezmis, Sunjuri Sun, Dr Otto Simonsson, Dr Koen Evers, Dr Quishi Xia, Dr Cong Zhang, Dr Yong Chen, Dr Beinan Zhou, Dr Swetlana Schuster, Dr Yves Kwan, Dr Amartya Sanyal, David Luan, Junyan Song, Siyu Chen, Dr Hyei Jin Kim, Felix Krones, Maren Fichter, Zirui Huang, Ezio Xie, Songjun He, Chen Xie, Yang Fan, Yiwei Si, Jian Cui, Runyi Yao, Lan Duo, Yingrui Zhao, Hannan Xu for all our college formals, LingLunches, Fish Guy dinners, hot pots, festive celebrations, Switch games, Murder Mysteries, room escapes, travels, and many more fond memories.

Then Dr Oleh Stupak, for generously inviting me to write thesis in the Oxford Internet Institute, a building lives up to its name with extremely fast internet, high-quality monitors and a lovely bunch of friends, and for his love and care.

Lastly, my sweetest thanks go to my parents, Huayang Chen and Shuguang Xu, who have been the most diligent (and free) research assistants for me. They accompanied me to Changsha for my fieldwork, picked me up late night after I finished my recording sessions, and double checked some of the speech transcripts. I would like to thank them for taking care of themselves and other family members, especially during Covid. 爱你们。

Abstract

The city of Changsha, Hunan Province, China has seen an increase in the use of Mandarin in the past decade, overshadowing the local non-Mandarin variety, Changsha. A new variety “Plastic Mandarin”, mostly spoken by millennials and younger generations, has emerged. It is defined in this thesis as a non-standard Mandarin accent that features the speech of young urban residents in Changsha and that has crystallised over the past few decades.

This thesis presents a detailed phonetic investigation of the tonal system of Plastic Mandarin through a cross-varietal comparative approach, mainly divided into two streams: citation tones and neutral tones in contexts. The defining characteristic of the citation tone system for Plastic Mandarin is established first: a mid-level tone, a low to mid rising tone, a low falling tone, and a high rising tone. By comparing the citation tones of the three varieties that coexist in the city of Changsha, the thesis provides acoustic evidence that Plastic Mandarin may arise when Mandarin tones adapt the pitch pattern of some corresponding Changsha tones.

In addition to citation tones, this thesis disentangles the sources of variability in the syllable duration and f_0 contour of speech sequences containing neutral tone syllables, i.e. those do not have any of the four canonical lexical tones and often overlooked in prior studies of tones. The data show that f_0 contours converge at the end of two consecutive neutral tone syllables at a low pitch in both Mandarin varieties. It suggests that a neutral tone or a sequence of consecutive neutral tones tends to be associated with a low pitch target, despite the varying f_0 shapes largely predicted by the preceding lexical tone. The thesis proposes a probabilistic target-approaching model for Mandarin tones in connected speech, in which pitch targets may be fewer than the number of syllables. While the phonetic realisation of the four lexical tones in Plastic Mandarin is consistently different from that in Standard Mandarin, the pitch target of neutral tone syllables tends to remain constant in this process of Mandarin variation and change, which may be attributed to the stable transfer of prosodic structure.

Keywords: Plastic Mandarin, Phonetics, neutral tone, Mandarin variation, citation tone, prosody, f_0 contour, mixed models

Contents

List of Figures

List of Tables

List of Abbreviations

Tone notations:

T . . . . Tone.

T1, T2... . . . Tone categories of a modern Chinese variety.

I, II, III, IV . Middle Chinese tone categories: *ping*, *shang*, *qu*, *ru*, respectively.

Ia, Ib, IIa... . Middle Chinese tone categories after register split. Register a: *yin* (high) and register b: *yang* (low).

˧, ˥˩, ˦˥**...** . . . . Tone letters devised by Chao (1930), also included in IPA Chart (International Phonetic Association, 1999). Each tone letter is consisted of a vertical reference line, to which a simplified pitch-time curve of the tone is attached.

33, 51, 35... . . Numerical notation of tone patterns in accordance with the tone letters by Chao (1930), indicating the pitch value on a 5-point scale, where 1 and 5 correspond to low and high respectively.

H, M, L . . . . High, mid, low.

Syntactic glosses:

CL . . . . Classifier.

DE . . . . Particle *de* marking relative clause in Chinese.

DUR . . . . . . Durative aspect.

GEN . . . . . . Genitive case.

NEG . . . . . . Negation.

NOM . . . . . Nominal suffix.

PL . . . . Plural (e.g. 3PL = 3rd person plural).

PROG . . . . . Progressive aspect.

SG . . . . Singular.

Statistic terms:

LME . . . . . . Linear Mixed Effects.

GAMM . . . . Generalised Additive Mixed Model.

Other abbreviations:

C, V . . . . . . Consonant, vowel.

CASS . . . . . Chinese Academy of Social Sciences.

CS . . . . Changsha.

IPA . . . . . . International Phonetic Alphabet.

PM . . . . Plastic Mandarin.

SM . . . . Standard Mandarin.

1 Introduction

Contents

As the underground train networks sprawled under Changsha city, the capital of Hunan Province in China, and the speed of railways skyrocketed, the way we speak has also had rapid changes there in the past few decades. In addition to the unprecedented skyscraper boom on the banks of the Xiang River, the official national language, Standard Mandarin, has gained momentum in the repertoire of a large part of population which hitherto had spoken the regional vernacular, Changsha[1], one of the Xiang varieties. The addition of the standard variety, both the written and spoken form, in Changsha society has been accelerated by education and mass media, apart from the greater human mobility. The increasingly intensified contact between Standard Mandarin and Changsha was accompanied by the emergence of **Plastic Mandarin**. It has been speedily spreading among younger generations

[1] In this thesis, Changsha is used to refer to the local language or the city.

and rapidly reshaping the local speech patterns. Although Plastic Mandarin has had some media exposure, it has received little scholarly attention.

The thesis is devoted to the investigation of the tonal system of Plastic Mandarin, using a cross-varietal comparative approach, considering the influential role of Standard Mandarin and Changsha may have played in the emergence of Plastic Mandarin. Having established the phonetic description for the citation tones of Plastic Mandarin, the thesis further explores the development of Plastic Mandarin tones, presents a new analysis of tones in connected speech in varied Mandarin varieties, and illuminates the structure and stability in Mandarin variation and change. It examines the phonetic realisation of lexical tone and neutral tone syllables in various contexts through quantitative modelling methods, incorporating the influence of multiple sources in speech production. The comparative approach and variation perspective provide insights into the prosodic organisation of Mandarin and highlight an aspect of prosodic constancy in this Standard–Plastic variation of Mandarin.

This chapter, first and foremost, provides a comprehensive introduction to Plastic Mandarin ranging from the anecdotal origin of its name and the social contexts of its use to young speakers' attitudes towards it, and discusses Plastic Mandarin in the academic discourse of new dialect formation. Then it delineates the significance of the research and the core research questions, followed by an outline of the organisation of the thesis.

1.1 What is Plastic Mandarin?

1.1.1 Origin of the name: An inauthentic Mandarin

The term 'Plastic Mandarin', a literal translation of *sùliào pǔtōnghuà* (塑料普通话) in Mandarin Chinese, or *sù pǔ* (塑普) in its short form, describes a Mandarin variety spoken mainly in Changsha, Hunan Province[2], China. Different from the names of other local Mandarin varieties, which are named after their locality such as Tianjin Mandarin and Jinan Mandarin, this term is rich in meanings and has multiple connotations: first, the pronunciation of *sù* in Changsha, [so], is a homophone of

[2]Hunan Province is also known as Xiang.

another morpheme meaning ‘bad’, thus *sù pǔ* indicates ‘bad Mandarin’; second, *sùliào* ‘plastic’ in Chinese connotes the negative sense of being artificial and fake, in opposition to authenticity. The term hence implies ‘a sham or parody of the high prestige Standard Mandarin’, which reflects Changsha speakers’ self-mocking tone of their Mandarin pronunciation. A summary of the meanings of ‘Plastic Mandarin’ is provided in Table 1.1.

The earliest mention of this term in academic work was in Y. Wu (2005, p. 22), referring to “a variety that combines the features of the local Changsha dialect with features of Standard Mandarin”[3]. The term ‘Plastic Mandarin’ was originally used as an umbrella term covering various kinds of Mandarin-vernacular-hybrid varieties including L2 Mandarin accents when adult monolingual Xiang speakers started to learn Mandarin. Chapter 2 §2.2.2 will further illustrate the major language varieties in the city of Changsha with examples.

In recent decades, the referent of Plastic Mandarin has become crystallised and increasingly identified as a particular non-standard Mandarin with distinct and consistent tonal patterns spoken by young people. This was validated in X. Xu et al. (2012)’s study: about 90% of their 104 survey respondents in Changsha correctly identified the recording of Plastic Mandarin (narrow sense) among recordings of four major varieties in Changsha including Standard Mandarin, Plastic Mandarin, Changsha-accented L2 Mandarin, and Changsha; 55.8% of them differentiated Plastic Mandarin from Changsha-accented L2 Mandarin and some participants considered Changsha-accented L2 Mandarin as a variant of Plastic Mandarin (broad sense). Despite the absence of authoritative definition or standard, most respondents could recognise and were highly aware of the particular Plastic Mandarin. Furthermore, Jing and Niu (2010) observed that the most speakers of such Mandarin were below 30 years old, and those with high academic qualifications were capable of and inclined to code-switching between Standard Mandarin, Plastic Mandarin, and Changsha. In this thesis, Plastic Mandarin refers to the emerged and crystallised Mandarin of the young generations in Changsha.

[3] It is still a mystery when this term first appeared in Changsha and who originated it.

Plastic Mandarin *sùliào pǔtōnghuà* 塑料普通话	Connotations	Bad or fake Mandarin
	Broad sense	Non-standard Mandarin varieties (in Hunan Province)
	Narrow sense	A particular non-standard Mandarin with distinct tonal patterns in Changsha, Hunan

Table 1.1: Summary of the meanings of 'Plastic Mandarin'.

1.1.2 Domains and speakers: A crystallised urban youth speech

Plastic Mandarin is a young and vibrant Mandarin variety that is spreading to broader geographic and social space and to a wider range of age groups, which has come to be linked increasingly to Changsha. Considering that older generations in Changsha do not speak Plastic Mandarin, it was probably formed in the recent two or three decades, the period when Standard Mandarin was massively promoted through education and media under the language legislation and strong language policy. Since 1999, the third week of every September has been designated as the "Putonghua (Standard Mandarin) Promotion Week" across the country, a national annual event jointly guided by nine central government bureaus including the Ministry of Education. Activities and competitions for the annual event are organised through the public school system and by various state and cultural institutions. I had the experience of participating in poetry recitations and drama performance characterised by speaking Standard Mandarin in mid 2000s. Schools are one of the sites where Standard Mandarin came into contact with the vernacular.

1.1.2.1 From de facto lingua franca in schools to public service

From my personal recollection of growing up in Changsha since 2006 and continuing observations, Plastic Mandarin has been predominantly used in schools, particularly in secondary schools, in urban Changsha. It thrives on the realistic gap between policy and practice — despite that Standard Mandarin is expected and required in schools, the accented Mandarin produced by many experienced local teachers

Interlocutor	Number of participants
Classmates and friends in school	20/21
Teachers after classes	16/21
Teachers during classes	13/21
Grandparents	3/21
Parents	8/21
Strangers outside schools	11/21

Table 1.2: Summary of the number of participants who often or always speak Plastic Mandarin to different interlocutors. Strangers outside schools can be taxi drivers, cashiers, shop assistants, etc.

who are adult learners of Mandarin is (has to be) tolerated. Plastic Mandarin is generally considered more standard and intelligible than the accented Mandarin among Mandarin speakers, from the perspective of Standard Mandarin acquisition. Speaking Plastic Mandarin is hence widely accepted in comfort and serves as a covert legitimised professional conduct.

To consolidate my observations, I administered a brief survey about daily language use (see Appendix A.5) to my participants during my fieldwork visits to Changsha in 2020. Nineteen high school students and two undergraduates who graduated from the same school completed the questionnaire. Table 1.2 summarises the number of participants who self-reported that they often or always[4] speak Plastic Mandarin in various scenarios. According to their responses, 20 out of 21 (95%) indicate that they always or often communicate with their classmates and friends in school in Plastic Mandarin, 16 out 21 (76%) indicate that they always or often talk to their teachers in Plastic Mandarin after classes, and 13 out of 21 (62%) indicate that they always or often speak to their teachers in Plastic Mandarin during classes. Such frequent uses of Plastic Mandarin in school, especially with their peers whose native tongues may be mutually unintelligible Xiang varieties, imply that Plastic Mandarin has become the de facto lingua franca in school. Y. Wang (2018) also reported that 36 out of 37 subjects in their study attended secondary schools where the major medium of communication was Plastic Mandarin.

[4]They chose 4 or 5 in a 5-point Likert scale.

From the survey, we also know that outside school these teenagers speak differently depending on the social interactions. 16 out of 21 (76%) indicate that they always or often communicate with their grandparents in one of the Xiang vernaculars, since most of the older generations grew up speaking the vernaculars. When they talk to their parents, Plastic Mandarin, Xiang vernaculars, and a mix of both are the usual options. Standard Mandarin is seldom used in their communication with family. The fact that fewer participants speak to their parents or grandparents in Plastic Mandarin compared to their classmates and friends suggests that Plastic Mandarin be the sociolect of the young. Plastic Mandarin may not be the dominant variety students encounter on university campuses unless their social networks are locally-based, since university admissions are typically a mixed intake of students from different parts of China.

The use of Plastic Mandarin extends beyond the realm of education and co-exists with Standard Mandarin and Changsha in many public spaces. About half the adolescent participants always or often speak Plastic Mandarin in public space too. Jing and Niu (2010) interviewed 42 Changsha locals who are competent 'trilinguals' in Standard Mandarin, Plastic Mandarin and Changsha, and found that for these speakers, Plastic Mandarin was also frequently used in workplace and public sphere. Especially in high-end department stores, restaurants, hotels, post offices, banks, airports, and train stations, more than three-quarters of them preferred to communicate in Plastic Mandarin.

1.1.2.2 From adolescents to wider age groups

Plastic Mandarin is first discernible in the speech of adolescents and young adults. X. Xu et al. (2012) surveyed 104 Changsha residents aged from 15 to 55 years old, and less than 30% of them used Plastic Mandarin in various contexts. The lower usage rate of Plastic Mandarin in their study compared to Jing and Niu (2010) is mostly likely attributed to the lower percentage of young people in the participants (only one secondary school student).

The popular youth speech, Plastic Mandarin, however, is no longer confined to teenagers. The generation born after 1980s and 1990s and proficient in Plastic Mandarin may have been reshaping the linguistic landscape of urban Changsha when they entered the workplace and started families. The presence of Plastic Mandarin was noticeable in the speech of primary school students on the reality television show *Grade One* produced by Hunan Broadcasting System, which was aired in 2014. The show featured seven students, aged 5 and 6, as they embarked on their first semester in a Changsha elementary boarding school. Through 120 cameras the show documented the interactions of these students with their parents, classmates, and teachers. A significant number of young children on the show used Plastic Mandarin, even during interactions with their parents at home or when expressing strong emotions such as crying or losing their temper in conflicts. These children might have acquired Plastic Mandarin from their caregivers and were accustomed to speaking Plastic Mandarin. While it is important to note that the students on the reality television show may not be representative of all young children in Changsha, and that the potential effects of camera lens and editing should be taken into consideration, the show does afford a window on the language practices of young children from relatively well-off families in Changsha.

The use of the vernacular Changsha, unfortunately, has been decreasing. Only 13 out of 21 participants rated 4 or 5 out of 5 in their Changsha proficiency, despite my efforts in recruiting subjects who speak both Changsha and Plastic Mandarin. Further empirical research is needed to document language use in Hunan, but that is not the purpose of this thesis.

1.1.3 Attitudes: A Changsha Mandarin

Through informal conversations with my participants I gathered their language attitudes towards Plastic Mandarin, in the hope of profiling Plastic Mandarin from a socially meaningful perspective. In spite of the negative connotations of the term ‘Plastic Mandarin’, this variety is instead associated with positive traits. Table 1.3 presents the translations of expressions that participants used to describe Plastic

Theme	Translation
Solidarity	"Plastic Mandarin **brings us closer**, reducing our social distance."
	"**Everybody** speaks plastic Mandarin here. If I did not speak it, it would be embarrassing."
	"It's Changsha. It is **inevitable** to speak Plastic Mandarin."
Amiability	"Speaking Plastic Mandarin is **down-to-earth without affectation**."
	"Speaking Plastic Mandarin makes us **affable** and **amiable**."
	"It is a **friendly** and **pleasant** gesture to speak Plastic Mandarin."
	"It depends on my mood. When I'm **happy**, I speak Plastic Mandarin."
Easiness	"There is **no pressure** to speak Plastic Mandarin."
	"I feel **relaxed** and **unconcerned** when I speak Plastic Mandarin."
	"Plastic Mandarin is **easy** and **comfortable** to speak."
	"Speaking Plastic Mandarin is **entertaining** with a sense of **humour**."
	"It is our habit to speak Plastic Mandarin."
Cultural Identity	"I am **proud** of Plastic Mandarin."
	"Plastic Mandarin relates to our **hometown**."
	"Plastic Mandarin **represents Changsha**....is characteristic of **Hunan**."
	"Plastic Mandarin combines Putonghua and our dialect."

Table 1.3: Categorisation of opinions about Plastic Mandarin by participants.

Mandarin[5] and categorises them by their theme. Rather than perceive Plastic Mandarin as a 'bad Mandarin' that needs to be corrected, they have a favourable regard for Plastic Mandarin on the whole.

Language attitudes predominantly reflect evaluative beliefs about speakers of a language variety, typically coalescing along two main dimensions of status and solidarity. While the use of Standard Mandarin tends to be associated with culture, education and high prestige (X. Xu et al., 2012), the use of Plastic Mandarin tends to be associated with maintaining social networks and building identities. As shown in Table 1.3, speaking Plastic Mandarin in the school is a gesture of being friendly and amiable, as well as a sign of solidarity and reduced social distance. It is the

[5]Some of the responses were from my MPhil work (C. Xu, 2018).

norm of communication within the social networks in school. Moreover, Plastic Mandarin has been an inclusive variety that is spoken by both Changsha local and non-local students. As the provincial capital with great educational resources, Changsha attracted many students born in other cities in Hunan province. Often these students are speakers of Xiang varieties other than Changsha, but they all speak Plastic Mandarin in the school. Plastic Mandarin is, in some respect, similar to Multicultural London English, an English variety associated with adolescents, whose linguistic characteristics are related to their multi-ethnic friendship networks at school (Kerswill et al., 2007).

When moving away from the family sphere, adolescents are engaged with constructing identities independently, partially through deploying variables of the vernacular, which contributes to the advancement of linguistic change from below (Eckert, 2017). The adolescents' awareness of the local cultural identity and their positive orientations to the local environment are projected in their choice of Plastic Mandarin as the preferred variety in school. They acknowledged Plastic Mandarin as a marker of their hometown Changsha, or broadly, Hunan (Xiang) province. Through speaking Plastic Mandarin, they were signalling a shared cultural heritage and making more self-conscious regional identity claims.

1.1.4 Contact and acquisition: A new Mandarin dialect

The scholarly discourse on the nature of Plastic Mandarin revolved around two theoretical strands: (1) Plastic Mandarin as a fossilised interlanguage (Jing & Niu, 2010) and (2) Plastic Mandarin as a contact-induced new variety (L. Fu, 2010). The former concept of interlanguage in the field of second language acquisition is specific to individual non-native speakers, while the latter concept in the field of sociolinguistics places the focus on collective linguistic behaviour arising in multilingual communities. Both contact and acquisition are essential factors in the formation of Plastic Mandarin.

Interlanguage refers to the language system that learners develop in the process of acquiring a target language and has a structurally intermediate status between the native language and the target language (Selinker, 1972). The accented Mandarin produced by older adults in Changsha tends to exhibit characteristics of interlanguage, i.e. permeability, dynamism, and systematicity. There is a certain degree of uniformity and a high degree of variability in the accented Mandarin, as evident in the classification of heavy, medium, and light accents in L. Fu (2010)'s examination of accented Mandarin produced by 200 native Changsha speakers aged from 16 to 55. The homogeneous linguistic forms deviant from the norm of Standard Mandarin in Plastic Mandarin, nevertheless, should not be analysed as persistent errors or failed approximations as in the fossilised interlanguage account, since most Plastic Mandarin speakers in this study were aware of the mismatch between Plastic Mandarin and Standard Mandarin and capable of producing more standard forms.

The sociolinguistic perspective of contact and new dialect formation is highly relevant to the case of Plastic Mandarin. In my generation and earlier generations, the norms of Plastic Mandarin were not introduced by our parents nor grandparents. The lack of continuity between generations serves as a marker of a new dialect (Kerswill, 2020). The contact between Changsha and Standard Mandarin in the city of Changsha may not be as apparent, given that there has not been large-scale migration of Standard Mandarin speakers to Changsha in recent years. The contact accompanied the process of adopting Standard Mandarin under the language policy, and was sustained and accelerated in the digital space including mass media and digital communication. Considering Changsha and Standard Mandarin, one might regard Changsha as a case of extended diglossia (Fishman, 1967) with bilingualism, where two varieties coexist in the society but are used in different domains: Standard Mandarin with a highly codified system is the High variety used in written form and formal domains and Changsha is the Low variety marginalised in formal education and thriving in ordinary conversations at home and wet market. The compartmentalisation by usage domains of the two varieties in a diglossic situation is often not clear-cut or stable in reality (Sayahi, 2020), especially when coupled with non-native competence. It appears to be a fluid diglossic situation, or diaglossia in Auer (2011)'s term since the linguistic landscape in Changsha is also filled up with intermediate variants (see §2.2.2 in Chapter 2).

The rise of homogeneous Plastic Mandarin during the contact, similar to the focusing stage of the new variety in Trudgill's three-stage theory of new-dialect formation (Trudgill, 2004), complicates the functional distribution of varieties. Plastic Mandarin tends to encroach upon the High variety in contexts such as education and corporate meetings as well as infiltrate the Low variety in some informal contexts.

1.2 Motivations and Research Questions

The impetus for this work lies in my own Chinese language experience in communicating with peers and friends marked by three stages: I was a fluent Xiang vernacular speaker when I was in a local primary school which most children in the neighbourhood attended; later in the middle and high school in Changsha where a large influx of students from diverse regions in Hunan Province joined for high quality education, we spoke little of our mutually unintelligible vernaculars but mostly conversed in Plastic Mandarin; then during my undergraduate studies in Hong Kong, I spoke Standard Mandarin with friends from various parts of China and started to learn Cantonese. The major shift of my social language was due to the changing geographical environments and social networks, and it also reflects concomitant language change in the society over the years. Now it appears to be the case that younger people in Changsha grew up with Plastic Mandarin, many of whom are no longer fluent in any Xiang vernacular. Similar to the way in which regional "Englishes" have sprouted up all around the world since English has become the global lingua franca, new urban varieties of Mandarin such as Plastic Mandarin may be emerging in other cities too. One motivation of the thesis is therefore to document such language variation since Plastic Mandarin is spoken by an increasing number of speakers, potentially in millions.

Variation studies have concentrated overwhelmingly on segmental phonetic and phonological features. There have also been growing interests in morphosyntactic, grammatical, and discourse features subject to diffusion and levelling (e.g. Cheshire, 2007). Linguistic analyses of contact variation in prosody, nevertheless, remain relatively scarce. Prosody in a spoken language, often decomposed into patterns of timing, melody, intensity, and so on, is indispensable to both the structure and meaning in speech, and conveys a plethora of communicative functions. It encodes phonological contrasts such as lexical tones, as well as indexes speaker identity ranging from gender and age to the sociolectal and dialectal status. Another motivation is to fill a lacuna in our understanding of contact and prosodic variation in tonal languages, through a case study of Plastic Mandarin.

The aim of this thesis is to investigate **the tonal system of Plastic Mandarin**, since little is known about this 'plastic' accent. A few studies such as Jing and Niu (2010), X. Xu et al. (2012), Y. Wang (2018), and S. Shen (2019) examined the citation tones of Plastic Mandarin, but they were merely impressionistic description or based on the fundamental frequency (f_0) measurements of one speaker. None of these studies examined neutral tones in Plastic Mandarin. It is concerned that the traditional approach to tones of Chinese varieties relies heavily on impressionistic data and neglects neutral tones[6]. This thesis provides comparative analysis of Plastic Mandarin and Standard Mandarin, incorporates lexical tones and neutral tone, and investigates the interplay of contextual tonal variation and dialectal variation. Through detailed acoustic-phonetic analyses of tonal and durational patterns of monosyllables and polysyllabic phrases containing neutral tone syllables, this thesis seeks to explore the following main questions:

RQ1. Citation tones: What are the citation tones of Plastic Mandarin? What are the potential sources of sound change?

RQ2. Neutral tone: What are the phonetic patterns of neutral tone in Plastic Mandarin? How do they compare to those of neutral tone in Standard Mandarin?

[6] Standard Mandarin has received extensive attention in tonal scholarship and is comparatively well-researched. This will be introduced in Chapter 2.

RQ3. How do full lexical tones and neutral tone interact with the Standard-Plastic language variation?

The thesis contributes to the documentation of a new Mandarin variety, the analysis of prosodic features in language contact and variation, and the discussion on theoretical representation of Mandarin neutral tone. Methodologically, this thesis demonstrates two mixed-model approaches to time series f_0 data: functional data analysis combining Legendre polynomial modelling with linear mixed effects models, and generalised additive mixed models. The former with the advantage of interpretable coefficients can be a valuable tool in establishing tonal space and clustering citation tone patterns in comparison; the latter with the flexibility of inclusion of time variable in dynamic phonetic patterns can reveal how varying conditions differ with respect to how f_0 unfolds over time.

1.3 Thesis Organisation

Chapter 2 provides the necessary background of this study and reviews previous work related to the production, representation, and modelling of tones and tone sequences. It begins with an introduction of Chinese as a constellation of different varieties and then situates the study in the social context of the popularisation of Standard Mandarin, and in the geographical context where the city of Changsha, its speakers, and its accents are portrayed. Then it continues into the academic area of tones, ranging from the historical processes to computational modelling of surface realisation. In particular, neutral tone, one of the main subjects of investigation, is introduced.

Chapters 3 to 5 present three studies that make up the empirical core of the thesis. As part of my research, I spent time in a middle school in Changsha collecting speech data from adolescent speakers of Plastic Mandarin. Comparative Standard Mandarin data were also collected from Mandarin speakers in Oxford during the period of the COVID travel restrictions. These recordings constitute the main speech data for this study.

Chapter 3 focuses on the citation tones, using the first-hand speech data of three varieties that coexist in the city of Changsha: Plastic Mandarin, Changsha, and Standard Mandarin. The chapter provides a thorough description of the data collection procedure and highlights the importance of fieldwork methods in eliciting a non-standard variety. The data show that Plastic Mandarin exhibits the same citation tone categories as Standard Mandarin, but with each tone category featuring a distinct pitch pattern compared to that of Standard Mandarin. The duration of citation tone syllables in Plastic Mandarin also differs from that in Standard Mandarin, especially Tone 3 syllables. The chapter summarises the prototypical pitch contour for each citation tone in Plastic Mandarin. Through a comparative analyses of citation tones in the three varieties, this chapter provides acoustic evidence that Plastic Mandarin arose by adapting some pitch patterns of Changsha tones into the corresponding Mandarin tone categories.

Chapter 4 presents a subsidiary investigation that involves the presence and realisation of neutral tone in Standard Mandarin using corpus data, so as to establish a benchmark for comparison with Plastic Mandarin. It explores the acoustics of thousands of phrases containing neutral tone syllables in a corpus of Standard Mandarin Broadcast News Speech, since most previous studies of neutral tone are either phonological discussion based on impressionistic pitch heights or acoustic studies over a few types of phrases in lab read speech. Despite the absence of well-controlled linguistic contexts, there were strong statistical correlations between the preceding lexical tone and the pitch contour shape of neutral tone in the corpus data. The analysis shows that neutral tone in Standard Mandarin news speech tends to converge towards a low pitch in the course of two consecutive neutral tone syllables, and suggests a probabilistic target approaching view of tones.

Chapter 5 examines the variability in the phonetic realisation of neutral tones, through phrases containing neutral tone syllables in the fieldwork data of semi-spontaneous speech in both Mandarin varieties. It probes into the extent of the influence of the preceding lexical tone, the following tone, and overall duration on f_0 patterns of neutral tone. The analysis shows extensive carryover effect from the preceding lexical tone but little effect from the following tone in both Mandarin varieties. It reveals how duration affects f_0 contours and manifests the important role of syllable duration in the f_0 approximation to a pitch target. Despite differences in the realisation of neutral tone pitch contours between the two Mandarin varieties, Plastic Mandarin resembled Standard Mandarin in the systematic approximation of a low pitch target (not necessarily the lowest pitch) among various neutral tone sequences containing one or more consecutive neutral tones. Furthermore, the pitch target of neutral tone tends to be the same for both Mandarin varieties.

Chapter 6 revisits the research questions, summarises, and contextualises the key findings from the previous chapters. It explores the possible mechanisms that may have shaped the Plastic Mandarin tone system and accounts of neutral tone representation and realisation in both Mandarin varieties. It also extends the discussion on prosody in language variation and change, methods of analysing phonetic data, potential directions of future research, and lastly, open science and reproducible practices.

2

Background and Literature

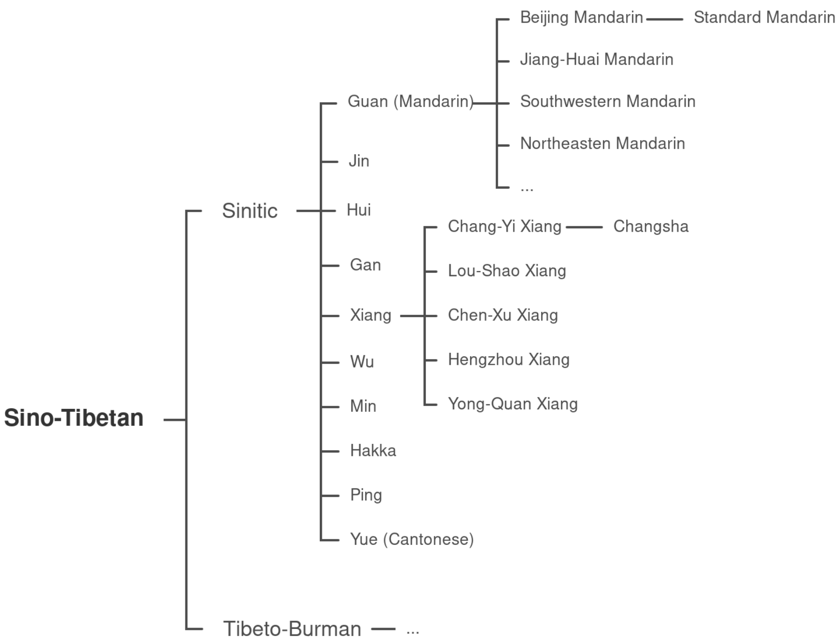
Contents

This chapter delineates the social, geographical, and linguistic context for our investigation of Plastic Mandarin, and reviews literature and theories pertaining to the phonetics and phonology of tones in Chinese languages. It starts by a brief introduction to Chinese languages and the rise of Standard Mandarin in §2.1, before focusing in on the Changsha city, the local community, and its accents in §2.2. Then §2.3 gives a sketch of the historical and phonological studies of lexical tones of Chinese languages, and defines a few key terms used in the thesis. §2.4 moves on to tones in contexts and introduces speech models of phonetic realisation of tones. §2.5 focuses on neutral tone, ranging from its properties and categorisation to existing analyses of its realisation.

2.1 Kaleidoscope of Chinese Varieties and the Lingua Franca

Chinese is a branch of the Sino-Tibetan language family. Typologically, Chinese is a lexical tone language, as well as an isolating or analytic language, with a logographic writing system. In other words, a morpheme in Chinese corresponds to a syllable in sound structure and a character in writing. Each stressed syllable has a particular pitch pattern, which differentiates the meaning of a morpheme. Instead of having extensive affixes, grammatical relations are indicated mainly by word order and morpheme particles. Chinese is sometimes classified as monosyllabic since the majority of morphemes, free or bound, are represented by single syllables throughout stages of development (Norman, 1988). The majority of words, though, are disyllabic in modern Chinese (P. C. Yip, 2000).

Chinese is often renowned for having the largest number of native speakers in the world, although this estimate depends on how exactly 'Chinese' is defined. Chinese is a language complex consisting of many disparate varieties. Beijing and Shanghai, for example, in a well-conceived analogy drawn by Kratochvíl (1968) and Chao (1968), may sound as divergent from each other as Portuguese from French in Romance languages. Linguists categorised Chinese into, mostly seven to ten groups, still with controversies. Yuan (1960) proposed seven groups: Mandarin (Northern), Wu, Xiang, Gan, Hakka, Yue (Cantonese), and Min. Later, Jin (separated out from Mandarin), Hui, and Ping were added to the list by *Language Atlas of China* (1987). Figure 2.1 specifies the ten groups of Chinese and the sub-groups of Mandarin and Xiang, so as to reveal the genetic link between Standard Mandarin and Changsha.



Source: Based on Matisoff (1991), Institute of Linguistics CASS et al. (2012), and H. X. Bao and Chen (2005).

Figure 2.1: Standard Mandarin and Changsha in the Chinese language family. Other nodes are not expanded.

The diverse Chinese varieties seem not to fit the common western perception of the terms ‘language’ and ‘dialect’: a plethora of regional varieties, unified by the common logographic writing system, are largely mutually unintelligible and phonologically distinct. Chinese scholarship (e.g. Chao, 1968; Yiu, 2013) has often adopted ‘dialects’ as the translation to the Chinese term *fāngyán* (方言) when referring to these varieties, which involves considerations of deeply-ingrained bonding history and culture of various speech communities, as well as a political undertone of unity. DeFrancis (1986) coined a term ‘regionalects’ to describe those mutually unintelligible varieties. This study chiefly adopts an all-encompassing term ‘**variety**’ for members in the Chinese language family, highlighting the very nature of variation and change.

Apart from the mosaic of Chinese varieties, up to 130 non-Chinese languages from other language families, mainly spoken by the minorities, are officially recognised in China, according to the newly updated Language Atlas of China (Institute of Linguistics CASS et al., 2012). Given the sheer linguistic diversity in the vast territory, the practice of designating a standard language dates back to the Western Zhou Dynasty, 1046-256 BC (Kurpaska, 2010), and has been exerting influence on the evolution of languages. In Ming and Qing dynasties (1368-1912), *guānhuà*[1], literally the speech of officials or bureaucrats, was the standard language to administer the empire.

At present, *pǔtōnghuà*, 'common tongue', also known as Standard Chinese or Standard Mandarin, has been declared as the common language of the People's Republic of China by law[2] since 2001, which is a continuation and consolidation of its status as the national language since 1930s in the years of Republic of China in Mainland China[3]. It is formed on the basis of Beijing Mandarin and Northern (Mandarin) dialects, and its popularisation has been one of the main goals of the Chinese language policy (Kurpaska, 2010). Standard Mandarin serving as a domestic lingua franca indeed seems indispensable in an era of extreme information and human mobility.

Legislation has enforced the use of Standard Mandarin in formal public domains including government administration, media, and education. This highly codified variety has thus been regarded as the esteemed tongue with high social prestige and great career prospects. The availability of mass education has effectively added this standard variety to the repertoire of a substantial proportion of the population which hitherto may have only spoken the vernacular. Nowadays, Mandarin has gained wide currency in China, where over 80%of the population speak Mandarin (E. Zhao & Wu, 2020). Most urban communities are at least diglossic, in which the majority of young generations are native bilinguals[4] of Standard Mandarin and their vernaculars (J. Wu, 2015). In reality though, Mandarin in local operations is very likely to be divergent, to greater or lesser extent, from the form used in the capital Beijing.

[1] In Qing dynasty *guānhuà* referred to Manchu and imperial Mandarin.

[2] *Zhonghua Renmin Gongheguo Guojia Tongyong Yuyan Wenzi Fa* 《中华人民共和国国家通用语言文字法》（PRC's Law of Common Language and Writing)

[3] It was called *Guóyǔ*, 'national language', and this term is still used in Taiwan.

[4] Many have studied English through schools at an early age too.

As Standard Mandarin progressively spreads under the vigorous policy in recent years, it has been overshadowing local varieties in terms of social functions: the use of the latter has been decreasing, especially among the young generations. Discourses of endangerment of Chinese varieties have emerged (J. Zhang, 2014) along with an increasing global concern over language death. Meanwhile, the contact and interaction between Standard Mandarin and vernaculars leads to language variation and change. New non-standard Mandarin varieties across various regions are observed, among which Plastic Mandarin has become prominent and evolved into a stable variety.

2.2 The City of Changsha and its Accents

Plastic Mandarin is mainly spoken in Changsha, Hunan Province, China. The rise of Plastic Mandarin is concomitant with the rapid urbanisation and economic growth of Changsha, and attributed to the Mandarin education and media. The emergence of this new variety can be considered a case of language contact and change in the context of a metropolis, and in this process the role of adolescents tend to be central. Another example of such language variation is Multicultural London English, although its development rooted in widespread migration and integration of diverse ethnic and cultural backgrounds is very different from Plastic Mandarin.

2.2.1 Changsha, Hunan

Changsha is the capital of Hunan province, in the central-south of China (see the map in Figure 2.2). Situated on the Xiang river valley plain, it is bounded by mountains on the east, west and south, and borders Dongting lake on the north. The Xiang River, one of the major tributaries of Yangtze river, runs from the north to the south throughout Changsha.

Historically, Changsha has often been a regional centre for administration, transportation, commerce, and culture. For example, it was the capital of Changsha State in the Han Dynasty (206BC - 220AD), and the capital of Chu State (907 - 951) in Ten Kingdoms period. Nowadays covered by extensive bus and rail networks,

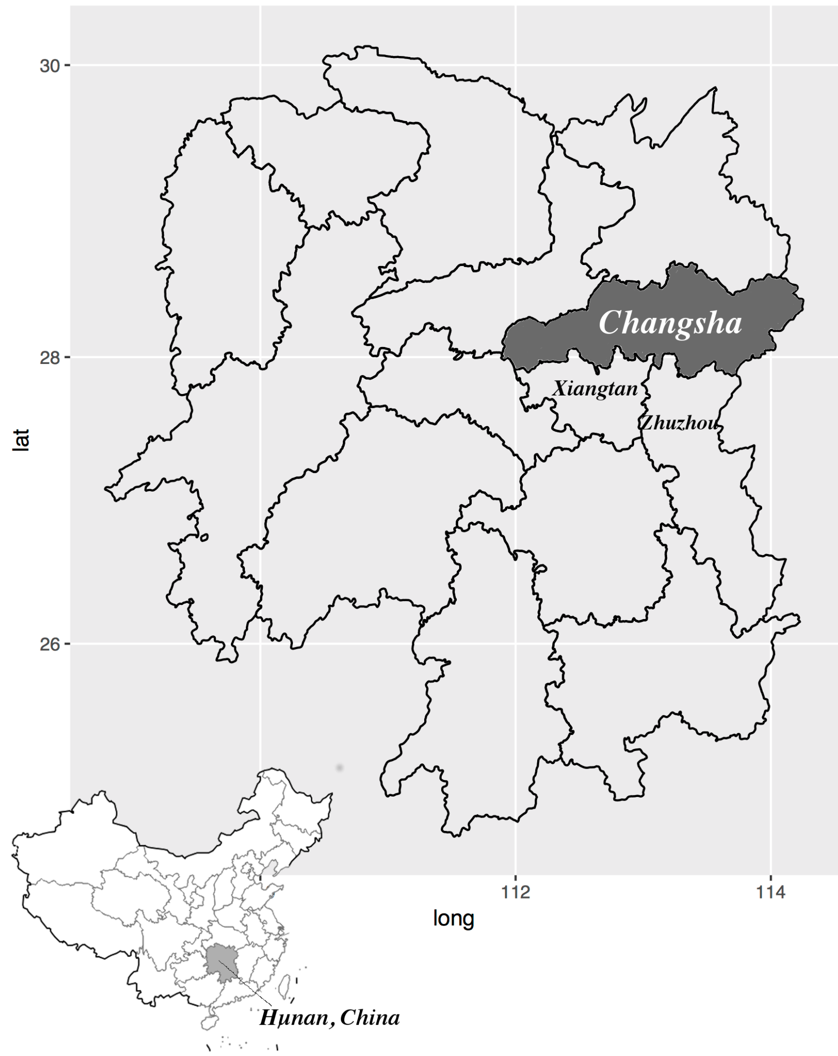


Figure 2.2: The location of Changsha, Hunan Province, China.

waterways, and air travel, Changsha is well-known for its service sector, mechanical engineering industry, as well as media and communications industry.

Changsha and another two bordered cities on the south Zhuzhou and Xiangtan (also marked on Figure 2.2) constitute the Greater Changsha Metropolitan Region (also known as Chang-Zhu-Tan Megalopolis). Since 2009, the communications and information networks of the three cities were integrated and the intercity high-speed rail was built in 2016. To give a concrete example, it takes only 12 minutes to arrive in Changsha from Xiangtan by the high-speed rail. These integration policies have greatly extended the urban area in the three cities. Such metropolitan region is highly likely to lead language change, as urbanisation may create new social-geographical space through reestablishment of communities and migration flows. J. Zhang (2014) observes that China's urbanisation process is geographically unbalanced and mostly affects the Chinese-dialect-speaking regions.

Among a residential population of more than 10 million in Changsha in the year of 2021, about 5.98 million live in the inner-city consisting of six districts[5] on both banks of the Xiang river (Hunan Provincial Bureau of Statistics, 2021). In Changsha, the urban population has grown 42.71 % since 2010 (Changsha Statistics Office, 2021). Changsha is the only city in Hunan Province that has positive net migration rate over the years (Hunan Provincial Bureau of Statistics, 2017), and the great majority of migrants are from other regions within the province, thanks to its impressive economic growth in recent years. In 2021, Changsha's GDP was 1.3 trillion RMB (208 billion USD). A surge of migrants in the labour market yields increasing complexity of the local linguistic repertoire, since there are numerous linguistic varieties within Hunan, many of which are mutually unintelligible.

Apart from work and business opportunities, the high quality education and medical resources of Changsha also attract migrants. Being the capital, it provides the best education in the province. Many students, especially top students from other cities or villages, move to Changsha to attend schools, where leading educational resources are available in Mandarin.

2.2.2 Changsha Xiang, Standard Mandarin, and everything in between

Changsha is the Chinese variety predominantly spoken in the city of Changsha and it is not mutually intelligible with Standard Mandarin. As shown in the previous tree diagram in Figure 2.1 and the map of language groups in Figure 2.3, Changsha is affiliated with Chang-yi subgroup (H. X. Bao & Chen, 2005), also known as New Xiang subgroup, under the Xiang group (Yuan, 1960).

[5]These are Furong district, Tianxin district, Yuelu district, Kaifu district, Yuhua district, and Wangcheng district.

The classification of language groups by dialectologists is mostly based on the shared sets of phonetic, phonological, lexical, and syntactic features of modern Chinese varieties in relation to Middle Chinese (circa AD600), the recorded historical standard variety. Most Xiang varieties have non-aspirated stops and affricates (Norman, 1988), among which the Chang-yi subgroup has only voiceless stops and affricates as syllable onsets (H. X. Bao & Chen, 2005) because the voicing contrast originally present in Middle Chinese has been lost.

From Figure 2.3, we can see that greater linguistic heterogeneity is concentrated in southern part of China, and Xiang is in the transitional zone between the northern (Mandarin and Jin) and southern groups (Xiang, Gan, Hui, Wu, Min, Hakka, Yue, and Ping). Three out of four sides of Xiang borders Mandarin, and the other neighbour is Gan. Xiang, especially the Chang-yi subgroup in northern Xiang area, is hence much influenced by Mandarin. According to Y. Wu (2005), Xiang contains a melange of northern and southern features, as well as classical and modern characteristics in terms of grammatical constructions. Since pitch dynamics is most relevant to this study, the next section (§2.3.2) will provide a more detailed introduction to the tones of Changsha.

Changsha, though, is not the only Chinese variety in the city. There are other Chang-yi Xiang varieties such as Ningxiang spoken in Ningxiang in western Changsha, and Liuyang spoken in Liuyang [6] in eastern Changsha. Liuyang is also home to some Hakka speakers and Gan speakers. The integration of three cities in the Greater Changsha Metropolitan Region, i.e. Changsha, Zhuzhou, and Xiangtan, has also brought speakers of different Xiang varieties into Changsha.

Another Chinese variety in the city of Changsha is Standard Mandarin. While Standard Mandarin is officially required in radios, television, films, lectures, formal meetings, and other formal public situations, Changsha has been dominating the communication of everyday life, and the local opera *Huāgǔxì*, the Flower Drum Opera. In real practices, language use in formal and informal domains is not clear-cut, and the level of Standard Mandarin proficiency of people in Changsha varies

[6] Varieties are named after the place. Both Ningxiang and Liuyang are county-level cities under the administration of the prefecture-level city of Changsha.

Source: Adapted from two source maps in Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Map_of_sinitic_languages_full-en.svg and https://commons.wikimedia.org/wiki/File:Classification_of_Xiang_2005_Bao.png(retrieved on 15 March 2018).

Figure 2.3: Geographical Distribution of Xiang Chinese and Chang-yi Xiang subgroup. Different shades of grey colour indicate different language groups. Some major cities are marked in black dots. The background map is cropped due to the size of the figure. The overlaying map at the bottom left corner zooms into Hunan Province, on which the shaded area roughly corresponds to Xiang region on the background map.

greatly. Television programmes and series in Changsha or other Xiang varieties are sometimes broadcast in Hunan Province, mainly through provincial television channels. Non-standard Mandarin accents in Hunan, despite its occasional use in the media, appears to be widely recognised across Mainland China. This is probably due to the exposure from social media and from the nationwide Hunan Satellite Television, which is currently the second most-watched channel in China, second to CCTV1 by China Central Television.

The scope and definition of the categories of local varieties and accents can be vague and vary in the literature. Y. Wu (2005, p. 22) identified three spoken codes of Changsha dialect: Spoken Changsha, Reading Changsha, and Plastic Mandarin. In her description, Spoken Changsha refers to Changsha used in normal daily conversations, Reading Changsha preserves the conventional and literary pronunciation for reading out aloud, dictating, or certain oral performances, and Plastic Mandarin describes "a variety that combines the features of the local Changsha dialect with the features of standard Mandarin". This was one of the very first appearances of the term 'Plastic Mandarin' in academic work to my knowledge. Y. Wu (2005) emphasised that the most salient feature of Plastic Mandarin pronunciation is the use of a more Standard Mandarin-like tonal system with Changsha segmental phonology. It is worth mentioning that Y. Wu (2005)'s definition of Plastic Mandarin is **not** adopted in this thesis. In my observation, Plastic Mandarin now spoken by the young generations has evolved into a variety that has a very similar segmental inventory to that of Standard Mandarin but features distinct tonal patterns. Her categorisation of Spoken Changsha and Reading Changsha highlights the important role of formality or mode of communication in speech, and sheds light on considerate methods of elicitation in fieldwork.

The language situation on the Standard–vernacular dimension in the city of Changsha can indeed be summarised as a spectrum of varieties blending elements of Xiang vernaculars (mainly Changsha) and Standard Mandarin in varying degrees, schematised in a continuum in Figure 2.4. The interaction between Xiang vernaculars and Standard Mandarin takes place on multiple layers of the language structure.

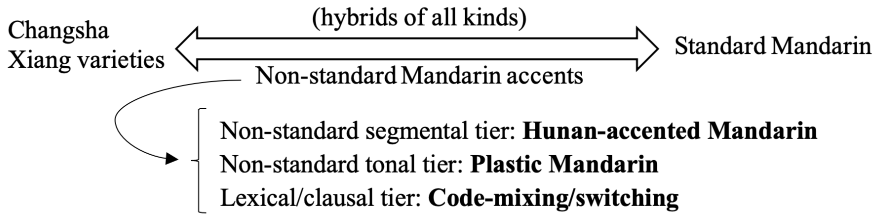


Figure 2.4: Continuum of Oral Chinese Repertoire of Changsha Speakers.

On the individual level, an idiosyncratic speech may contain an eclectic mix of Standard and vernacular characteristics in multiple tiers. Factors such as age, education, and occupation largely determine the proficiency level of Standard Mandarin spoken by Changsha speakers. The self-claimed Standard Mandarin spoken by an average middle-aged Changsha speaker, for example, is highly likely to fall within the middle ground of the continuum, while the elderly is highly likely to speak a variant on the continuum close to the end of Changsha or even only Changsha. It is typically a different case for the young generations, who have the capacity of commanding a wider range of varieties on the continuum and acquired a high level of Standard Mandarin due to their early exposure to omnipresent Standard Mandarin and their Standard Mandarin education since kindergarten. They can usually switch among speech styles or accents along the continuum.

On the group level, it is thus possible to identify and propose two broad categories of non-standard Mandarin produced by Changsha speakers: Hunan-accented Mandarin and Plastic Mandarin, depending on whether their accent adopts Standard Mandarin tone patterns.

Hunan-accented Mandarin is characterised by Standard Mandarin tone patterns in addition to perceptual and production differences in segmental phonology. For example, some Changsha speakers are not perceptually sensitive to the distinction between /ʂ/ and /s/ in Mandarin, a non-native contrast in Xiang vernaculars including Changsha, similar to the phenomenon that native Japanese speakers hardly perceptually differentiate English /r/ and /l/. This is illustrated in two Hunan-accented Mandarin examples [1] *shī shēng* and [2] *dú shū* by a Changsha speaker taken from a transcript in Y. Wu (2005, p. 29). The corresponding

[1]	师生
	‘teacher(s) and student(s)’
	shī shēng
Standard Mandarin	/ʂʐ̩˥ ʂəŋ˥/
Changsha	[sz̩˧ sən˧]
Hunan-accented Mandarin	[sz̩˥ sən˥]
Plastic Mandarin	[ʂʐ̩˧ ʂəŋ˧]

[2]	读书
	‘read book(s)’
	dú shū
Standard Mandarin	/tu˧˥ ʂu˥/
Changsha	[təʊ˨˦ ɕy˧]
Hunan-accented Mandarin	[tu˧˥ ɕy˥]
Plastic Mandarin	[tu˨˧ ʂu˧]

Standard Mandarin, Changsha, and Plastic Mandarin pronunciations are provided for comparison[7]. [1] and [2] demonstrate that the Changsha speaker employs either alveolar fricative [s] or alveolo-palatal fricative [ɕ], the acoustically similar segments that are available in vernacular Changsha, as alternatives to the retroflex fricative [ʂ] that exists in Standard Mandarin, whereas keeping the tones identical to Standard Mandarin. Other Mandarin phoneme pairs including syllable onsets /n/-/l/, /ʈʂ/-/ts/, /ʈʂʰ/-/tsʰ/, /x/-/f/, and syllable coda /n/-/ŋ/[8] are often not distinguished by these Changsha speakers. Hunan-accented Mandarin varies among speakers, dependent on their levels of proficiency in Standard Mandarin. Current middle-aged generations (born in the 1960s and 1970s) tend to speak such Hunan-accented Mandarin from my observations.

Plastic Mandarin in this study, as introduced in §1.1, refers to the Mandarin variety with distinct tonal patterns, spoken by the young generations who were exposed to Standard Mandarin from an early age. It is also illustrated in the

[7]The transcription of Changsha and Plastic Mandarin equivalents was based on my pronunciation.

[8]The preceding vowel is often /i/ or /ə/.

previous examples [1] and [2], where the segments in Plastic Mandarin are the same as those in Standard Mandarin but the tones differ. Some native Standard Mandarin participants stated in our debriefing conversations that the frequent pitch rises in Plastic Mandarin made it discernibly distinct from Standard Mandarin and drew their attention to this accent. Plastic Mandarin can easily and systematically adapts new words and expressions.

According to Trudgill (1986), the levelling mechanism, that is, the reduction or attrition of differences between varieties, is omnipresent in the processes of dialect contacts, and operates on both cross-dialectal and standard-dialect dimensions (Hinskens, 1998). Both Hunan-accented Mandarin and Plastic Mandarin may be considered as levelled forms, converging towards the standard variety in not only lexical and syntactic domains, but also segmental and tonal domains.

2.3 Chinese Tones in the Past and Present

Chinese varieties vary considerably in their tonal system and most contain three to ten contrastive tones. But first, what is a tone? In general terms, tone is the linguistic use of pitch and voice quality in marking lexical items and the primary acoustic correlate is the fundamental frequency (f_0), the frequency of the vocal fold vibration jointly determined by the subglottal pressure (Titze, 1989) and the tension of the vocal folds. The term **pitch** in the field of speech perception often refers to the perceived or subjective experience of the tonal height of a signal (Reetz, 1996). In this thesis, pitch is mostly used interchangeably with the **log-transformed** f_0.

The tonal system of a Chinese variety consists of two parts: **tone categories** and **tone patterns**. It is ambiguous when expressions such as 'varieties X and Y share the same tone' are used, without specifying whether it is the tone category (phonological grouping) or the tone pattern (phonetic realisation) or both that is shared. Hence I would like to propose a bag-of-characters (or morphemes) metaphor to facilitate our understanding of tonal systems and tonal processes. We can conceive each tone category as a bag (set) of Chinese characters or morphemes, and this is illustrated in Figure 2.5. Each bag usually features a unique pitch pattern,

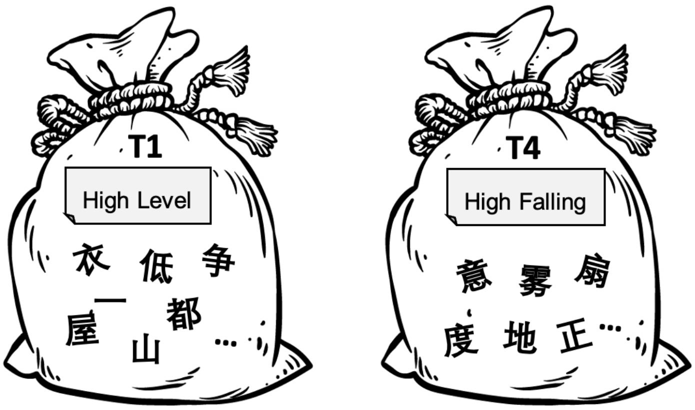


Figure 2.5: The bag-of-characters/morphemes metaphor for a tonal system.

indicated by the label on the bag (e.g. high falling), and all characters within a bag have the same pitch pattern. In tonal analysis, arbitrary numbers are often assigned to each bag to index the set of morphemes, which are referred to as tone category. Under this metaphor framework, tonal changes such as mergers and splits can be represented as repacking the morphemes in decreasing or increasing number of bags and removing or creating labels accordingly. Another type of tonal change lies in the variation of the tone pattern of a specific set of morphemes and can be represented as replacing the label for a bag, so that the citation tone of all morphemes inside this bag is consistently changed. In the literature, when two varieties have a shared tone category, it often means that each variety has a bag in which the majority of the morphemes match those in the other. In other words, the varieties contain a similar or identical set of morphemes in one of their respective tone categories (bags). It is often the case that these two tone categories evolved from the same historical tone category.

In addition to pitch pattern, phonation such as voicing, aspiration, creaky and breathy voices may play a role in the tone category as well as the tone pattern. An elementary knowledge of historical Chinese phonology is crucial to the understanding of tonal systems in modern varieties, thus §2.3.1 begins by introducing the tonal categories established in Middle Chinese.

2.3.1 Historical tonal category

It is well established in the historical linguistic literature that the tone systems of modern-day Chinese varieties evolved from a four-by-two system, as shown in Table 2.1, which has been employed as a reference frame that facilitates diachronic and synchronic comparison.

Index	Categories	Translation	Derived tones	
			voiceless onset (a)	voiced onset (b)
I	平 *píng*	'level'	*yīn píng*	*yáng píng*
II	上 *shǎng*	'rising'	*yīn shǎng*	*yáng shǎng*
III	去 *qù*	'departing'	*yīn qù*	*yáng qù*
IV	入 *rù*	'entering'	*yīn rù*	*yáng rù*

Table 2.1: Summary of Middle Chinese tone categories.

The four primary tone categories (labelled in Roman numerals), *ping* (I), *shang* (II), *qu* (III), *ru* (IV), were from Middle Chinese, a historical Chinese variety recorded in the landmark *qiēyùn* dictionary compiled in AD601 (E. G. Pulleyblank, 1991). Several attempts such as Karlgren (1922), Tsu-lin (1970), and Ting (1996) have been made to reconstruct the Middle Chinese tonal values, yet their results were not in accord. Some properties of these historical tones can be inferred from their classification. Except the *píng* or level tone (I), others (II to IV) are contour tones and classified as 'oblique tones' (*zè shēng*). There is also the robust dichotomy of 'smooth' or 'checked' syllables. The *rù* or entering tone (IV) in Middle Chinese is also called the 'checked tone' because it occurs exclusively with 'checked syllables' that end in an occlusive coda containing /p, t, k/, which are often not released [p̚, t̚, k̚] or reduced to a glottal stop [ʔ] (M. Y. Chen, 2000). The *rù* tone is often regarded to be relatively short and abrupt in reconstruction (e.g. Ting, 1996). Other tones (I to III) occur with 'non-checked (smooth) syllables' that are either an open syllable or closed by a nasal stop (or possibly a non-glottalised sonorant consonant) in Middle Chinese (M. Y. Chen, 2000). Such dichotomy is robust not only in Middle Chinese but across many other tone languages in Southeast Asia. Nevertheless, checked syllables have been lost in many contemporary Mandarin varieties including Standard Mandarin and Changsha.

In the field of tonogenesis, it is widely accepted that tonal contrasts most commonly derive either from the loss of laryngeal segments such as [ʔ] and [h] or from phonation variation including voicing, aspiration, breathiness, and creakiness (Hyman, 2018). According to the theory of Haudricourt (1954) and E. G. Pulleyblank (1978), the Middle Chinese tone categories arise from different syllable structures in Archaic Chinese. Middle Chinese tones have undergone assorted 'splits' and 'mergers' in forming the modern varieties, normally conditioned on certain phonological features at certain positions in a syllable. Many diachronic studies illuminate how syllable structure and segmental features interact with tonal development (e.g. Thurgood, 2002; E. G. Pulleyblank, 1991).

The most notable case is that the voicing contrast in the syllable onset leads to tone splits. Each of the four primary tonal categories was first split into two non-phonemic high-low registers, known as *yīn* (a) and *yáng* (b) in the traditional nomenclature, subsequently yielding the primitive eight-tone system in Table 2.1. The *yīn* register is associated with voiceless onsets and has a higher pitch than the *yáng* register associated with the voiced counterpart. The partition into *yīn* and *yáng* registers is highly likely to be phonetically motivated (Hombert et al., 1979) and related to the intrinsic pitch variations of consonant onset pairs differentiated by voicing. During the split, the categorisation of syllables with sonorant onsets including nasals, liquids, and glides in the absence of their voiceless counterparts and null-onset syllables is not as neat and clear-cut as syllables with obstruent onsets, in that the former may pattern with voiced obstruents in the *yáng* register (e.g. Changsha), or with voiceless obstruents in the *yīn* register (e.g. Standard Mandarin). The categorisation of syllables with sonorant and null onsets vary across tonal categories as well as different varieties. Subsequent voicing loss or devoicing of the onsets in some varieties leads to phonemic register distinction (Norman, 1988).

Major mergers in the development of many modern varieties involve *yáng shǎng* (tone IIb) splintered off from tone II and melded with tone III (e.g. in both Standard Mandarin and Changsha), and the redistribution of *rù* (IV) tone, the checked tone, such that the whole category IV disappears and its elements are merged with other categories due to the weakening or loss of occlusive coda (e.g. in Standard Mandarin).

2.3.2 Tonal system of Standard Mandarin and Changsha in literature

This section introduces the tone systems of Standard Mandarin and Changsha in the literature, and summarises the correspondence relations between lexical tone categories of Standard Mandarin and Changsha. While the tones of Standard Mandarin are well-established and studied (e.g. Lee & Zee, 2003), the phonetic description of Changsha tones is based on the seminal work of Firth and Rogers (1937), Beijing University (1995), Y. Wu (2005), and H. X. Bao and Chen (2005).

Tables 2.2 and 2.3 display the tones of Standard Mandarin and Changsha respectively. The phonetic description of tones adopts both the tone letters and digits that were devised by (Chao, 1930), since this system has been extensively used in the literature on tones of sinitic languages. In Chao (1930)'s numeric tone stave system, the tonal range is idealised as a five-point scale on which 1 represents the lowest pitch and 5 the highest, and the juxtaposition of digits from left to right signals the tone contour: increasing digits such as 35 indicates rising tone and decreasing digits such as 21 indicates falling tone. The tone categories of modern Chinese varieties are labelled in Arabic numerals (column 'Tone Index') for easy reference. These numbers are arbitrary but follow the convention in the literature.

2.3.2.1 Correspondence relations between lexical tone categories

In Tables 2.2 and 2.3, the column 'Historical Category' traces the Middle Chinese tone category from which the tone category of the modern variety evolves. This indicates that the majority of morphemes in the modern tone category belong to the corresponding historical category in Middle Chinese. The tone categories in Changsha and Standard Mandarin that evolved from the same Middle Chinese

Tone Index	Historical Category	Phonetic Description		Example	Phonology
		Lee and Zee (2003)		[ma]	Duanmu (2000a)
1	*yīn píng* (Ia)	˥ 55	high level	妈 'mother' 抹 'wipe'	H
2	*yáng píng* (Ib)	˧˥ 35	mid to high rising	麻 'hemp'	LH
3	*yīn shǎng* (IIa)	˨˩˦ 214 ˨˩ 21	low dipping low falling	马 'horse'	L
4	*qù* (III)	˥˩ 51	high falling	骂 'scold'	HL

Table 2.2: Lexical tone system of Standard Mandarin.

Tone Index	Historical Category	Phonetic Description		Example	
		R. Li (1998)		[ma](*[pʰa])	[i]
1	*yīn píng* (Ia)	˧ 33	mid level	妈 'mother'	衣 'clothes'
2	*yáng píng* (Ib)	˩˧ 13	low to mid rising	麻 'hemp'	姨 'aunt'
3	*yīn shǎng* (IIa)	˦˩ 41	high falling	马 'horse'	蚁 'ant'
4	*yīn qù* (IIIa)	˦˥ 45	high rising	怕* 'fear'	意 'meaning'
5	*yáng qù* (IIIb)	˨˩ 21	low falling	骂 'scold'	易 'easy'
6	*rù* (IV)	˨˦ 24	mid rising	抹 'wipe'	一 'one'

Table 2.3: Lexical tone system of Changsha.

category are largely labelled with the same number to indicate their connection. For instance, Tone 2 in Standard Mandarin and Tone 2 in Changsha are both evolved from the *yáng píng* (Ib) category of Middle Chinese.

Each tone type in Standard Mandarin is demonstrated with at least one [ma] syllable. An equivalent set of [ma] syllables is also included in Changsha as examples, though [ma] in Tone 4 is not available. An extra set of [i] syllables is, therefore, included. The tonotactic gap in the [ma] set reflects that Changsha evolved from the 'register distinction' of *qù* (III) tone in Middle Chinese. In other words, Tone 4 in Changsha usually co-occur with syllables with a voiceless obstruent onsets. In Table 2.3, [pʰa] with a voiceless bilabial stop onset is employed instead to illustrate

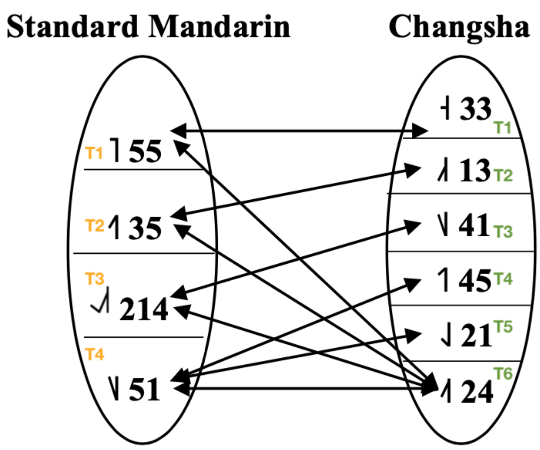


Figure 2.6: Correspondence relation between the lexical tone categories of Standard Mandarin and Changsha.

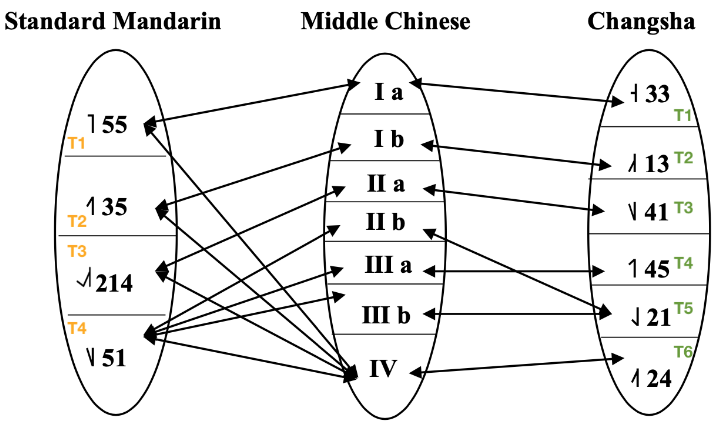


Figure 2.7: Historical correspondence relation between the lexical tone categories of Standard Mandarin and Changsha through Middle Chinese.

the *yīn qù* (IIIa) tone in Changsha. In Standard Mandarin, Tone 3 has two variants without considering tone sandhi: a full form ˨˩˦ and a reduced form ˨˩, although the full form is the prescriptive form in textbooks. The full form is typically reduced to a low falling tone without the rising tail in non-utterance-final positions (Chao, 1968).

The morphemes ‘wipe’ (抹) and ‘mother’ (妈) are affiliated with different tones in Changsha (T6 and T1 in Table 2.3) yet the same tone in Standard Mandarin (T1 in Table 2.2). In fact, the correspondence relation between the lexical tones of Standard Mandarin and Changsha is complicated, as schematised in Figure 2.6. Characters or morphemes in the Tone 4 category [˥˩] in Standard Mandarin, for instance, may bear Tone 4 [˦˥], Tone 5 [˨˩], or Tone 6 [˨˦] in Changsha.

Standard Mandarin	Middle Chinese	Changsha
˥ 55[a], ˧˥ 35[a]	píng (I)	˧ 33[a], ˩˧ 13[a]
˨˩˦ 214[ae], ˥˩ 51[ab]	shǎng (II)	˦˩ 41[a], ˨˩ 21[abh]
˥˩ 51[bf]	qù (III)	˦˥ 45[i], ˨˩ 21[bhi]
˥ 55[dg], ˧˥ 35[dg], ˨˩˦ 214[dg], ˥˩ 51[dg]	rù (IV)	˨˦ 24[jd]

[a] Phonemic register split of tone I and II

[b] Tone IIb merges with tone III

[c] Devoicing and, in some cases, aspiration of voiced obstruents

[d] Loss of obstruent codas

[e] Sonorant onsets in tone II pattern with voiceless obstruents as IIa

[f] Tone III remains as a single cohesive tonal category

[g] Redistribution of tone IV among other tonal categories

[h] Sonorant onsets in tone II, III pattern with voiced obstruents as IIb, IIIb

[i] Register split of tone III

[j] Tone IV remains as a single cohesive tonal ctageory with a distinctive tone contour

Source: Y. Wu, 2005, p.30; M. Y. Chen, 1976, 2000

Table 2.4: Historical processes in the development of tones of Standard Mandarin and Changsha. Each superscript notates the corresponding historical process, among which a, b, c, d are common processes for both varieties and the remaining letters indicate distinct processes related to the development of Standard Mandarin (e, f, g) and Changsha (h, i, j) respectively.

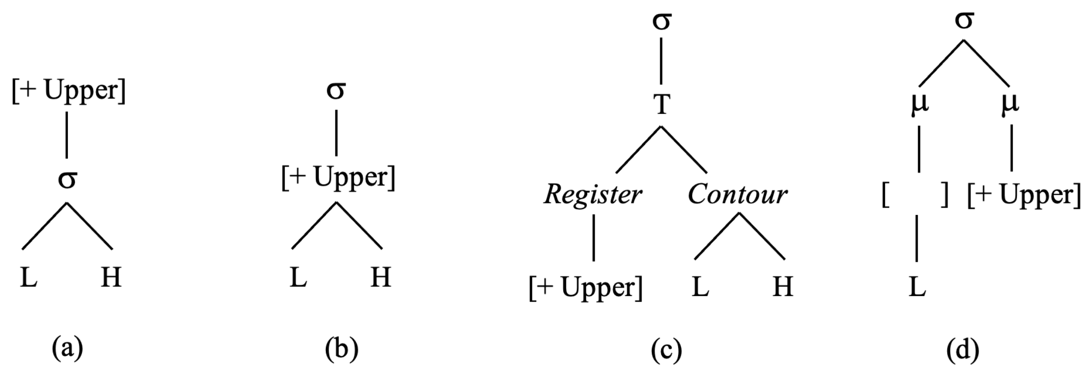
The seemingly arbitrary synchronic correspondence patterns between the tones of Standard Mandarin and Changsha can be better understood from a diachronic perspective. Figure 2.7 illustrates the correspondence between the tonal categories of Middle Chinese (Table 2.1) and modern tones of Standard Mandarin and Changsha (Tables 2.2 and 2.3). The correspondence relation to their common ancestor Middle Chinese entails the divergent processes of tonal change in the development of Standard Mandarin and Changsha. For example, 'wipe' (抹) originally in the *rù* (IV) tone category of Middle Chinese acquired a distinctive tone contour because tone IV evolved into Tone 6 in Changsha, whereas in Standard Mandarin it was redistributed to Tone 1, the category evolved from *yīn píng* (Ia), and acquired the high level tone, homophonous to 'mother' (妈), because the entire *rù* (IV) tone disappeared in Standard Mandarin.

Table 2.4 summarises the major historical processes involved in forming the tone system of Standard Mandarin and Changsha. Such mapping of historical tone categories and paths of tonal changes across related language varieties has been a fruitful method in defining and structuring dialect groups (Ratliff, 2015). It also enables us to predict what morphemes may pattern together.

2.3.3 Representations of citation tones

Despite the flexibility and practical convenience, Chao's tone system has its weaknesses including lack of precision. The distinction of one degree (e.g. 44 and 55) is usually not significant and the choice of five levels was not based on any phonological principles, which makes it lie in a dubious status between a phonetic system or a phonemic one (Duanmu, 2007). In Table 2.2, a phonological representation of Standard Mandarin citation tones (Duanmu, 2000a) is also included, in which four pitch patterns are represented by two tone levels, H (high) and L (low), and their combinations. Analyses of tone feature system were, for the most part, based on autosegmental framework (Goldsmith, 1976), representing tonal features in parallel to the tier of phonological segments, mediated through a tier of tone-bearing units, often syllables or moras.

In Duanmu (2000a)'s model of Standard Mandarin, each tone also has a register feature (e.g. [+mur]) in addition to a pitch feature (e.g. L). Such register system was first proposed in M. Yip (1980), with two tonal features, *Register* [± Upper] and *Tone* [± high], and subsequently modified by a number of tonal geometry models such as D. G. Pulleyblank (1986), Hyman (1993), Z. Bao (1999), apart from Duanmu (2000a). The definition of 'register', however, varies from one model to another. For example, *Register* in M. Yip (1980) divided the pitch range into two halves, while in Z. Bao (1999) and Duanmu (2000a) it denoted articulartory mechanisms: the former



Source: Adapted from M. Yip (2002, pp. 52–53) and J. Wang (1997, p. 165)

Figure 2.8: Feature geometry models of a high rising tone. (a) no Tonal Node dominating *Register* and *Tone* features, e.g. M. Yip (1980); (b) *Register* feature is the Tonal Node, dominating the *Tone* features, e.g. M. Yip (1989); (c) *Tone* features are dominated by a node, which is a sister to the *Register* feature, and both are dominated by a Tonal Node (T), e.g. Z. Bao (1999); (d) Each half of the contour tone is independent, the tone bearing unit is mora μ, *Register* dominates *Tone* features, and empty brackets indicate underspecification, e.g. J. Wang (1997).

adopted [± stiff] as register features to indicate overall tension of vocal folds and the latter used [± mur] to capture murmur or breathiness related to slack vocal folds.

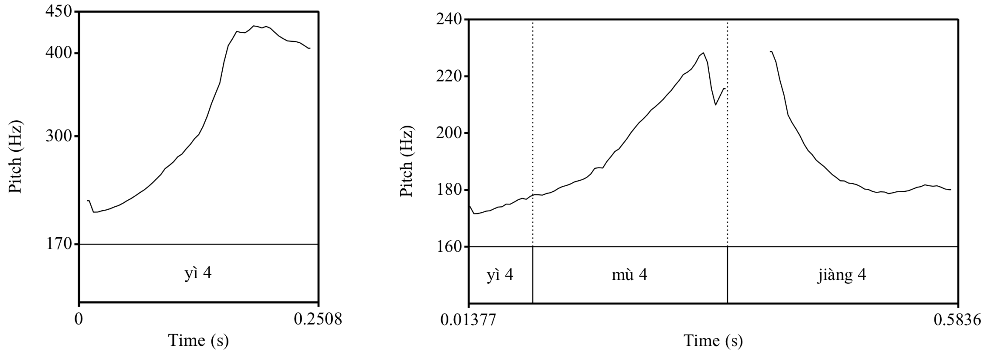
The structural relation between *Register* and *Tone* also diverges in these models: *Register* dominates *Tone* (e.g. M. Yip, 1980; Hyman, 1993) or *Register* and *Tone* are sister nodes (e.g. Duanmu, 2007; Z. Bao, 1999). Figure 2.8 illustrates a few feature geometry models of a high rising tone. M. Yip (2002) argues that model (c) can better account for tonal spreading at three levels including register spreading, contour spreading, and whole tone spreading or copying. J. Wang (1997) follows the underspecification theory (see Kiparsky, 1982; Lahiri and Reetz, 2008) and incorporates the default values [−Upper] and H proposed by D. G. Pulleyblank (1986) into her tonal model of Standard Mandarin in underlying representation. As shown in (d), only [+Upper] and L are specified. One implication of J. Wang (1997)'s proposal is that for Tone 3 in Standard Mandarin, the low tone with L L *Tone* features, the dominating *Register* tier is entirely underspecified. Yet neurolinguistic studies including Event-related potential studies have offered controversial findings regarding the underspecification of Tone 3 (see Politzer-Ahles et al., 2016; Chien et al., 2021; Y. Zhang, 2022). The underlying representation of Tone 3 also varies in the models in terms of whether the rising tail is part of the representation.

There are, however, still phonetic patterns which these register systems failed to accurately characterise such as mid tones (see discussion in M. Yip, 2001). Many models were built based on the four lexical tones of Standard Mandarin, which has a relatively simple tone system compared to other Southern Chinese varieties. The binary tone levels alone are not adequate to represent complex tonal systems, especially those with more than four tones or three discrete tone heights. In addition, tones with the same featural representation cross-linguistically may be perceptually very different. For example, the high rising tone in Standard Mandarin (Tone 2) and the high rising tone in Changsha (Tone 4) differ greatly in the f_0 pattern, which will be evident in Chapter 3. There is also analysis of tones under Optimality Theory (see M. Yip, 2002) including an instantiation of Dispersion Theory (Flemming, 1995).

While H and L are often regarded as pitch targets in phonetic implementation, Y. Xu and Wang (2001) proposes dynamic pitch targets with linear movement specification including Rising and Falling for Standard Mandarin, based on acoustic evidence in the shape of f_0 contours in tonal coarticulation. In fact, the citation form of a syllable is highly likely to be distinguished from its forms in connected speech. This leads to our next section about tones in various contexts.

2.4 Tones in Contexts

While §2.3 is dedicated to the citation tones of monosyllables, this section focuses on tonal patterns in connected speech. Connected speech refers to a natural stream of continuous speech consisting of a sequence of syllables (tones), usually characterised by various features including interaction between adjacent tones. The surface f_0 realisation of connected Mandarin speech is never as straightforward as concatenating contours of the corresponding citation tones (Kochanski et al., 2003). The tone pattern of a syllable embedded in larger units may be drastically different from its monosyllabic citation form (see Kochanski et al., 2003 for Standard



(a) T4 *yi* 'meaning' **(b)** Trisyllabic phrase *yi mu jiang* 'Carpenter Yi'

Figure 2.9: f_0 contours of Tone 4 morphemes in Plastic Mandarin by the same speaker.

Mandarin examples). This phenomenon is illustrated in Figure 2.9 with Plastic Mandarin data from my MPhil study (C. Xu, 2018).

Figure 2.9 (b) shows the f_0 contour of a trisyllabic phrase consisting of three consecutive Tone 4 syllables *yi mu jiang* ('Carpenter Yi') in the utterance-initial focus position, spoken by a sixteen-year-old male speaker. Figure 2.9 (a) presents the f_0 contour of Tone 4 *yi* ('meaning') in its citation form, produced by the same speaker. No f_0 curves of the Tone 4 syllables in the trisyllabic phrase are the same as the reference citation contour, which has a steep rise to around 430 Hz. Conversely, they have much reduced pitch range, up to about 230 Hz. Only the second syllable *mu* in focus has a relatively large f_0 rise resembling the citation contour, and the T4 syllable *jiang* even has a falling pattern.

If the f_0 curve of every Tone 4 syllable in the trisyllabic phrase was replaced with the one shown in (a), then there would be a big jump at both boundaries of the middle syllable: from around 430 Hz to 200 Hz. Such a sharp transition is difficult and rare in speech due to physical constraints of the human articulatory apparatus. Y. Xu (2009, p. 908) refers to 'aspects of timing that are obliged by articulatory mechanisms' as 'obligatory timing', such as minimum duration of articulatory movements, which constraints how fast the pitch can change. Another physical consideration concerns the economy of effort and energy conservation (Lindblom, 1990) such that speakers minimise muscle force as long as the output is perceivable. Continuous f_0 change in speech should, therefore, be smooth and physically reasonable (Kochanski et al., 2003).

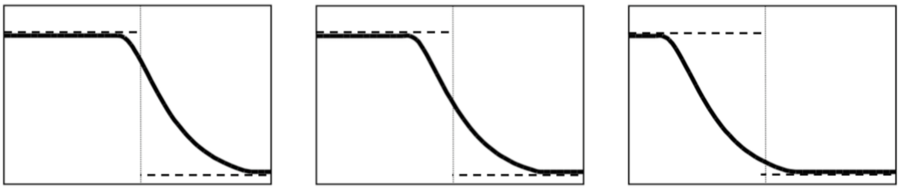
This phenomenon of varied surface pitch realisation has become a site of integrated research involving biomechanical aspects of speech production, phonological aspects of the language, and informational structural aspects in the communication. Especially in total languages, pitch contours are of extreme importance in distinguishing otherwise identical morphemes. The following section §2.4.1 reviews sources and effects of tonal variation in the literature.

2.4.1 Coarticulation, tone sandhi, and intonation

The citation tone contours or some properties thereof such as the pitch height of turning points have usually been conceived as the underlying tonal templates or targets, whose surface realisation is partially modulated by physiological constraints. The constraint that limits f_0's rate of change makes tone contours sensitive to the local tonal context (the preceding and succeeding tones). In fact, adjacent tones have been identified as one of the major sources of influence on tonal variation (X. S. Shen, 1990; Kochanski and Shih, 2000; Kochanski et al., 2003). This intrinsic effect of speech production, however, is not sufficient to explain the preferred patterns in the surface variation of a language. There is language-specific arbitrariness in the directionality and magnitude of coarticulatory tonal effects (e.g. Potisuk et al., 1997; Y. Xu, 1997), parallel to the cross-linguistic differences in segmental coarticulation (Keating & Cohn, 1988).

2.4.1.1 Tonal coarticulation

Shih (1988) undertook a comprehensive study of disyllabic expressions in Standard Mandarin and came up with three tonal coarticulation rules: first, the high falling Tone 4 ends in the mid-range rather than the low target of the baseline of the pitch range when it is in non-utterance-final positions; second, the mid rising Tone 2 does not reach its high target when preceding a high tone (Tone 1 or Tone 4); and third, the low mid beginning of the low falling variant of Tone 3 assimilates to a



Source: Flemming, 2007

Figure 2.10: Schematic illustration of possible timing patterns of tonal coarticulation. The vertical bar in the middle represents the syllable boundary.

preceding high target. These three rules are in fact all assimilatory effects when tones are connected. Y. Xu (1997) also studied disyllabic sequences in Standard Mandarin, and further specified that the assimilatory effect is asymmetric in timing such that the transition to the next tone usually does not start until the onset of the next syllable, whose initial f_0 may have a large excursion away from its target. A substantial portion of the next syllable, as much as two thirds or more, is thus dedicated to the coarticulatory variation. Such f_0 pattern reflects the carryover or perseverative effect of the preceding tone, which is illustrated in (a) of Figure 2.10, where three types of possible timing patterns are schematised. Anticipatory effect of the succeeding tone is also reported in Standard Mandarin, although it is more subtle than the carryover effect. It can be dissimilatory in nature, manifested by a low onset tone target raising the maximum f_0 value of the preceding tone (Y. Xu, 1997).

The predominance of carryover coarticulation tends to be common across tonal languages in south-east Asia such as Cantonese (Y. Li et al., 2004), Thai (Potisuk et al., 1997), and Vietnamese (Brunelle, 2003). Flemming (2007) noticed that the predominant rightward coarticulation in Asian tonal languages is analogous to the rightward bias in phonological tone spreading in African tonal languages (Hyman, 2007) demonstrated in (1), where H and L indicate high and low tones respectively, HL/LH indicates contour tones, and ‘ . ’ indicates syllable boundary.

(1) a. (natural) L.H → L.LH; H.L → H.HL

b. (unnatural) L.H → LH.H; H.L → HL.L

The perseverative tone spread, shown in (1a), is extremely common or 'natural' in Hyman's terms. Hyman (2007) offered a perspective of phonologisation of carryover coarticulation to account for the rightward bias in tone spreading. Flemming (2007) adopted the approach of constraints prioritisation in optimality theory and proposed that both tone spreading and coarticulation are shaped by the same articulatory and perceptually-motivated faithfulness constraints.

2.4.1.2 Tone sandhi and tone stability

The varied surface realisation of tones in connected speech in many Chinese varieties may result from tone sandhi, a consistent tonal alternation triggered by phonological or morphosyntactic environments (M. Y. Chen, 2000). The classic Tone 3 sandhi rule (2) in Standard Mandarin, for instance, turns T3, the low dipping tone (/˨˩˦/ or /˨˩/), into a rising tone T2 ([˧˥]) when preceding another T3.

(2) T3 → T2 / ___ T3

Compared to progressive tone spreading or coarticulation, the anticipatory tonal variation of T3 is less predictable from the neighbouring tones but follows a rule or pattern that governs how the tone changes based on its position in a word or phrase.

Changsha tone sandhi is understudied and whether there is tone sandhi in this variety has been controversial. While X. S. Shen (1991) claimed that no tone sandhi occurs in Changsha, H. Lin (2011) proposed that Changsha exhibits complex bi-directional sandhi patterns, conditioned on the morphosyntactic structure of the word or phrase. The reported tonal phenomena in Changsha were later analysed as a result of the interaction of stress and tone by M. Zhang (2023).

Segmental ellipsis and autosegmental tone stability can also lead to tonal variation. As demonstrated in (3) in Cantonese, the high tone target of the deleted syllable [jat˺] docks on the preceding vowel to form a rising tone (M. Y. Chen, 2000).

(3) *si yat si* → *si si*
[sɪ jat˺ sɪ] [sɪ sɪ]
M H M MH M
'Give it a try.'

2.4.1.3 Downtrends

Pitch lowering in the course of an utterance is a ubiquitous phenomenon cross-linguistically. A theory-neutral term 'downtrend' (Pierrehumbert, 1988, p. 57) is adopted to describe the general f_0 decrease that may result from various effects. This section briefly introduces three downtrends: downstep, declination, and final lowering.

Downstep or catathesis, usually refers to the lowering of a high tone to the right of a sequence of a high tone and a low tone (H-L), which has been observed in many languages such as Kikuyu (Clements & Ford, 1979), English (Pierrehumbert, 1980), Japanese (Poser, 1984), Mandarin (Shih, 1988), and Yoruba (Laniran & Clements, 2003), among others. Different from the L-to-H carryover coarticulation that usually gives rise to an initial f_0 decline, downstep lowers the ceiling of the following H tone, and results in a stepwise descending of H tones in sequences of alternating H and L tones where downstep can be applied in a chain. Pierrehumbert (1988) emphasised that in Japanese and English not any H-L sequence induces downstep, but those in particular structural position such as the accent HL in Japanese and the HL in a single pitch accent in English. Downstep, therefore, is 'a local phonologically controlled modification of pitch range' (Pierrehumbert, 1988, p. 57).

In Shih (1988)'s production experiment of Standard Mandarin, four five-syllable sequences in the format of alternating two tones HLHLH were adopted to examine the downstep. All the H targets in the sequence were Tone 1 (˥), while the intervening L positions were filled with the one of the four tones. The average f_0 of the high level tones in a sequence was measured. This revealed that all other three contour tones triggered the downstep effect, and the Tone 3 variant (˨˩) with lowest pitch level in the L positions led to the greatest downstep effect. In addition, Shih's data showed that the downstep effect of the final H syllable in the sequence was smaller than that of the medial H syllable. Such decrease of the downstep effect over an utterance was also observed in English, and modelled as a decaying exponential function by Liberman and Pierrehumbert (1984). Downstep was also observed in Plastic Mandarin and interacted with focal prominence (C. Xu, 2018).

Declination is another source of downtrend with smaller magnitude yet in larger domain — the f_0 gradually drifts downward in the course of an utterance, and this is not phonologically triggered (Liberman & Pierrehumbert, 1984). There is also evidence in the literature of perceptual compensation for such a downtrend (Pierrehumbert, 1979). Some contested the existence of such a residue global downtrend when downstep effects are factored out. To resolve the potential confound of downstep, Shih (2000) made use of sequences of Mandarin high level tones where no downstep was triggered to study delineation. She identified and modelled the declination effect as an exponential decay.

The third source of downtrend, final lowering or terminal fall, is a consistent phonetic process associated with the boundary of a declarative utterance cross-linguistically, often accompanied by lengthening. Lieberman (1967) offered physiological explanations including the reduction of subglottal pressure. It is often analysed as a low boundary tone marking the finality of an intonation unit (Pierrehumbert, 1980).

2.4.1.4 Focus

In the Plastic Mandarin trisyllabic example of T4 syllables in Figure 2.9, only the medial syllable *mu* among the others has a resembling shape of its citation form. In fact, *mu* was in focus, and was the new information in the discourse.

Focus is the discourse function that highlights a piece of information against the shared information (Y. Xu & Xu, 2005) and this prominent information is often manifested through prosodic variations including f_0. In Cooper et al. (1985)'s work on English, the focused word in sentence-initial and sentence-medial positions had a durational increase of 40% and the f_0 peaks of post-focus words were significantly lowered. When the focus was sentence-final, the acoustic cues of focus were less prominent. This suggests that the focus effect interacts with the position of the focused unit in a utterance. According to Y. Xu and Xu (2005), the on-focus stressed syllables in English had expanded pitch range, the post-focus syllables had reduced pitch range, and the pre-focus syllables had normal pitch range.

Post-focus effects in Standard Mandarin were 'multifaceted manifestations of the weak implementation of post-focus tonal targets' (Y. Chen, 2010, p. 524). Y. Chen (2010) indicated that the degree of distinctiveness of post-focus lexical tones was reduced severely when the focused tone was high. On the contrary, lexical tones of syllables under emphasis in Standard Mandarin were produced with enhanced distinctiveness (Y. Chen & Gussenhoven, 2008), reflected roughly through f_0 range expansion, duration increase, or other adjustments depending on the tone type. Under the circumstance of emphasis, f_0 contours were generally very informative.

2.4.2 Speech models of variable tone patterns

Quantitative methods have been increasingly prevalent in linguistic pitch studies over the past decades. Quantitative modelling of the fundamental frequency enriches our understanding of prosodic phenomena by testing explicit theoretical postulations and using standard mathematical methods as the toolkit. Previous approaches in general can be divided into two categories according to Prom-on et al. (2009), those that model f_0 contours directly and those that simulate underlying mechanisms of f_0 production.

The former includes, for instance, Pierrehumbert (1981) that used linear and sagging functions as transition functions to connect target f_0 values, Hirst and Espesser (1993) that interpolated the target f_0 points with quadratic spline functions, Taylor (2000)'s tilt model that produced f_0 through continuous parameters to control local f_0 rise-fall shapes in intonational events of pitch accents and boundary tones, and Ni et al. (2006)'s constrained tone transformation technique that superimposed global intonation on local tone shapes whose tonal targets were connected by truncated second-order response functions. The more recent work by Hadjipantelis et al. (2012) built well-formed f_0 prototypes of Mandarin tones through functional principal component analysis (FPCA) and linear mixed effect (LME) modelling. As it employs an entirely acoustic approach to the representation of tone, the phonological interpretability seems limited. The major strength is that the linguistic effects of neighbouring tones, consonants, vowels, and breaks were manipulated in their comprehensive LME model.

The various models in the first category can produce satisfying synthetic f_0 contours; however, they hardly differentiate 'surface patterns that carry intended information from those that are due to articulatory mechanisms' (Prom-on et al., 2009, p. 406). The second category, therefore, mostly incorporated articulation-related model components and are physically motivated. Three intonational models on Standard Mandarin in the spirit of analysis-by-synthesis are introduced below. Kochanski and Shih's Soft Template Mark-Up Language (Stem-ML) model (2000;2003) made use of 'soft template' modulated by a set of mathematically defined tags for tone shapes, compromising between articulation effort and communication accuracy. The balance was accounted for by the parameter of 'prosodic strength' at word level. The Command Response (CR) model by Fujisaki (Fujisaki et al., 2005) simulated the control mechanisms of relevant muscles. Xu's Parallel Encoding and Target Approximation (PENTA) model (2005;2009) was acoustically motivated, which conceptualised f_0 realisation as approximating local pitch targets in syllable-bound time structure. On the whole, these models successfully captured variations of the same phonemes in varied prosodic contexts, although these approaches had different predetermined assumptions about f_0 production involving the working mode of muscle groups and the directionality of f_0 smoothing (or the symmetricity of tonal coarticulation).

The thesis falls into the first category and will model f_0 contours using standard mathematical methods. The following section delves into another phenomenon associated with tones in connected speech in many Chinese varieties, neutral tone.

2.5 Neutral Tone

Apart from the categories of four lexical tones in Mandarin Chinese as introduced in Table 2.2, some syllables tend to be light, short, and do not carry any of the four tones. These are traditionally referred to as *qīng shēng* (literally, 'light or soft tone') or neutral tone syllables (Chao, 1968). While the name 'light tone' describes the articulation – these syllables are said to be pronounced softly with less intensity (Hu & Xu, 1995), the term 'neutral tone' reflects Chao's view of the neutralisation of contrasts among lexical tones in unstressed positions. This section introduces and reviews descriptions and discussions of neutral tone, although the majority of the research on neutral tone was based on Standard Mandarin.

2.5.1 Properties of neutral tone syllables

As the seminal work on word-prosodic typology by Hyman (2006) suggested, both tone and stress may co-occur in languages. Stress, however, is defined varyingly in different phonological theories and realised differently across various languages. Despite being known as tone languages, both Standard Mandarin and Changsha also display stress-like phenomena, where certain syllables are more prominent than others in an utterance. Some scholars proposed that they have a stress system (see Duanmu (2000b) and H. Lin (2000) for Standard Mandarin and M. Zhang (2023) for Changsha), and the presence of neutral tone has been a crucial piece of evidence for stress. It has been well established that neutral tone syllables show how stress plays a role in Mandarin, in that they are always unstressed with reduced duration, loudness, and vowel contrasts (Chao, 1968; Shih, 1987; M. Yip, 2002).

2.5.1.1 Vowel, duration, and pitch patterns

Two examples below demonstrate the centralised vowel of a syllable when it becomes a neutral tone syllable in the unstressed position in a disyllabic word: in (4a), the back mid unrounded /ɤ/ is realised as a mid-central vowel [ə]; in (4b), an example from Duanmu (2017), the velar nasal is dropped, nasally coarticulated with the vowel, and a back low unrounded /ɑ/ is realised as a mid-central vowel [ə]. Moreover, the vowel can even be deleted, as in (4c). The four lexical tones in Standard Mandarin are, therefore, sometimes regarded as full tones (e.g. in Y. Chen and Xu, 2006; Tang et al., 2017), while neutral tone being the reduced tone. The degree of reduction tends to depend on the speaking rate, formality, focus, emphasis, and so on. In line with the *Pinyin*[9] transcription of Standard Mandarin in the dictionary (Institute

[9] The official romanisation system for Standard Mandarin in Mainland China.

of Linguistics CASS, 2016), no diacritics denoting tones are assigned for neutral tone syllables (in bold) here. Apart from the vowel centralisation or deletion and nasal coda deletion, some voiceless onsets, especially voiceless unaspirated affricates, in neutral tone syllables may be voiced (Cao, 1986).

(4)	a.	*gè*	/kɤ˥˩/	CL	*zhè**ge***	这个	[ʈʂeɪ˥˩ kə]	‘this one’
	b.	*fāng*	/faŋ˥/	‘direction’	*dì**fang***	地方	[ti˥˩ fə̃]	‘place’
	c.	*fǔ*	/fu˨˩˦/	‘rotten’	*dòu**fu***	豆腐	[toʊf˥˩]	‘bean curd’

In terms of duration, M. Lin and Yan (1980) reported that the duration of a neutral tone syllable was about half of the duration of the same syllable in lexical tone at the same position in disyllabic phrases. It was approximately 60% (Cao, 1986) or varying from 33% to 60% (Jing, 2002) of the duration of the preceding lexical tone syllable. The latter two studies also found that the intensity of neutral tone syllables is not always lower than that of lexical tone syllables.

Neutral tone has also been considered as a contextual tone (M. Yip, 2002) with varied and predictable pitch realisations. It never occurs in initial positions, mostly enclitically, and its pitch patterns tend to depend heavily on the preceding syllables (Chao, 1968; M. Y. Chen, 2000; M. Yip, 2002). The phonetic description of neutral tone in Standard Mandarin, however, is not consistent across different studies. Table 2.5 presents the findings of four influential studies amongst many others (e.g. Dreher and Lee, 1968; M. Lin and Yan, 1980): two impressionistic descriptions of the pitch value of neutral tone by Chao (1968) and M. Y. Chen (2000); and two acoustic analyses by Shih (1987) and Lee and Zee (2008). While the impressionistic studies only reported one pitch value of neutral tone in different contexts, the acoustic studies included the pitch movement details. Although their conclusions collide over the phonetic details, one common characteristic is that neutral tone in Standard Mandarin is falling or overall lower than the offset of the preceding tone when preceded by H, LH, HL tones, but rising or higher than the offset of the preceding tone when preceded by the L (T3) tone.

Preceding tone	Neutral tone			
	Chao, 1968	Shih, 1987	Chen, 2000	Lee and Zee, 2008
1 H /˥/	half-low	starts high, then falls	mid	mid falling
2 LH /˧˥/	mid	starts high, then falls, but not as low as after H	mid	high falling
3 L /˨˩˦/	half-high	starts fairly low, then rises	half-high	mid-level
4 HL /˥˩/	low	starts fairly low, and falls even lower	low	low falling

Table 2.5: Phonetic descriptions of neutral tones in four previous studies.

2.5.1.2 Occurrence, obligatoriness, and functions

A small and limited number of grammatical morphemes including suffixes and particles are the typical examples of syllables bearing neutral tone. These include the nominaliser or possessive/genitive particle *de* [tə], the perfective or inchoative particle *le* [lə], the classifier particle *ge* [kə], the diminutive suffix *zi* [tsz̩], and so on. Some are obligatory instances of neutral tone, *de* for example, because they are not assigned to any canonical lexical tone category. Their lexical tone category may be traced back diachronically (H. Lin, 2006). Table 2 lists a few more examples of morphemes with neutral tone, in which the sub-categories of suffix in the table are not exhaustive. Sometimes personal pronouns served as an object such as *ta* [tʰa] 'him/her' can be neutral tone syllables (Hu & Xu, 1995).

Neutral tones are also found in the second syllable of disyllabic reduplicated words such as kinship terms, and often, of some disyllabic compound nouns such as /po˥ li/ or verb compounds such as /kwo˥˩ lai/ in Table 2.6. These syllables in isolation do keep their lexical tones, as in Example (4). In these cases the neutral tone may not be obligatory and the meaning of a word may not change even when the neutral tone pronunciation is replaced by its corresponding full lexical tone pronunciation. For example, Institute of Linguistics CASS (2016, p. 865) prescribes that the second syllable for 'mom' *mā-ma* is a neutral tone syllable. Yet if a child is calling *mā-mā* with the second syllable being Tone 1, it is still intelligible and acceptable.

Type		Example				Productivity	Lexical tone Counterpart
Suffix	Possessive	*dà de*	/ta˥˩ tə/	大的	'big'	Productive	
	Tense/Aspect	*lái le*	/laɪ˧˥ lə/	来了	'came'	Productive	
	Interogative	*zǒu ma*	/zoʊ˨˩˦ ma/	走吗	'go?'	Productive	
	Classifier	*yī ge*	/i˧˥ kə/	一个	'one CL'	Productive	/kɤ˥˩/
	Noun suffix	*yǐ zi*	/i˨˩˦ tsz̩/	椅子	'chair'	Lexicalised	/tsz̩˨˩˦/
Reduplicative		*mā ma*	/ma˥ ma/	妈妈	'mom'	Lexicalised	/ma˥/
Compound Lexeme		*guò lai*	/kwo˥˩ laɪ/	过来	'cross'	Productive	/laɪ˧˥/
		bō li	/po˥ li/	玻璃	'glass'	Lexicalised	/li˧˥/

Source: Adapted from Tang et al. (2019)

Table 2.6: Classification of Mandarin neutral tone syllables with examples.

Dictionaries usually mark neutral tone syllables and whether they are optional. Examples (5a) and (5b) are two instances of optional neutral tone syllables, and both pronunciations listed for the same words are acceptable. Dictionaries, however, can be inconsistent about the neutral tone status of syllables in some words (K. Zhang, 1998), and the same dictionary may change the neutral tone marking in newer editions (Ma and Wang, 2015; Y. Lin and Li, 2017). There is some arbitrariness or lexical idiosyncrasy about tonal reduction or neutral tones. Each pair of words in (5c) and (5d), for instance, has the same morphological structure with a shared morpheme (i.e./ *lǎo* 'old' and *guā* 'melon or gourd'), but the second syllable in only one word is pronounced as a neutral tone syllable by some speakers and never in the other word.

(5) a. *tǐ liàng* 体谅 /tʰi˨˩˦ ljaŋ˥˩/ 'be considerate'
tǐ ***liang*** 体谅 /tʰi˨˩˦ ljaŋ/
Source: X. Chen (2004, p. 114)

b. *tuǒ dàng* 妥当 /tʰwo˨˩˦ taŋ˥˩/ 'properly'
tuǒ ***dang*** 妥当 /tʰwo˨˩˦ taŋ/
Source: Institute of Linguistics CASS (2016, p. 1338)

c. *lǎo shǔ* 老鼠 /laʊ˧˥ ʂu˨˩˦/ 'rat'
lǎo ***hu*** 老虎 /laʊ˧˥ xu/ 'tiger'
Source: M. Y. Chen (2000, p. 385)

d. *nán guā* 南瓜 /nan˧˥ kwa˥/ ‘pumpkin’
*huáng **gua*** 黄瓜 /xwɑŋ˧˥ kwa/ ‘cucumber’
Source: X. Chen (2004, p. 118)

While neutral tone can serve as an allophonic variant to a full tone in some words, it leads to phonemic distinctions in others. In (6), five minimal pairs are shown as examples. The internal word structure differs between each pair. In (6a), *qí zǐ* is a compound word in which the low-dipping tone *zǐ* refers to small pieces, while the neutral tone *zi* in *qí zi* is a semantically bleached suffix to form a disyllabic word. Similarly in (6b), *dōng xī* with two high tones is a compound word linking two directions and *dōng xi* with a neutral tone denotes a more abstract concept, which is less transparent or accessible from the semantics of each morpheme. However, not all compounds retain the full tones for both syllables. In (6c) we have a verb-noun compound *bào chóu* (‘repay enmity’) and a noun-noun compound *bào chou* (‘monetary reward’) with neutral tone. One common characteristic among the first four pairs in (6) is that the meaning of the disyllabic word with a neutral tone is less straightforward than simply concatenating the semantics of two morphemes – there is a semantically bleached suffix in *qí zi* (6a), a metonymic sense [10] in *dōng xi* (6b), a sharpened meaning by taking the intersection of the meaning of two morphemes in *bào chou* (6c), or a meaning by inference[11] in *dà yi* (6d). (6e) demonstrates neutral tone syllables in a trisyllabic word. The stressed verb-verb compound *qǐ lái* conveys the spatial meaning of going up, while the neutral tone counterpart *qi lai* extends it to the temporal domain with the inchoative interpretation. This suggests that the development of neutral tone may be related to lexical semantic change.

(6) a. *qí zǐ* 棋子 /tɕʰi˧˥ tsz̩˨˩˦/ ‘chess piece’
*qí **zi*** 旗子 /tɕʰi˧˥ tsz̩/ ‘flag’

b. *dōng xī* 东西 /tʊŋ˥ ɕi˥/ ‘east and west’
*dōng **xi*** 东西 /tʊŋ˥ ɕi/ ‘things’

c. *bào chóu* 报仇 /paʊ˥˩ ʈʂoʊ˧˥/ ‘revenge’
*bào **chou*** 报酬 /paʊ˥˩ ʈʂoʊ/ ‘reward, payment’

[10] *East and west* represents any substance within east, west, south, and north of the world.
[11] A person is *careless* about details, when only focusing on the *general idea*.

d. *dà yì* 大意 /ta˥˩ i˥˩/ 'general idea'
dà ***yi*** 大意 /ta˥˩ i/ 'careless'

e. *xiǎng qǐ lái* 想起来 /ɕiɑŋ˧˥ tɕʰi˨˩ lai˧˥/ 'want to get up'
xiǎng ***qi lai*** 想起来 /ɕiɑŋ˧˥ tɕʰi lai/ 'recall'

Neutral tone can also distinguish the part of speech of a word. In (7), when the second syllable is a neutral tone syllable, *dà qi* can be an adjective or a noun; otherwise *dà qì* is a noun. Only about 4% of words rely on neutral tone syllables to differentiate meanings or part of speech (Hu & Xu, 1995). In these words, neutral tone plays a crucial role and is often compulsory in the dictionary. The number of words with a compulsory neutral tone syllable, though, has been decreasing (Ma and Wang, 2015; Y. Lin and Li, 2017).

(7) *dà qì* 大气 /ta˥˩ tɕʰi˥˩/ 'generosity'
dà ***qi*** 大气 /ta˥˩ tɕʰi/ 'generous, classy'
Source: Liang (2010)

Iwata (2018) further discussed the occurrence of neutral tone in different prosodic domains. Given the shorter duration of neutral tone syllables, Iwata (2018) argued that disyllabic words or compounds containing a neutral tone syllable have trochaic stress, while the phrasal stress in Standard Mandarin is iambic and it can predominate the trochaic lexical stress[12]. More specifically, the predominance of phrasal stress is indicated by reviving the original lexical tone of the underlyingly toned neutral tone syllables. This is demonstrated in the following examples in (8), in which the second syllables in the disyllabic words are neural tone syllables (in bold). When the disyllabic words are extended to trisyllabic and quadrisyllabic compounds by affixation such as the addition of an adjective *dà* 'big' in (8a) and a negation marker *bù* in (8b and c) or reduplication in (8d and e), the morphemes in the final position, which are the same morphemes bearing neutral tone in disyllabic words, carry their lexical tones.

[12]The phrasal stress can indeed be reanalysed as a trochee too, instead of an iamb. For instance, the trisyllabic phrase in (8c) can have the foot structure: ($\acute{\sigma}$ σ) $\acute{\sigma}$, so that the first and the third syllable are stressed.

(8) a. *xué* ***sheng*** 学生 /ɕɥe˦˥ ʂəŋ/ 'student'
dà xué shēng 大学生 /ta˥˩ ɕɥe˦˥ ʂəŋ˥/ 'university student'

b. *zhī* ***dao*** 知道 /ʈʂɻ̩˥ taʊ/ 'know'
bù zhī dào 不知道 /pu˥˩ ʈʂɻ̩˥ taʊ˥˩/ 'not know'

c. *kàn* ***jian*** 看见 /kʰan˥˩ tɕian/ 'see'
kàn ***bu*** *jiàn* 看不见 /ʰan˥˩ pu tɕian˥˩/ 'unable to see'

d. *huāng* ***zhang*** 慌张 /xwɑŋ˥ ʈʂɑŋ˥/ 'panicked'
huāng ***huang zhang*** *zhāng* 慌慌张张 'panicked'

e. *dì* ***dao*** 地道 /ti˥˩ taʊ˥˩/ 'genuine'
dì ***di dao*** *dào* 地地道道 'really genuine'

Source: Iwata (2018, p. 187)

The final syllable in trisyllabic phrases, however, are not necessarily a lexical tone syllable, as in (9). The occurrence of successive neutral tones mentioned in Z. Li (2003) is also frequent. Iwata (2018) stated that trisyllabic phrases with two consecutive neutral tone syllables such as (9b) are limited in type frequency, although show high token frequency. The fact that tonal contrasts are only preserved (not neutralised) in stressed syllables in Optimality Theory can be analysed as positional faithfulness constraint so that the underlying tones are preserved in prosodic heads (see M. Yip (2002) for details).

(9) a. *gē* ***bo*** *zhǒu* 胳膊肘 /kɤ˥ po ʈʂoʊ˨˩˦/ 'elbow'

b. *xué* ***sheng men*** 学生们 /ɕɥe˦˥ ʂəŋ mən/ 'students'

Chao (1968, p. 39) reported that tonal reduction is hardly predictable on phonological, syntactic, or semantic grounds, and historical factors often come into play. The formality of speech tends to play a role too. In connected colloquial speech, neutral tone occurs rather frequently. About one-third of all syllables can be neutral tones, according to (Duanmu, 2007). In written texts, neutral tone syllables are estimated to be about 15%-20% (C. N. Li & Thompson, 1981).

2.5.1.3 Neutral tone in other varieties

Neutral tone can serve as a dialectal cue for different Mandarin varieties. First, the number of neutral tone morphemes tend to differ among Mandarin varieties (H. Lin, 2000). Neutral tone typically plays a predominant role in Mandarin varieties in the north of China, in particular Beijing Mandarin, but is less common in Mandarin varieties in the southwest (Yuan, 1960). Second, the realisation of neutral tone can be specific to a Mandarin variety. Ürümqi Mandarin in Xinjiang, for example, is different from Beijing Mandarin in that the neutral tone syllables in Ürümqi Mandarin only undergo tone loss, but not reduction in duration (Wei, 2011). The toneless and some other frequently used neutral tone syllables in Southwestern Mandarin bear a high level tone, regardless of the preceding tone categories, and their duration is not as short as that of neutral tone syllables in Standard Mandarin (X. Li, 2004). Neutral tone syllables in Taiwan Mandarin are reported to have a mid-low pitch target and the occurrence of neutral tone is less frequent in Taiwan Mandarin than in Standard Mandarin (K. Huang, 2018).

Citation tone	Broad Neutralisation	Narrow neutralisation
˧ 33	˧ 3	˥ 5
˩˧ 13	˧ 3	˧ 3 or ˥ 5
˦˨ 42	˦ 4	˧ 3 or ˥ 5
˦˥ 45	˥ 5	˧ 3
˨˩ 21	˨ 2	˧ 3 or ˥ 5
˨˦ 24	˦ 4	˧ 3 or ˥ 5

Table 2.7: Tonal neutralisation in Changsha by Zhong (2003).

Neutral tone in Changsha was reported in Zhong (2003), which was based on the impressionistic transcription of tones in the speech of one female Changsha speaker born in 1970. According to Zhong (2003), there are two neutralisation patterns in Changsha as shown in Table 2.7, where the broad neutralisation applies to most words in Changsha and the narrow neutralisation is found in a minority of words, mostly nouns. Unlike neutral tone in Standard Mandarin whose pitch patterns were largely predicted by the preceding tone, the neutral tone of a syllable in Changsha depend on the (underlying) citation tone of the syllable regardless of the preceding and following tones (Zhong, 2003). From Table 2.7 we can see that neutral tone in Changsha can exhibit four levels of pitch height (broad neutralisation) and a neutral tone broadly retains a part of the citation tone of the syllable. More empirical research on Changsha is needed, although that is beyond the scope of the current thesis.

2.5.2 Neutral tone typology

X. S. Shen (1992) proposed a three-way typology for neutral tones in Standard Mandarin: toneless, detonic and atonic. The functional morphemes such as the particles *de*, *le*, and *ma* in Table 2.6 were considered underlyingly *toneless*, whose pitch does not belong to or derive from the four lexical tones. The other two groups are underlyingly toned. The obligatory neutral tones in minimal pairs such as (6) where neutral tone leads to phonemic distinction were *detonic* and used in approximately 200 minimal pairs (X. S. Shen, 1992). The idiosyncratic and optional neutral tones resulting from fast careless speech were *atonic*. Such classification differentiated diachronic tone loss and synchronic de-stressing, which might be different stages in the development of neutral tone.

Y. Zhang (2022) probed into the historical development of many neutral tone syllables and supported the categorisation based on the underlying forms, that is, underlyingly toneless and underlyingly toned. Y. Zhang (2022) named the two categories as *intrinsic* and *derived* neutral tones. The former corresponds to the toneless functional particles in X. S. Shen (1992) and W. Li (1981), and these were lexicalised early in the history, before Ming dynasty (mid-14th century or earlier). The latter may be the intrinsic neutral tone in development, triggered by the stress movement in compounds and phrases motivated by semantic shifts, disappearance of *qù* (IV) tone, and contacts with Manchu. The developmental process, though, tends to have slowed down or even reversed, considering the decline of the number of derived neutral tone syllables in the newer editions of dictionaries. Through processing experiments, Y. Zhang (2022) suggested that intrinsic neutral tone syllables have an underspecified tonal target while derived ones are underlyingly the same as their lexical tone counterparts.

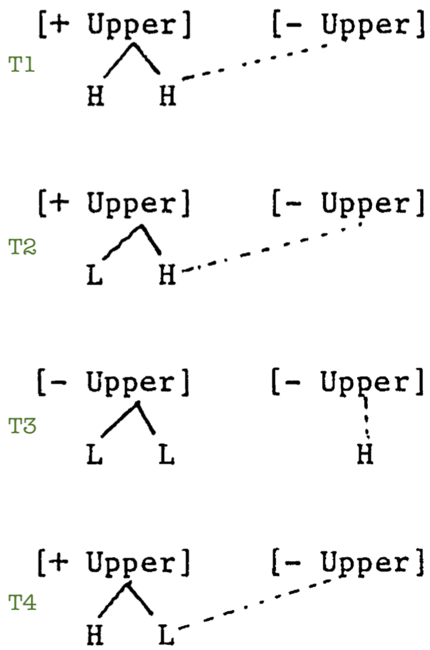
Nevertheless, whether the different types of neutral tone have different phonetic representations were not addressed by X. S. Shen (1992). Y. Zhang (2022) reported that focus increased the similarity between the derived neutral tones and their corresponding underlying (citation) lexical tones, but did not result in significant change in average f_0 range and height of the intrinsic neutral tones, which exhibited a falling contour regardless of the focus conditions. In reality, whether neutral tone is underlyingly toneless has been controversial in the literature. In the following §2.5.3, we will revisit and discuss the representations and realisations of Standard Mandarin neutral tone under different theoretical frameworks.

2.5.3 Approaches to neutral tone

Neutral tone in Standard Mandarin has been analysed as the fifth lexical tone category, a kind of tone sandhi phenomenon (M. Y. Chen, 2000), neutralisation phenomenon (Iwata, 2018), or weak tonal implementation with reduced articulatory effort. Views diverge in terms of whether neutral tone in Standard Mandarin has any phonological specification, whether there is any specific phonetic target, and what is the mechanism of f_0 realisation.

2.5.3.1 Feature spreading

The contextual dependency of neutral tone in Standard Mandarin revealed by previous impressionistic observations and earlier acoustic studies has led to few attempts on the phonological specifications of neutral tone as an independent tonal category. Within the framework of autosegmental phonology, M. Yip (1980) proposed that neutral tone was specified with a register feature [−upper] and the surface values of neutral tone were realised through left-to-right feature spreading from the preceding tone, illustrated in Figure 2.11. Recall that Yip's tonal framework



Source: M. Yip (1980, p. 253)

Figure 2.11: Derivation of surface values of neutral tone by M. Yip (1980).

consists of two tiers: register tier [+/−upper] and tonal tier [H/L], which combine to define the four lexical tones. The neutral tone in the first three cases in Figure 2.11 has the final form of [−upper, H], which corresponds to a mid-range pitch value. The neutral tone after the falling tone has the final form of [−upper, L], a low pitch. This roughly matches the phonetic descriptions in Table 2.5, though [−upper, H] covers a range of slightly different mid-range pitch values. In Yip's proposal though, a special rule was needed, as shown in Figure 2.11, where an H feature was inserted (instead of being linked to the preceding tone) when a low tone (T3) preceded a neutral tone, so as to account for the pitch rise after a low tone.

X. S. Shen (1992) adopted a similar feature spreading view as M. Yip (1980), linking neutral tone to the end component of the preceding tone (i.e. [H] or [L]), since she observed that the f_0 contour of neutral tones is in congruence with the f_0 direction of preceding Tones 1, 2, and 4. But her model differs in that toneless neutral tone such as *de* is **not** specified with any features. The tones in X. S. Shen (1992) were not represented in two tiers but in matrix tone, that is, [HH], [LH], [LLH], and [HL] for Tones 1 to 4 respectively. The underlyingly toned neutral tone syllables are first de-stressed so that the underlying tone is deleted. Another major difference between these two proposals is the controversial underlying representation of Tone 3, and how the two variants of Tone 3 illustrated in Table 2.2 and Tone 3 sandhi are derived through rules. Instead of the insertion of an external H tone by an ad hoc rule, X. S. Shen (1992) considered that a floating H tone was part of the underlying representation of Tone 3, the low-dipping variant with a rising tail [LLH]. This floating H component can dock on a following toneless neutral tone, or be lost via a deletion rule to derive the low variant [LL] in certain contexts.

2.5.3.2 Underlying or boundary L tone

Given that neutral tones after Tones 1, 2, and 4 are typically falling, they may be associated with an L tone. Apart from acquiring a tonal target from the preceding tone through progressive assimilation, H. Lin (2006) posited an underlying L tone for neutral tone. The higher or rising neutral tone after Tone 3, an L tone, was accounted for by dissimilation appealing to the Obligatory Contour Principle.

Z. Li (2003) also assigned an L target to neutral tone, but via a different mechanism. The positional constraint that neutral tone syllables are never in the initial position of a phrase and always cliticised to the preceding lexical tone syllable motivates the unit of a prosodic word, which can consist of a neutral tone syllable or a sequence of neutral tone syllables and the preceding lexical tone syllable regardless of their morphosyntactic structure. In Z. Li (2003)'s account, the pitch target of neutral tone syllables is derived by inserting a post-lexical prosodic word boundary tone L at the right edge of the last neutral tone.

2.5.3.3 Target interpolation and approximation

The phonological analyses are based on various tonal frameworks that define the four lexical tones in Standard Mandarin in different ways and the derived tone specifications do not represent the instrumental data very well. Some experimental studies also proposed tonal targets for neutral tone. In van Santen et al. (1998)'s tone sequence approach in speech synthesis, the neutral tone has a single mid target slightly before the end of a syllable and the target shifts from mid to low when following a falling tone. Target interpolation has been employed or assumed by many theories in phonetic realisation such that the surface f_0 of neutral tone is interpolated between the preceding tone target and the neutral tone target, the boundary tone target, or the following lexical tone target (Shih, 1987).

Another approach is target approximation, where a tonal target is the intended goal within the duration of the tone-carrying syllable (Y. Xu, 1997). Y. Chen and Xu (2006) identified a static and mid target for neutral tone and explained that the greater surface variability of neutral tone is due to the much slower or less robust implementation of the target and a heavy carryover influence of the preceding syllable. An advantage of the target approximation approach is that the short duration of neutral tone is integrated in their analysis — the short duration leads to inadequate target attainment, and thus, varied pitch patterns. In their study, four Beijing speakers produced sentences containing zero to three consecutive neutral tone syllables with varying preceding and following tones. When there were more than one consecutive neutral tones, the f_0 data of neutral tone sequences following different lexical tones displayed converging trends, which provided evidence for the existence of a tonal target for neutral tone. Q. Li and Chen (2019) and Sun and Shih (2021) adopted speech materials similar to Y. Chen and Xu (2006), to investigate the effects of focus and boundary in f_0 of neutral tone in Tianjin Mandarin and anticipatory tonal coarticulation in neutral tone in Standard Mandarin.

2.6 Summary

Following an introduction of Plastic Mandarin in Chapter 1, this chapter has further provided the necessary background for understanding and discussing Plastic Mandarin by delineating its social and geographical contexts and incorporating the linguistic foundations to Chinese tones. The knowledge of tones in Standard Mandarin and Changsha affords insights into tonal variation and change from a diachronic perspective and portrays the linguistic forms available in the local repertoire of downtown Changsha in recent decades.

In order to profile and investigate the tonal system of Plastic Mandarin, the thesis bifurcates into two strands featuring citation tones (Chapter 3) and neutral tones (Chapter 5) respectively. The study of neutral tones is less straightforward compared to isolated tones in citation form because there are no stand-alone neutral tones in speech. The inevitable connected speech phenomena such as coarticulation, downtrends, and focus should, therefore, be taken into consideration in the data collection and analysis of neutral tones. Chapter 4 provides a subsidiary investigation to Standard Mandarin neutral tone using corpus approach to establish a benchmark for comparison with Plastic Mandarin. The variable patterns of neutral tone in Standard Mandarin manifest the complexity in tonal representation, structure, and realisation. A comparative study of neutral tones in Plastic Mandarin and Standard Mandarin may reveal how prosodic features interact with language variation.

3
Citation Tones of Plastic Mandarin

Contents

3.1 Introduction

Plastic Mandarin is a new variety of Mandarin that may still be quite variable but that features a core of distinctive pitch patterns. This chapter examines the acoustics of the citation tones of Plastic Mandarin as well as two other Chinese varieties including Standard Mandarin and Changsha (Xiang).

The primary objective of this chapter is to establish the tone system of this emerging variety, Plastic Mandarin, since it has received limited scholarly attention in the past. While the pilot study in my MPhil thesis (C. Xu, 2018) reported on f_0 data of a single utterance of one syllable ([iː]) in each of the four tone categories from a few speakers, S. Shen (2019) reported on f_0 data of a single utterance of 60 arbitrary Chinese characters, 15 each tone category. The latter, however, was based on only one speaker. This chapter, therefore, provides a thorough analysis of the citation tone system of Plastic Mandarin by utilising speech data from multiple speakers, containing multiple repetitions of various syllables in each tone category. The use of multi-speaker and multi-item (syllable) data allows for an examination of the inter-speaker and intra-speaker consistency of Plastic Mandarin tones, so as to determine whether Plastic Mandarin is a system with stability or an idiosyncratic accent unique to an individual.

The second objective of this chapter is to explore the development of Plastic Mandarin tones using a comparative approach. Historically tone often arises from laryngeal features such as voicing and aspiration spreading from consonants to the neighbouring vowels (Humbert, 1978), or, in some cases, from other features of a syllable such as vowel length (e.g. Svantesson, 1991), vowel quality, and voice quality.

These features can contribute to changes in tone patterns and tone inventories. Language contact is another factor that underlies tonal change, though how exactly tones were transferred or stimulated to change remains unclear. Contact may even lead to tone emergence in non-tone languages (Thurgood, 1999) or tone loss in tone languages (Salmons, 1990). Tone has been identified as one of the prosodic aspects of language that is highly diffusible through language contact (Ratliff, 2015). The rise of Plastic Mandarin coincides with increasingly intense contact between Standard Mandarin and Changsha in the city of Changsha. This empirical study presents a case of Mandarin tonal change under contact in a multilingual community.

Given the linguistic context in the city of Changsha (§2.2), this chapter hypothesises that the emergence of Plastic Mandarin is closely related to the two varieties in contact, Standard Mandarin and Changsha. By employing comparative datasets of the three varieties and comparing the acoustic manifestations of tones in Plastic Mandarin with those in Standard Mandarin and in Changsha, this chapter answers the central questions: How are the lexical tones of Plastic Mandarin compared with those of Standard Mandarin and Changsha?

This chapter examines the duration and f_0 contours of monosyllables produced in isolation in Plastic Mandarin, Standard Mandarin, and Changsha. In particular, it offers a systematic approach to characterise and compare f_0 contours of lexical tones in tone languages. On the one hand, it documents some basic phonetic characteristics of these three varieties and situates Plastic Mandarin among the other two established varieties. On the other hand, the duration and f_0 contour of citation tones constitute a baseline for the study of tone sequences in the later chapters. The study of a crystallising regional mandarin contributes to our understanding of dialect levelling and language variation in contact. Apart from the group patterns, individuals' speech is also examined to better understand Plastic Mandarin from the perspective of language variation and language acquisition.

This chapter addresses the following specific questions:

1) How long is a syllable in isolation in Plastic Mandarin, Standard Mandarin, and Changsha?

2) Does tone category and syllable structure have an impact on syllable duration?
3) What is the prototypical pitch contour of each tone in Plastic Mandarin and how does it vary?
4) What is the prototypical pitch contour of each tone in Changsha spoken by the same group of Plastic Mandarin participants?
5) What are the similarities and differences of the pitch contours among tones of Plastic Mandarin, Standard Mandarin, and Changsha?
6) What is the likely path of change from Standard Mandarin to this innovative variety Plastic Mandarin?

This chapter is organised as follows: §3.2, §3.3 and §3.4 present the methodological and analytical decisions for data collection and analysis. This chapter is based on the monosyllabic part of the speech materials (see §3.2.7) collected from my linguistic fieldwork introduced in detail in §3.2. Analysis of other parts of the speech materials is presented in chapter 5. The following section §3.3 describes and justifies the acoustic preprocessing procedure pertaining to the monosyllabic dataset including the transcription, segmentation, extraction, and acoustic measurement. Then §3.4 outlines the course of statistical analysis and addresses the statistical methods. §3.5 and §3.6 report on the syllable duration and pitch contour of citation tones in the three varieties respectively. Under each section, the specific statistical models are specified prior to the presentation of results and discussion. §3.7 explores the origin of Plastic Mandarin tones by conducting comparative analysis of tones, examining the individuals' tone patterns, and discussing the potential formation of Plastic Mandarin tones.

3.2 Speech Data

This empirical study constructed and employed the specialised speech corpora of monosyllables in Plastic Mandarin, Standard Mandarin, and Changsha. This section delineates relevant details of the speech data collection.

3.2.1 Data sources

3.2.1.1 Fieldwork visits in Changsha, Hunan

The Plastic Mandarin and Changsha data were mainly collected over the course of a fieldwork visit to Changsha, Hunan, China in January 2020. Prior to this, I conducted a preliminary fieldwork on a smaller scale during my MPhil studies in September 2017. Some of the quotes expressing attitudes toward Plastic Mandarin presented in §1.1.3 were from teenage participants in the preliminary fieldwork.

Senior high school students are the most ideal sources of Plastic Mandarin data. Most of these competent youths are Changsha residents, who are fluent and frequent users of Plastic Mandarin. Even if some students speak a different dialect at home, Plastic Mandarin has been the de facto lingua franca at schools in downtown Changsha. The linguistic situation may be different in local colleges and universities, since their nationwide admission often results in a diverse student body from all parts of China, among whom Plastic Mandarin may not be the common language. This makes university students less ideal subjects.

The fieldwork site was Changsha Nanya Middle School, a state secondary school located in downtown Changsha (see Figure 3.1). It is one of the boarding school campuses of Yali School that was established in 1906. It covers both junior and senior high school education from grades 7 to 12, with over 6000 students. As one of the most prestigious high schools in Hunan province, it attracts province-wide top applicants and its admission is extremely competitive. Its nature of a boarding school tends to create speech communities with dense and multiplex social networks and high frequencies of interaction. Students are tied together through not only classes and studies, but also extracurricular, social, and daily activities. The classmates who attend all classes together in the same classroom may also be roommates and join the same sports clubs or societies.

Thanks to the support and assistance from my former teachers in this school, I distributed recruitment materials, interacted with students, secured a room for recording, and arranged recording sessions amid their tight schedule in the boarding school. Students' speech was recorded, and written consent had been

Source: Google map search

Figure 3.1: Map of downtown Changsha and the location of Changsha Nanya Middle School.

given in accordance with the ethical approval the study had obtained (Central University Research Ethics Committee (CUREC) – Social Sciences and Humanities Interdivisional Research Ethics Committee Approval Reference: R67136/RE001).

3.2.1.2 Recording in the Phonetics Laboratory, Oxford

The comparative Standard Mandarin data were collected at the Phonetics Laboratory, University of Oxford. Due to the COVID19 pandemic lockdowns, data collection in China was not feasible. The Phonetics Laboratory at Oxford equipped with professional audio studios stood out as an alternative choice. In Oxford, the recording sessions continued on and off in December 2020, May and October 2021, following the COVID regulations, safety guidelines, and participants' consents.

Figure 3.2: The music room with porous sound-absorbing wall materials for sound recording in Changsha Nanya Middle School, Hunan, China. On the door, it says *Piano Room B*.

3.2.2 Premises

Schools are the exact context in which natural and spontaneous conversations in Plastic Mandarin predominate. While organising recording sessions in a school situates speakers in the familiar context of using Plastic Mandarin, it also ensures a safe and convenient participation for students.

In Changsha Nanya Middle school, the recording sessions were conducted in a music room in the school. The music room as shown in Figure 3.2, being small and furnished with soundproofing materials on the walls, was in fact ideal for recording, despite that the school does not have an audio studio with strict laboratory standards.

Figure 3.3: The audio studio with soft furnishings for sound recording in the Phonetics Laboratory, University of Oxford.

The music room sits on the top floor of the Arts Centre, an independent building separated from the noisy teaching zone of the school, and no activities were organised there during my visits. It is suitably quiet for speech production experiments. In this boarding school, students have a packed schedule of academic work under rigorous management. The sessions were arranged at their convenience, mostly in late afternoons and evenings after their last lecture of the day, so that their participation in this study did not affect their normal course schedule in the school.

In Oxford, the audio studio in the Phonetics Laboratory, as shown in Figure 3.3, was employed.

3.2.3 Recruitment

3.2.3.1 Plastic Mandarin / Changsha speakers

During the fieldwork visits, self-claimed Plastic Mandarin speakers were recruited with a preference for speakers who also speak Changsha and have spent the majority of their life in the city of Changsha.

Due to students' restricted access to electronic devices in school, the recording sessions were advertised through printed copies of a recruitment package, available in written Chinese and English in Appendix A, which contained a poster (A.1), information sheets for the participants (A.2.1 to A.2.2) and their parents (A.3), a written consent form (A.4), and a participant form (A.5.1). Recruitment materials provided a general background of this study without revealing specific research questions of the investigation. On the participant form, there was a brief survey about their language use in major social life including their interactions with close families, friends, and school teachers, apart from requesting some basic information including age, place of birth, and current residence.

Prior to the distribution of the recruitment packages, I delivered 50-minute talks in a few classes, with approximately 60 students per class. The talks were themed on sharing my student experiences in the universities in Hong Kong and Oxford, as well as tips about university-entrance exams, which attracted the senior high school students the most. During each talk, I also briefly talked about linguistics, asked students about their observations and understanding of Plastic Mandarin, and advertised my research fieldwork. The talks, on the one hand, enabled me to observe and verify students' frequent use of Plastic Mandarin and their awareness, and to explain the details of how to proceed with participation if they were interested. On the other hand, they helped me build a good rapport with students. My alumna identity and shared school-life experiences reduced our social distance. To be one of "them" was conducive to creating a friendly and relaxed atmosphere for the elicitation of a non-standard accent, which was conducted by me. This minimised the influence of my presence in the context, which should not alter their usual language code.

In addition, two of my Plastic Mandarin participants were recorded in the Phonetics Laboratory, University of Oxford. They graduated from Changsha Nanya Middle school and happened to study in the University of Oxford at the time of data collection. They voluntarily offered their time to take part in this study.

3.2.3.2 Standard Mandarin speakers

Standard Mandarin speakers were recruited in Oxford based on the criteria that they are Standard Mandarin or Beijing Mandarin native speakers who do not speak or speak little of other Chinese varieties. In Oxford, recruitment information was shared on WeChat, a social media app used by most Chinese students. For interested individuals, a information sheet for the participant (Appendix A.2.3) was provided in PDF format. They also filled in a participant form through Google Forms (Appendix A.5.2). They were paid £5 for their participation.

Participants were highly encouraged to attend the recording sessions with a friend of theirs who fits the specified criteria, so that they could speak in front of their close peers as they normally would. The presence of close peers is considered to be a norm-enforcement mechanism (Milroy, 1980). In this way, participants' social networks were also utilised in the recruitment, which is also known as 'snowball technique'.

3.2.4 Subjects

Thirty-five young adults participated in this study. No history of speech or hearing disorders was reported. Written informed consent was acquired from subjects prior to recording.

Twenty-one participants contributed to the elicitation of Plastic Mandarin (sixteen females and five males; mean age = 17.24 years, $\sigma = 0.70$ years). They were all students from Changsha Nanya Middle School, and recruited from pre-existing friendship groups. Table 3.1 presents the demographic profile of these participants. Fourteen of them were born and bred in Changsha; an additional four of them (speakers 105, 106, 108, and 110) almost lived their whole life in Changsha; the rest were born in nearby cities in Hunan province and moved to Changsha at an early age. On average, they had spent 15.71 years in Changsha by the time of elicitation.

Not all of these young adults from Changsha speak Changsha well. Some speak another different regional dialect. For instance, among the fourteen who were born in Changsha, one was from Ningxiang, a county-level city under the administration of Changsha, and speaks a regional dialect Ningxiang. The young generation's confidence with their Changsha proficiency seemed rather mild. Among these 21 Changsha residents, only about 19% (4 out of 21) claimed high level of competence (score 5 out of 5) in their Changsha proficiency in an informal self-assessment

Speaker ID	Sex	Age	Place of Birth	Years in Changsha	Self-rated Changsha proficiency
101	Female	17	Zhangjiajie	12	4
102	Female	17	Changsha	17	5
103	Female	17	Changsha	17	4
104	Female	17	Changsha	17	4
105	Male	17	Loudi	17	4
106	Male	18	Loudi	18	4
107	Male	17	Changsha	12	5
108	Female	17	yongzhou	17	2
109	Female	17	zhuzhou	6	3
110	Female	17	Yueyang	17	4
111	Female	17	Changsha	17	4
112	Female	17	Changsha	17	4
113	Female	17	Changsha	17	5
114	Female	17	Changsha	17	4
115	Female	17	Changsha	17	4
116	Female	17	Jishou	6	2
117	Female	17	Changsha	17	5
118	Male	17	Changsha	17	2
119	Male	17	Changsha	17	2
120	Female	18	Changsha	18	2
121	Female	20	Changsha	20	3

Table 3.1: Demographic profile for Plastic Mandarin participants. The self-assessment of their Changsha proficiency was on a five-level Likert scale from "*I don't speak Changsha at all*" (represented by 1) to "*I speak Changsha fluently*" (represented by 5).

survey, while about 67% (14 out of 21) claimed moderately high level (score 4 out of 5). On average, they reported a score of 3.62 ($\sigma = 1.07$) for their Changsha proficiency, while a score of 4.05 ($\sigma = 0.74$) for their Standard Mandarin proficiency. A paired t-test of the mean differences of the distribution of self-reported scores for the two varieties $t = -1.37, df = 20, p = 0.19$ suggests that the average scores do not significantly differ. While the informal survey about language proficiency is preliminary and small in sample size, it reveals that Mandarin has become a language to be reckoned with among young people in Changsha.

Among those who rated 4 or 5 for their Changsha proficiency, thirteen participants (highlighted in grey in Table 3.1, mean age = 17.08 years, $\sigma = 0.28$ years) speak little other regional varieties and, therefore, were selected to contribute to

Speaker ID	Sex	Age	Place of Birth	Years in Beijing	Self-rated SM proficiency
201	Female	23	Shandong	18	5
202	Female	24	Beijing	11	5
203	Female	26	Beijing	17	5
204	Female	22	Hubei	4	5
205	Female	24	Shandong	0	5
206	Male	24	Liaoning	0	5
207	Male	26	Beijing	25	5
208	Female	22	Shandong	22	5
209	Male	23	Beijing	23	4
210	Female	23	Shandong	2	5
211	Female	22	Hebei	0	5
212	Female	23	Beijing	18	4
213	Male	29	Hebei	0	5
214	Male	25	Beijing	18	4

Table 3.2: Demographic profile for Standard Mandarin (SM) participants. The self-assessment of their Standard Mandarin proficiency was on a five-level Likert scale from "*I don't speak Standard Mandarin at all*" (represented by 1) to "*I speak Standard Mandarin fluently*" (represented by 5).

the task where speaking Changsha was needed.

Another fourteen young adults (nine females and five males; mean age = 24 years, $\sigma = 1.96$ years), recruited in Oxford, United Kingdom, participated in the elicitation of Standard Mandarin. Table 3.2 presents the demographic profile of these participants. All of them are from the Mandarin Chinese region in northern China, among which eight spent more than 10 years in Beijing. None of them claimed to speak or frequently use another Chinese variety or Mandarin dialect[1]. Their average score of self-reported Standard Mandarin proficiency was 4.79 ($\sigma = 0.43$).

3.2.5 Equipment

Subjects attended the recording session in friendship pairs. Each speaker in the pair had a Rode Lavalier microphone clipped to their collar or lapel below the speaker's chin, and the speech of each speaker was recorded onto a separate channel of a Marantz digital audio recorder (PMD661 MK.II) at a sampling rate of 44,100Hz

[1] All subjects speak English at various levels, since English has been a compulsory subject in schools in China. It is very hard to find young monolingual Standard Mandarin speakers.

and 16-bit quantisation. The same set of equipment was used for all pairs of participants. The recordings were initially stored on a Secure Digital 32GB Memory Card before being transferred to a MacBook Pro (Early 2015) laptop for analysis and to the file server of the Oxford Phonetics Laboratory for backup. Volume was constantly monitored throughout the recordings to ensure that no high amplitudes in the speech signal were clipped.

A hard copy of overall instructions and main speech materials was given to participants before the recording started, while instructions and prompts for the speech production experiment were also available on a laptop screen to facilitate the elicitation throughout a session.

3.2.6 Procedure

The recording sessions were mostly interactive in nature, drawing from a range of elicitation techniques. The key to eliciting Plastic Mandarin was to create a relaxing environment where informants speak naturally as they normally do with their friends in everyday life. This translates to two aspects of the design of elicitation. First, the setting of a peer group pair for each recording session was established, as Labov (1972, p. 115) stated, "to allow the interaction of actual peer group itself to control the level of language produced." Second, an informal context in which less attention is paid to speech was preferable. The more formal the situation, the less likely students are to speak a non-standard accent, as they have been trained to speak Standard Mandarin in formal settings for years through education. Traditional passage reading tasks such as Deterding (2006) were avoided because such literary activities are not typical in the realm of Plastic Mandarin speaking. On the other extreme, speech data from completely undirected conversations would be hard to compare across subjects and language varieties, given that phonetic realisation is influenced by various factors. Hence, verbal games embedded with carefully designed speech stimuli were devised for the purpose of acquiring controlled yet spontaneous speech.

This study adopted a four-step procedure for each recording session, illustrated briefly in Figure 3.4, beginning with warm-up informal open conversations (A) and

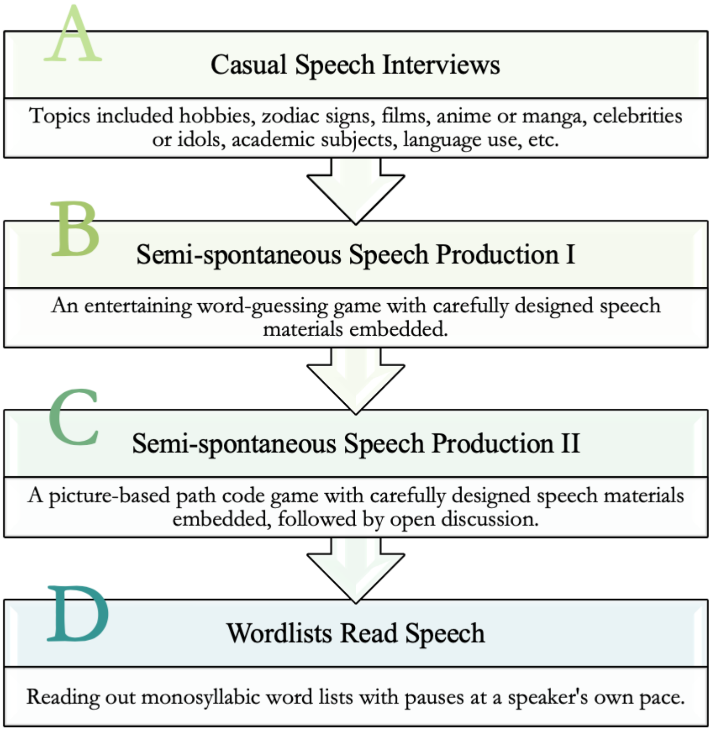


Figure 3.4: Flow chart of the recording sessions. A, B, C, D are the task code.

progressing to verbal games (B and C) and, lastly, reading out word lists (D). In this way, datasets featuring different types of speech were collected. The Plastic Mandarin subjects and Standard Mandarin subjects followed the same procedure but spoke in their respective language. As an audience and the moderator of the recording sessions, I used Standard Mandarin throughout the elicitation sessions of Standard Mandarin, and tailored my own accent to Plastic Mandarin in the course of the Plastic Mandarin sessions. This was to prevent Plastic Mandarin subjects accommodating towards a standard accent given my presence. Before the recording started, I greeted both subjects, attached a microphone to each subject, tested the recorder, and distributed the instruction sheets. I was in charge of the display of instructions and experimental prompts on the laptop screen throughout a session.

Before subjects left the room, they had chances to ask questions and know more about the research project. All recording sessions ranged in length from forty minutes to one hour.

This chapter concentrates on citation tones and reports on the word lists read speech, collected from the last task (D) of a recording session. The procedure of this task is further introduced in subsection §3.2.6.2. The ice-breaking task A is also described in more detail in subsection §3.2.6.1. Task B will be described in detail in chapter 5. The details of Task C are presented in Appendix B.2. Not all speech materials were analysed in this study, given the time and funding constraints.

3.2.6.1 Casual interview

Each session started with a brief casual interview, during which the pair of participants took turns to interview each other, guided by the questions and topics presented on the laptop screen. Despite being short, these warmup conversations were hierarchically structured, as suggested in Tagliamonte (2006), and mainly centred around lighthearted topics that appealed to teenagers and young adults. The topics ranged from individual traits such as nicknames and zodiac signs to personal pastimes and preferences such as favourite films, celebrities, and characters, before progressing into academic subjects and a language module in the end, where the subjects shared their observations of language use and opinions. The language module was intended to "obtain a record of overt attitudes towards language" (Labov, 1984, p. 33).

3.2.6.2 Word lists

The last step of the recording session was to capture subjects' pronunciation of words in isolation. In Labov (1964)'s seminal work of English in New York city, the minimal pair task at the end of the spectrum of formality was introduced last. In a similar vein, this task was administered after the free conversations and verbal games, in order to prevent subjects from focusing their attention to speech too soon. As they became more comfortable and familiar with the recording setting, the elicitation formality was increasingly elevated, yielding more self-conscious and careful speech in the end.

In this task, the subjects were asked to enunciate lists of monosyllabic words of all tones slowly. They were instructed to repeat each monosyllabic word six times, with pauses in between, and they were self-paced without interruptions by the experimenter. The instructions and monosyllabic stimuli were presented on a laptop screen in written Chinese. All utterances of word lists of a participant were recorded in one audio file.

This is a straightforward task for Standard Mandarin speakers, who have explicit knowledge of the four lexical tones of Standard Mandarin due to years of formal learning and training in schools. Plastic Mandarin speakers, however, were not trained to be aware of the citation tones of Plastic Mandarin since there have not been established conventions, despite being native speakers. Instead, they have undergone the same training of Standard Mandarin too. When reading out aloud syllables in isolation, they might have habitually switched to Standard Mandarin. Plastic Mandarin subjects were, thus, encouraged to think of or think aloud a word, a phrase, or any utterances that contain the target syllable in Plastic Mandarin before articulating the monosyllables.

Plastic Mandarin subjects who are native speakers of Changsha enunciated a separate set of Changsha monosyllabic words. There was a short break between pronouncing the Plastic Mandarin set and Changsha set. Subjects were given the freedom to choose which set to start.

3.2.7 Materials

The *Word lists* task was devoted to the comparative analysis of citation tones in Plastic Mandarin, Changsha and Standard Mandarin. This subsection presents the word lists used in this task. Speech materials used in Task B and Task C are presented in chapter 5 and Appendix B.

The impressionistic summaries of Plastic Mandarin tones (e.g. Y. Wu, 2005; Jing and Niu, 2010; Y. Wang, 2018), and acoustical analyses (e.g. C. Xu, 2018; S. Shen, 2019) had shown that Plastic Mandarin has the same tone categories as Standard Mandarin but the tone patterns of Plastic Mandarin were different from those of Standard Mandarin. These studies compared Plastic Mandarin to Standard Mandarin and Changsha, but their discussion of Standard Mandarin and Changsha tones was based on second-hand data in the literature instead of empirical data elicited in a similar controlled environment.

This study, therefore, adopted a more thorough design by including four sets of monosyllables that differ only in tone in the materials for the elicitation of Mandarin tones (see Table 3.3) and the corresponding four sets of syllables in Changsha (see Table 3.4). These sets of Mandarin syllables were selected because their Changsha counterparts (cognates) have similar segmental composition. Three main vowel contexts of Mandarin including /i/, /u/, and /a/ were considered and the majority of the syllables were fully voiced in order to avoid the potential f_0 perturbation brought up by voiceless consonants. One 'minimal quadruplet' set [tiː] with a voiceless unaspirated stop onset was included to allow us to detect the possible effects of an initial voiceless consonant. The same materials displayed in Table 3.3 were used for the elicitation of both Mandarin varieties. It is worth mentioning that participants were only presented with the **Chinese orthography without other information such as pinyin or tone categories**.

Vowel	IPA	Tone			
		T1	T2	T3	T4
/i/	[iː]	*yī* 衣	*yí* 姨	*yǐ* 蚁	*yì* 意
		clothes	aunt	ant	desire
	[tiː]	*dī* 低	*dí* 笛	*dǐ* 底	*dì* 地
		low	flute	bottom	earth
/u/	[uː]	*wū* 屋	*wú* 吴	*wǔ* 武	*wù* 雾
		house	Wu	martial	fog
/a/	[maː]	*mā* 妈	*má* 麻	*mǎ* 马	*mà* 骂
		mother	hemp	horse	curse

Table 3.3: Speech materials for the citation tones of both Standard Mandarin and Plastic Mandarin. The romanised transcripts here used the Pinyin system that is based on Standard Mandarin.

In Table 3.4, Standard Mandarin-based Pinyin system was adapted to represent these Changsha syllables in romanised form, but diacritics denoting tones were removed. The selection of these syllables was based on R. Li (1998). As Changsha retains the register distinction for the qù (III) tone (Norman, 1988), most syllables in the T4 group have voiceless consonant onsets or zero onset, meaning that the [maː] set with a sonorant onset, widely cited in Standard Mandarin research, might be inadequate to cover all six tones in Changsha (tonotactic gap). The syllable [pʰaː] that contains a voiceless stop onset with the same place of articulation as [m] and the same vowel as [maː] was hence included. It is relevant to note here that, although the same set of phonemes was employed in the IPA transcription of the Mandarin and Changsha syllables, the vowel quality of /a/ in Changsha, for instance, tends to be distinct from /a/ in Standard Mandarin ([a]). /a/ is sometimes realised in a more retracted and/or raised manner (e.g. [a̠] or [ɔ]) by some Changsha speakers.

Vowel	IPA	Tone					
		T1	T2	T3	T4	T5	T6
/i/	[iː]	*yi* 衣	*yi* 姨	*yi* 蚁	*yi* 意	*yi* 易	*yi* 一
		clothes	aunt	ant	desire	easy	one
	[tiː]	*di* 低	*di* 题	*di* 底	*di* 帝	*di* 地	*di* 笛
		low	question	bottom	emperor	earth	flute
/u/	[uː]	*wu* 乌	*wu* 吴	*wu* 武	*wu* 务	*wu* 雾	*wu* 屋
		black	Wu	martial	task	fog	house
/a/	[maː] (*[pʰaː])	*ma* 妈	*ma* 麻	*ma* 马	**pa* 怕	ma 骂	*ma* 抹
		mother	hemp	horse	fear	curse	wipe

Table 3.4: Speech materials for Changsha citation tones.

In addition, a set of toneless neutral tone syllables, as shown in Table 3.5, were also incorporated in the Mandarin materials, to explore how speakers produce the citation form of neutral tone syllables. All of these are function words, which usually do not occur stand-alone. Since Changsha has a distinct set of function words with different segmental structures, no Changsha counterparts were considered in this study.

[ma]	[də]	[lə]	[ʈʂə]	[mən]
ma 吗	*de* 的	*le* 了	*zhe* 着	*men* 们
INT	GEN	PRF	DUR	PL

Table 3.5: Extra neutral tone syllables in the Mandarin speech materials.

3.3 Acoustic Preprocessing

This section presents the stages and details of acoustic prepossessing for the monosyllabic speech data in this study[2]. The main acoustic features examined in this thesis are timing (syllable duration) and pitch (f_0) patterns.

3.3.1 Transcription

The raw recordings from the *Word lists* task were two-channel WAV files at a sampling rate of 44.1 kHz with 16-bit, signed integer PCM encoding, containing pronunciation of all repetitions of the monosyllables in the speech materials of a given language variety by a speaker. The main channel for each speaker was extracted and saved as a mono audio file. The audio was transcribed in Chinese orthography, where the pauses between monosyllables were represented by spaces between Chinese characters.

3.3.2 Forced alignment

The audio and transcript files were then fed to the Penn Phonetics Lab Forced Aligner (P2FA) for Mandarin Chinese (Yuan and Liberman, 2008; 2015) to achieve automatic phonetic segmentation. Prior to the forced alignment, the audio files were downsampled to 16 kHz to prepare for the use of P2FA. The output of the aligner was a TextGrid file for each audio file, in which the phones and syllables (represented by Chinese characters) were aligned to the audio. This study is chiefly interested in the syllable tier, from which we can identify where in the audio a syllable begins and ends.

[2]All relevant scripts are available as online supplemental materials at https://osf.io/ufhzt/.

Although P2FA's acoustic models were trained on Standard Mandarin, it can still be applied to Plastic Mandarin since the phoneme inventories of the two varieties are highly similar. Existing forced alignment systems even had potential for facilitating the segmentation for under-resourced languages that were different from the training language (DiCanio et al., 2013). P2FA here was also adapted to process the the monosyllabic Changsha data, by constructing a separate pronunciation dictionary for the Changsha syllables with customised grapheme to phoneme rules coded using the phone set specific to P2FA. After the visual inspection of the output TextGrids of some trial runs, it can be concluded that P2FA for Mandarin Chinese was successfully utilised for Plastic Mandarin and Changsha.

3.3.2.1 Problems in forced alignment

The TextGrid files generated by P2FA were verified against their corresponding audio files. The silence-to-phoneme and phoneme-to-silence transitions were our main concerns. The majority of the alignment were acceptable, but three main types of misalignment or errors were identified and marked:

1) Extended interval due to reverberation;
2) Misplaced interval due to environmental noise;
3) Truncated interval.

These are exemplified in Figures 3.5 to 3.7 of Praat editor interface, in which the waveform and wide-band spectrogram of an audio signal are presented together with the text annotations and boundaries generated by P2FA.

In Figure 3.5, both syllables of 的 *de* have a much delayed right boundary. In the spectrogram, the light grey area with faint formant patterns following each syllable, labelled with a schwa (represented by '&' in the phone tier), in fact reflects some reverberation captured by the microphone, which should not be identified as part of the vowel portion of the syllable. In Figure 3.6, an extra-short text interval 屋 *wū* in the second tier, highlighted in yellow background, mistakenly corresponds to some noise in the audio. Moreover, this resulted in subsequent misalignment in the remainder of the audio because the text labels were no longer matched with

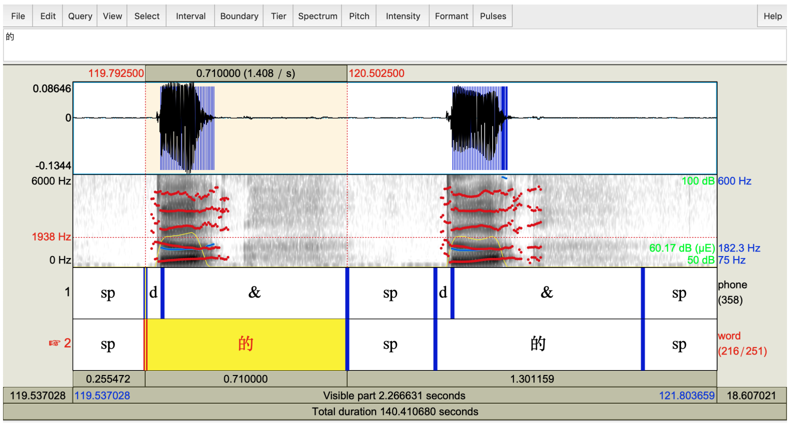


Figure 3.5: Expanded intervals of the syllable 的 *de* GEN by P2FA. One of the intervals is highlighted in yellow background in the word tier.

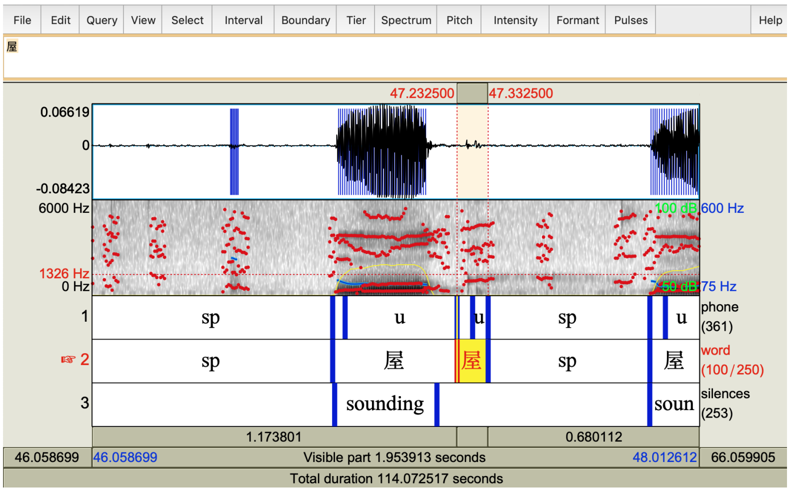


Figure 3.6: Misplaced boundaries of the syllable 屋 *wū* 'house' by P2FA. The problematic interval is highlighted in yellow background in the word tier.

the audio. In Figure 3.7, the left boundary of the syllable 笛 *dí* failed to cover the whole interval between the release of the unaspirated stop consonant [t], signalled by the burst spike followed by aperiodicity in the waveform, and the voice onset. Some other cases of truncated intervals include the right boundary being placed too early, cutting off the creaky end of a syllable.

3.3.2.2 Improving forced alignment for monosyllables

The first two types of misalignment tend to result from the presence of reverberation or noise in the recording room. The alignment can thus be improved by removing such reverberation and noise during pauses of speech. Pauses, or non-speech

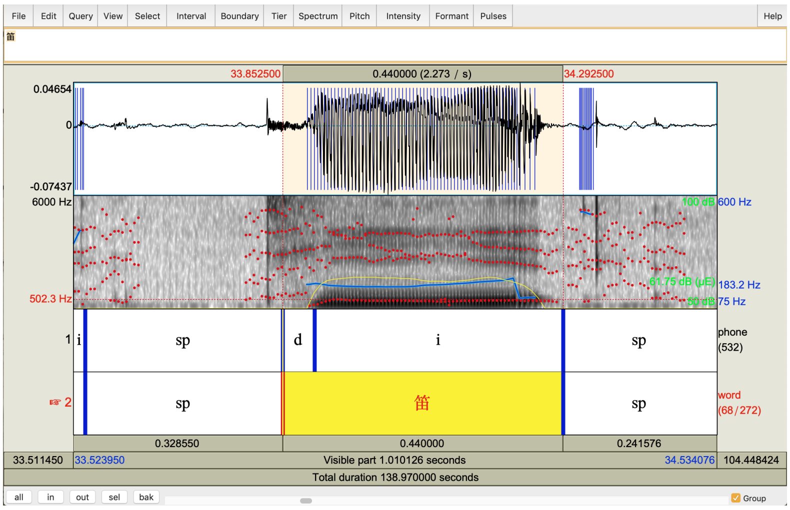


Figure 3.7: Contracted interval of the syllable 笛 *dí* 'flute' by P2FA. The interval is highlighted in yellow background in the word tier.

intervals, were first identified using Praat's algorithm *Sound: To TextGrid (silences)* and then muted automatically. The algorithm determined silent intervals from the intensity curve of the bandpass filtered (between 80 and 8000 Hz) waveform of a copy of the sound, and the silence threshold ranging from −40 dB to −30 dB was used depending on the intensity peaks of the audio[3]. The output of this algorithm was a tier of silence intervals in a TextGrid file, such as the third tier *silences* in Figure 3.9, in which the empty intervals were identified as non-speech intervals and muted. The resulting audio was double-checked to ensure that speech-related aperiodic intervals, which were relatively lower in intensity than periodic intervals, were not muted; otherwise the parameter settings of the algorithm were adjusted accordingly. The speech intervals remained untouched. Subsequently, P2FA was run the second time on the edited audio. The newly-obtained TextGrid files were compared to the previous ones, and as a result the alignment of vowel-only syllables such as [i] and [u] was improved, though syllables with a nasal onset [ma] in the new TextGrids tended to have slightly wider intervals. Finally, the most optimal TextGrid files were created by replacing the misaligned intervals and intervals of vowel-only syllables in

[3]The parameters used for each audio file were logged in a text file, available in the online supplementary materials.

the original TextGrid files with corresponding intervals with improved alignment. By doing so, the aforementioned phenomena of misalignment were largely eliminated.

The main motivation to improve the aligner's performance in above approach is that it is systematic, consistent, and reproducible, compared to hand corrections which are inherently subjective (Baghai-Ravary et al., 2009). This also leads to consistent measurement of syllable duration using an automatic aligner.

3.3.3 Sound interval extraction

3.3.3.1 Syllable timing database

The syllable (word) tier of each TextGrid file was converted into a more reader-friendly plain text (TXT) file where each row contained temporal information of a syllable. All the plain text files were assembled into a comma delimited CSV file to store the large tabular data. Rows of irrelevant syllables, non-speech, and silent intervals were removed from the CSV file, while metadata including the speaker ID and the enumerated sequence of syllable repetitions were added as columns. It was found that some participants occasionally repeated some syllables only five times or more than six times. This syllable timing database was further used for durational analysis.

3.3.3.2 Trimming and zero-padding

To extract the target syllables from the recordings, sound trimming scripts were created utilising the *trim* command from SoX and time stamps of sound intervals from the TextGrid files, and each output WAV file of a syllable was assigned a unique filename rich in contextual information (e.g. *dp101衣1s1.7025f2.2825*, which consists of the task code, speaker ID, Chinese character, enumerated sequence, starting and ending time stamps).

Sound intervals with finely-aligned boundaries, in particular short intervals, were not the most suitable for frame-based acoustic measurements such as f_0 and intensity, given the analysis window. When the minimum periodicity frequency is set to be the default 75 Hz, Praat (Boersma & Weenink, 2017) does not compute

f_0 values in the first or last 20 ms of each sound interval because the analysis requires a 40 ms window for every pitch frame. Likewise, the effective analysis window for intensity analysis is about $3.2/75 = 42.67$ ms[4] (Boersma & Weenink, 2017). Therefore, 50 ms of silence was added to both edges of each sound interval to prevent such information loss due to windowing.

These zero-padded speech tokens constitute the audio data for subsequent analysis. Tokens with intermittent laughter or overlapping speech were discarded. The slight imbalance in the number of repetitions can be taken care of in the statistical analysis.

3.3.4 Acoustic measurement

Apart from the syllable duration measurement from the forced aligner, f_0 was measured using Praat's algorithm *Sound: To Pitch*, which is based on an autocorrelation method (Boersma, 1993). The pitch floor and pitch ceiling of the algorithm were set to 75 Hz and 550 Hz respectively. The choice of a wide f_0 range was because teenager participants reached very high pitch, especially for high tones, during the elicitation. To facilitate such a process for over a large number of audio files, a Python script was created employing `Parselmouth`, a Python interface to Praat (Jadoul et al., 2018).

Two sets of f_0 estimates were obtained for each syllable via Python scripts: Set A, f_0 estimates in every 10 milliseconds (ms) throughout the voiced region; Set B, f_0 estimates at 21 equidistant points, which are at intervals of every 5% of the same region. Set A was the main dataset for f_0 curve fitting, which will be introduced in depth in §3.4.3, and Set B was used for initial exploratory data visualisation since it made f_0 data comparable across utterances at various time points.

The data of f_0 contours were stored in tabular format: in each row, an f_0 value corresponds to a time in second in Set A, and to a normalised time index ranging from 1 to 21 in Set B[5]. The corresponding intensity value in dB was also available

[4] See Praat manual https://www.fon.hum.uva.nl/praat/manual/Sound___To_Intensity___.html

[5] The f_0 data are in CSV files and available in the online supplementary materials.

in the same row. The metadata or contextual information coded for each sound token in §3.3.3 were then added to the databases of f_0 contours.

Manual scrutiny and correction of the raw f_0 data was an extremely labour-intensive process. The following subsections delineate how I identified and tackled f_0 measurement problems in the monosyllabic datasets using an innovative and semi-automated approach.

3.3.4.1 Problems of automatic f_0 measurement

Non-modal voice, especially creaky voice, tended to be the leading cause of problematic or unreliable f_0 measurement. For example, the f_0 estimates of the edges of many speech tokens often had a sudden jump to much higher frequencies, deviating from the trend of the main body contour. Examining the sound waveform and spectrogram of such tokens revealed that the upsurge in f_0 at the right edge was likely to result from inaccurate f_0 estimation of creaky phonation at the end of a syllable, related to the termination of an utterance. Creaky voice may be at much lower frequencies, and cannot be properly estimated when the f_0 values are below the pitch floor setting of 75 Hz. In this case, Praat might instead pick out small individual peaks in the highly damped long pulses of creaky voice, yielding substantially high f_0 estimates. Figure 3.9 in §3.3.4.2 demonstrates the high f_0 estimates (approximately 500 Hz) of the creaky voice at the end of a Plastic Mandarin syllable *wù* [u] 'fog' in black solid line. Without the setting of a much lower pitch floor, the f_0 estimates of a creaky voice portion are highly likely to be problematic.

Creaky voice is pervasive in the recordings. There are tokens of which almost the entirety is in creaky voice with some irregular pulses, and tokens that are partially or fully diplophonic in nature. Figure 3.8 presents a Standard Mandarin syllable *yǐ* [i] 'ant' with creaky voice. The middle of the syllable portrays low rate of vocal fold vibration and irregular f_0 pulses, and the ending part is also damped and irregular pulses. A large percentage (64 %) of the Tone 3 syllables in Standard Mandarin have creaky voice. Creaky voice is not an aberration but a canonical feature of Standard Mandarin Tone 3. It is, however, not exclusive to Tone 3, but

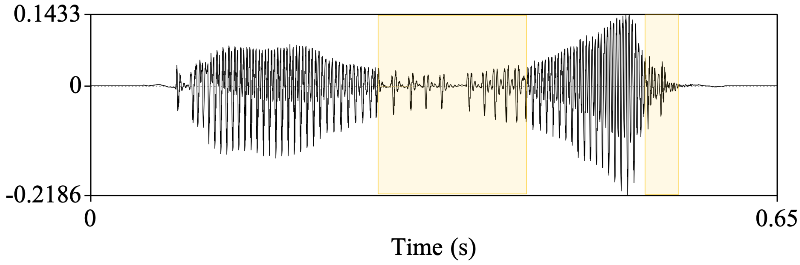


Figure 3.8: An example of irregular creaky voice pulses in the middle of the syllable *yǐ* in Standard Mandarin. The creaky voice portions were highlighted.

likely to occur whenever there is a low pitch target (Kuang, 2017). Such allophonic phonation variation in the production of tones added difficulties in pitch tracking. Even if Praat managed to generate accurate low f_0 estimates of some regular creaky pulses in an utterance, the large plunge in f_0 was usually not perceived as large pitch drop but coarseness in the voice. Such a large plunge would greatly influence the subsequent functional modelling of the f_0 curve.

Another cause of unreliable f_0 measurement was voiceless obstruents. Occasionally, there were exceptionally high f_0 estimates in intervals of burst or frication noise of an obstruent consonant.

3.3.4.2 Trimming f_0 measurements

Previous acoustic studies of f_0 often trim an arbitrary portion of the measurements at the edges of the vowel or syllable nuclei, to reduce f_0 perturbations from neighbouring consonants or in the initiation and termination of the speech. For example, in Keating and Kuo (2010), the f_0 measurements for the first and last 2% of each target interval were discarded; in Tang *et al.* (2019), the initial and final 5% of the vowel were excluded; in Rose (1987), it was the first and last 10% of the duration of each tone track; and in Stanford (2008), as much as 25% of the beginning portion of a tone token was trimmed. Removing a portion of contour edges can be helpful in excluding unwanted f_0 estimates. It is, however, purely arbitrary to decide how much f_0 values to omit and the optimal amount might vary among different utterances.

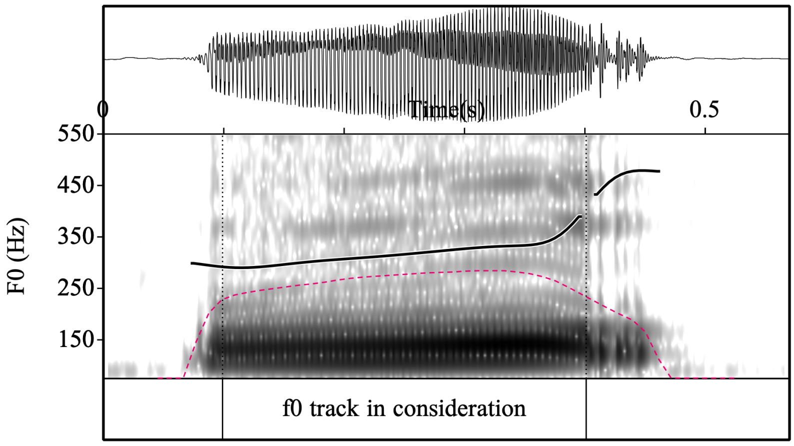


Figure 3.9: An example of trimmed f_0 track (black) of the syllable 雾 *wù* in Plastic Mandarin, marked in the text tier. f_0 track segments outside the marked interval were removed.

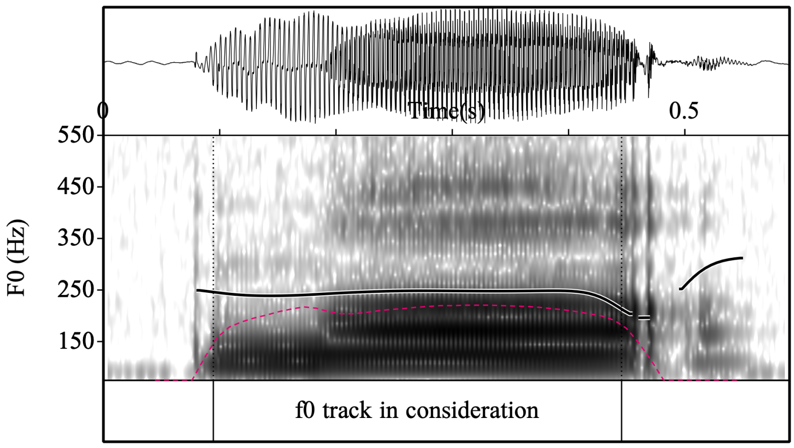


Figure 3.10: An example of trimmed f_0 track (black) of the neutral tone syllable 吗 *ma*, marked in the text tier. f_0 track segments outside the marked interval were removed.

This study adopted an intensity-dependent approach to trim the edges of f_0 tracks of monosyllables, whereby the intensity was measured along with each f_0 point. For monosyllables, the intensity curve usually resembles an inverse U-shape, where the beginning and end have a steep rise and a plunge respectively and the middle portion reaches a plateau (see the pink dashed line in 3.9). My algorithm filters out f_0 measurements in edge intervals of low intensity, by excluding f_0 estimates whose intensity falls below one standard deviation below the mean intensity of the token.

Empirically tested, this method was conducive to excluding inaccurate f_0 estimates at edges as well as attenuating effects attributed to a neighbouring voiceless consonant, since its intensity is normally lower than a vowel. Figure 3.9 and Figure 3.10 illustrate the trimmed f_0 tracks determined by the proposed algorithm. In both cases, a few voicing pulses were excluded at the beginning, as were the creaky pulses at the end of the vowel in Figure 3.9 and the reverberation pulses at the end of the syllable in Figure 3.10. This algorithm, though, was not perfectly effective. The presence of high-intensity creaky voice portions in a syllable may result in the inadvertent inclusion of inaccurate f_0 estimates. Thus, further checks were required.

3.3.4.3 Dealing with problematic f_0 tracks

The following procedure was conducted to identify likely problem areas in the trimmed f_0 tracks of monosyllables. Bespoke methods for addressing different f_0 tracking problems will then follow.

First, the mean, standard deviation, minimum, and maximum of f_0 for each token were calculated. Tokens with a large standard deviation (larger than 40 Hz), a low minimum (lower than 80 Hz), or a high maximum (higher than 350 Hz) were flagged for further check of their waveform and spectrogram. If f_0 values in a token were close to the pitch floor, we re-ran the f_0 tracking algorithm for this token and lowered the pitch floor setting so that f_0 estimates lower than 75 Hz were computed.

Second, extra short f_0 tracks or tracks with a large portion being removed were checked. The algorithm might have trimmed edges more than we wanted sometimes, especially when all first-pass f_0 measurements in a short token are “good”. In such

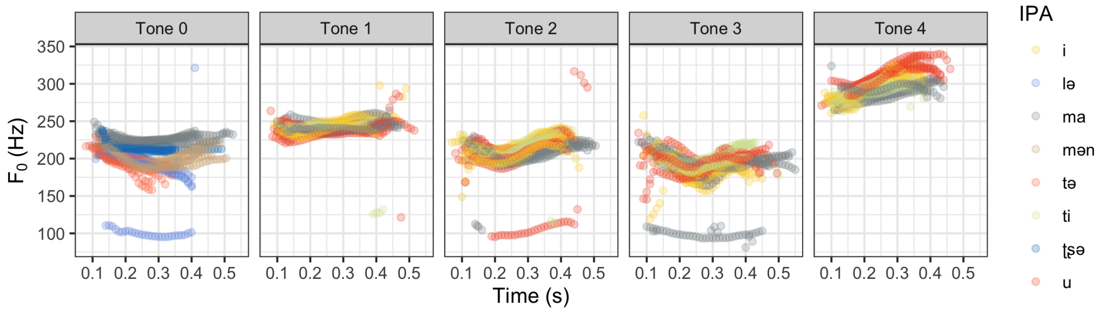


Figure 3.11: Demonstration of visual inspection of f_0 tracks of all syllables by Speaker 112 in Plastic Mandarin.

cases we skipped the trimming step in order to preserve these very limited good f_0 estimates in a track and prevent significant information loss for short utterances.

Third, f_0 tracks of each speaker grouped by tone categories were plotted in semi-transparent colours in an overlaying manner and visually inspected. This was needed because problems such as octave shifts were not easily identified using previous methods. As demonstrated in Figure 3.11, the whole f_0 track of three syllables including one [lə] in neutral tone, one [u] in Tone 2, and one [ma] in Tone 3 deviate from most of the f_0 curves in their tone categories. Their f_0 estimates are approximately half of the f_0 of other repetitions of the same syllables. These were highly likely to be pitch-halving errors or diplophonia. Moreover, other potential f_0 outliers such as sudden jumps to over 300 Hz or drops to below 150 Hz are markedly shown in Figure 3.11.

These remaining potential outliers and artefacts were further examined and manually corrected in Praat, via editing the f_0 points (the Pitch object) in the Pitch editor in Praat. For diplophonic or double pulsed utterances, the high f_0 values in the two concurrent periodicities were included because they resembled the f_0 values of the modal voice counterparts, although the resulting percept of diplophonic voice is often of an indeterminate pitch plus roughness (Keating et al., 2015). As demonstrated in 3.12, the path of pink disks indicate the raw f_0 estimates around 100 Hz, which is also signalled by the blue marks flanking each glottal pulse in the waveform. In the selected region highlighted in light yellow background, particularly from 0.11 s to 0.29 s, there is another line of digits around 200 Hz

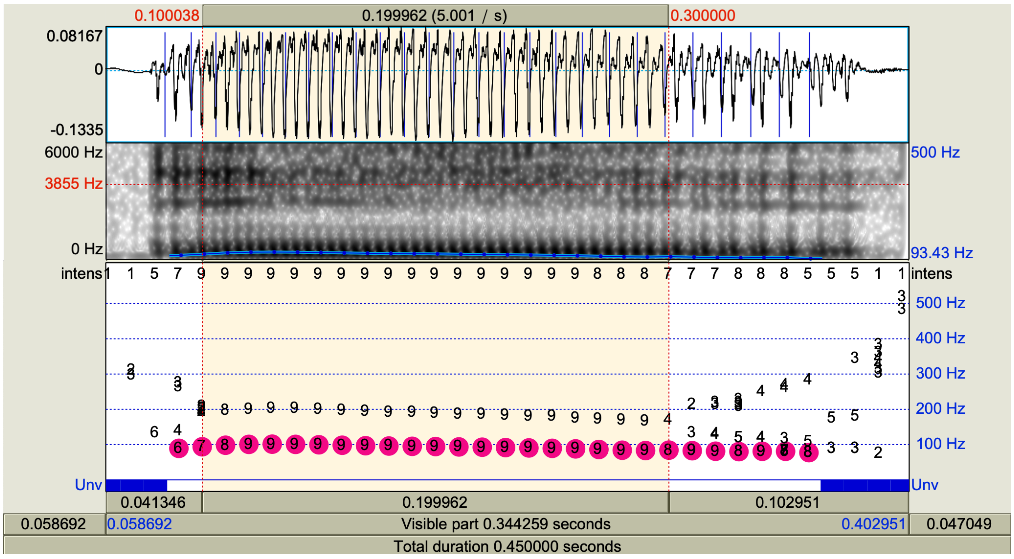


Figure 3.12: Editing f_0 estimates of 易 *yi* in Changsha by Speaker 117 in Praat. The upper panel is the oscillogram and spectrogram. The lower panel is the PitchEditor window in Praat. The digits in the PitchEditor represent the goodness of the candidate. Original pulse-marking is in blue vertical bars in the oscillogram.

with the same strength attributes regarding the degree of periodicity (mostly 9). The alternative contour is in line with f_0 estimates of other repetitions of the same syllable. This highlighted portion is highly likely to be diplophonic and the syllable-final portion after 0.3s is creaky too. For this token, the pink disks in selection were moved an octave up, which was the upper path of digits around 200 Hz, and the remaining f_0 estimates were discarded. Other adjustment of the Pitch object in the PitchEditor involved marking a few irregularly voiced pulses as unvoiced to exclude inaccurate f_0 estimates.

After modifying the Pitch object, a Praat script was created to re-extract the updated f_0 values in the desired time domain, for instance, 0.11s-0.29s in Figure 3.12. By manually setting the time domain for f_0 track of extra short utterances, overly trimmed edge f_0 values can be restored. Monosyllabic tokens with a large portion of aperiodic creaky voice without reliable f_0 data were excluded in the pitch analysis. It should be noted that mostly f_0 estimates of modal voice portions were included and not all f_0 contours were continuous. f_0 tracks may have a few missing points in the middle.

Despite the use of custom-written scripts to facilitate the identification and trimming of f_0 measurement errors, manual checking and editing was still needed and this process was extremely time-consuming. In all, about 32% of the Plastic Mandarin, 36% of the Changsha, and 13% of the Standard Mandarin monosyllables had minor adjustment of f_0 estimates, among which about 80% had irregular f_0 values arising from creaky or diplophonic voice and were not removed or corrected by the algorithm. The recording environment may have influenced the audio quality and f_0 measurement, since the Standard Mandarin data that were collected in the professional sound studio required less manual work.

3.4 Methods

Functional data analysis has been a valuable and fruitful collection of tools in various fields (Ramsay & Silverman, 1997). The f_0 data registered on a continuum of time can be considered as one-dimensional functional objects, curves. Hadjipantelis et al. (2012), for instance, treated f_0 contours of Mandarin syllables as bounded continuous curves. In this chapter, f_0 contours were represented by smooth functions which explained the most variation. Following Grabe et al. (2007), parametric orthogonal basis functions, Legendre polynomials, were employed as basis functions. Other alternatives include quadratic spline basis (Hirst & Espesser, 1993), discrete cosine transformation (DCT) (e.g. Watson and Harrington, 1999; Graham et al., 2016), or functional principal component analysis (FPCA)(Gubian et al., 2015). Nevertheless, Legendre polynomials were utilised because the f_0 contour of a single syllable, which was not expected to be very wiggly, can be easily represented by a few number of **interpretable and meaningful** coefficients. The coefficients were served as proxy data for the f_0 curves and used as the dependent variables in a series of linear mixed effects models.

This section summarises the data, introduces the workflow of data analysis, and outlines statistical methods, with a focus on f_0 curve data. The specific statistical models pertaining to a particular research question and dataset will be presented in the following result sections.

3.4.1 Data summary

There are 2656 Plastic Mandarin tokens, 1765 Standard Mandarin tokens, and 1853 Changsha tokens in total. Tables 3.6 and 3.8 present a breakdown of the number of speech tokens by categories of lexical tones and syllables in the datasets of Plastic Mandarin, Standard Mandarin, and Changsha. Table 3.7 presents the number of neutral tone tokens in Plastic Mandarin and Standard Mandarin. Apart from the metadata, acoustic measurements including the syllable duration, the intensity and f_0 trajectories are available for each token. The tables show that the datasets are slightly imbalanced, which should inform our choice of statistical analysis.

IPA	Tone Category				Count
Variety	T1	T2	T3	T4	
[iː]	衣	姨	蚁	意	
PM	125	124	126 (122)	125	500 (496)
SM	84	84	84 (69)	84	336 (321)
[tiː]	低	笛	底	地	
PM	128	126	126 (121)	127	507 (502)
SM	83	84	84 (74)	84	335 (325)
[uː]	屋	吴	武	雾	
PM	127	127	127 (123)	128	509 (505)
SM	84	84	83 (76)	84	335 (328)
[maː]	妈	麻	马	骂	
PM	122	122 (121)	127 (124)	127	498 (494)
SM	84	84	84 (75)	84	336 (327)
Count (PM)	502	499 (498)	506 (490)	507	2014 (1997)
Count (SM)	335	336	335 (294)	336	1342 (1301)

Table 3.6: The number of lexical tone tokens by categories in the Plastic Mandarin (PM) and Standard Mandarin (SM) monosyllabic dataset. The numbers in brackets indicate the numbers of tokens used for pitch analysis (Set A).

As a Plastic Mandarin speaker, I intuitively know that in terms of tone categories morphemes in Plastic Mandarin are grouped the same way as in Standard Mandarin,

	[ma] *ma* 吗	[də] *de* 的	[lə] *le* 了	[ʈʂə] *zhe* 着	[mən] *men* 们	Count
PM	124	129 (127)	130	129 (128)	130	642 (639)
SM	85	84	84	85	85	423

Table 3.7: The number of neutral tone tokens by categories in the monosyllabic datasets of Plastic Mandarin (PM) and Standard Mandarin (SM). The numbers in brackets indicate the numbers of tokens used for pitch analysis (Set A).

IPA	Tone						Count
	T1	T2	T3	T4	T5	T6	
[iː]	衣	姨	蚁	意	易	一	
	77	77	76	77 (76)	77	77	461 (460)
[tiː]	低	题	底	帝	地	笛	
	80	77	77	76	77	77	464
[uː]	乌	吴	武	务	雾	屋	
	77	79	77	78	76	76	463
[maː]	妈	麻	马	怕*	骂	抹	
(*[pʰaː])	76	78	77	79 (78)	76 (75)	79	465 (463)
Count	310	311	307	310 (308)	306 (305)	309	1853 (1850)

Table 3.8: The number of lexical tone tokens by categories in the Changsha monosyllabic dataset. The numbers in brackets indicate the numbers of tokens used for pitch analysis.

which can be verified by enumerating and classifying a large number of morphemes. This observation is consistent with the findings in previous studies such as Jing and Niu (2010). The Plastic Mandarin data of these 21 young speakers show that pitch contours of morphemes in each column of Table 3.6 pattern together. For instance, morphemes 衣 'clothes', 低 'low', 屋 'house' and 妈 'mother' share a highly similar pitch contour. It worth mentioning that the morpheme 屋 'house' does not pattern with the other three morphemes in tone in Changsha (see Table 3.8). This also indicates that Plastic Mandarin has the same tone categories (grouping of morphemes) as Standard Mandarin.

The number of tokens used for duration analysis and pitch (f_0) analysis are slightly different, given that a small percentage of tokens were produced almost entirely in aperiodic voice and thus had no f_0 measurement. For pitch analysis, only Set A monosyllabic datasets were prepared for statistical modelling, while Set B datasets were used to visualise the distribution of f_0 consecutively among speakers throughout the production of tones. Set A of Plastic Mandarin contains 2636 f_0 contours, and Set B 2618 f_0 contours. Set A of Changsha contains 1850 f_0 contours, and Set B 1848 f_0 contours. Set A of Standard Mandarin composes of 1724 f_0 contours. Set B contains slightly fewer contours because utterances with creaky portions in the middle of the vowel were excluded in Set B, due to the fact that in Set B f_0 at specified equidistant times was likely to be estimated inaccurately using the local linear interpolation method in Praat, susceptible to inaccurate or missing f_0 associated with creaky voice in the middle of a Pitch object. About 64% of Standard Mandarin Tone 3 syllables were creaky in the middle of an utterance, thus Set B of Standard Mandarin was very imbalanced and not considered further.

This study focuses on f_0 trajectories of utterances in modal voice, for the purpose of establishing the prototypical pitch pattern of citation tones. This does not make the excluded tokens in other voice quality less important in our understanding of the natural production of tones. I hope to explore the interaction of voice quality and tone in future research.

The f_0 data of each monosyllabic token were dynamic trajectories instead of measurements at a single time point or aggregated or averaged data over a timespan. The datasets were hierarchical in nature: while individual measurements at various time points were arranged into f_0 contours, the contours can be grouped according to syllables, tones, and speakers, resulting in complex dependencies among individual measurements. Figure 3.13 demonstrates the hierarchical data structure of the combined dataset of Plastic Mandarin and Changsha monosyllables produced by the same group of multilingual speakers. The statistical analysis should not neglect these dependencies among the measurements.

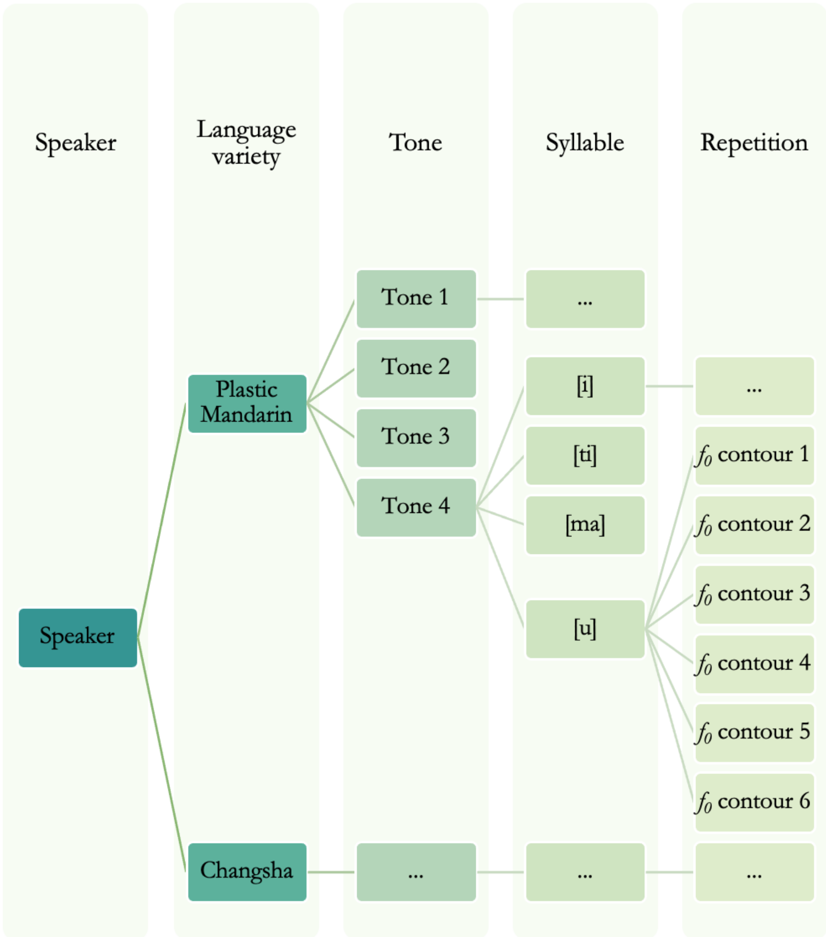


Figure 3.13: Hierarchical structure of the combined dataset of Plastic Mandarin and Changsha monosyllables. Each tier only shows one set of child nodes (others omitted to save space) and all parent nodes in the same tier have the same child nodes. f_0 contours 1 to 6 are the six repeated tokens of each kind of syllable.

3.4.2 Data normalisation

3.4.2.1 Pitch Normalisation

To allow inter-speaker comparison, all f_0 values of a given speaker in a language variety were normalised to semitones referenced on the speaker's mean f_0 in that variety. The speaker mean f_0, $\bar{f}_s$, was defined over the average of the f_0 mean of each tone in a language variety using the corresponding monosyllabic dataset (Equations 3.1 to 3.3).

$$\bar{f}_T = \frac{1}{n}\sum_{j=1}^{n} f_{0j} \tag{3.1}$$

Here $\bar{f}_T$ is the average of all n f_0 values of Tone T syllables in a given language variety by a given speaker ($T \in [0,4]$ for Mandarin varieties; $T \in [1,6]$ for Changsha; Tone 0 refers to the neutral tone). j is the index of each f_0 measurement of a tone T syllable in a given variety by a given speaker.

$$\bar{f}_s = \frac{1}{5}\sum_{T=0}^{4} \bar{f}_T \text{ (Mandarin)} \tag{3.2}$$

$$\bar{f}_s = \frac{1}{6}\sum_{T=1}^{6} \bar{f}_T \text{ (Changsha)} \tag{3.3}$$

In this way, the speaker mean $\bar{f}_s$ was not biased due to an imbalanced number of tokens of each tone in the dataset. The $\bar{f}_s$ value was in fact very close to the simple average of all f_0s of a speaker ($\bar{f}_0$) in the monosyllabic dataset. Figure 3.14, for instance, presents the histogram of f_0s by each speaker in the Plastic Mandarin data, in which the two means (i.e. $\bar{f}_s$ and $\bar{f}_0$), indicated by dotted vertical lines in different colours, were mostly overlapping. The differences in Hz between the two means were less than 3 Hz for all but two speakers.

Then, each f_0 value in Hz was transformed into semitone using Equation 3.4:

$$f_0^* \text{ (semitone)} = 12\log_2\left(\frac{f_0}{\bar{f}_s}\right) \tag{3.4}$$

The reference frequency in this semitone transformation equation is the speaker mean $\bar{f}_s$, thus an f_0 equates to the defined speaker mean when its semitone counterpart is zero ($f_0^* = 0$). Equation 3.4 is essentially the semitone difference between an f_0 value and the speaker's mean f_0. Semitones can be used to approximate the perceived pitch (Nolan, 2003). It should be noted that although the musical scale has psychophysical validity for pure tones, it does not directly measure the perceptual pitch of the speaking voice (Menn & Boyce, 1982). This normalisation method was considered the best among many others in preserving phonemic variation as well as sociolinguistic variation, according to J. Zhang (2018).

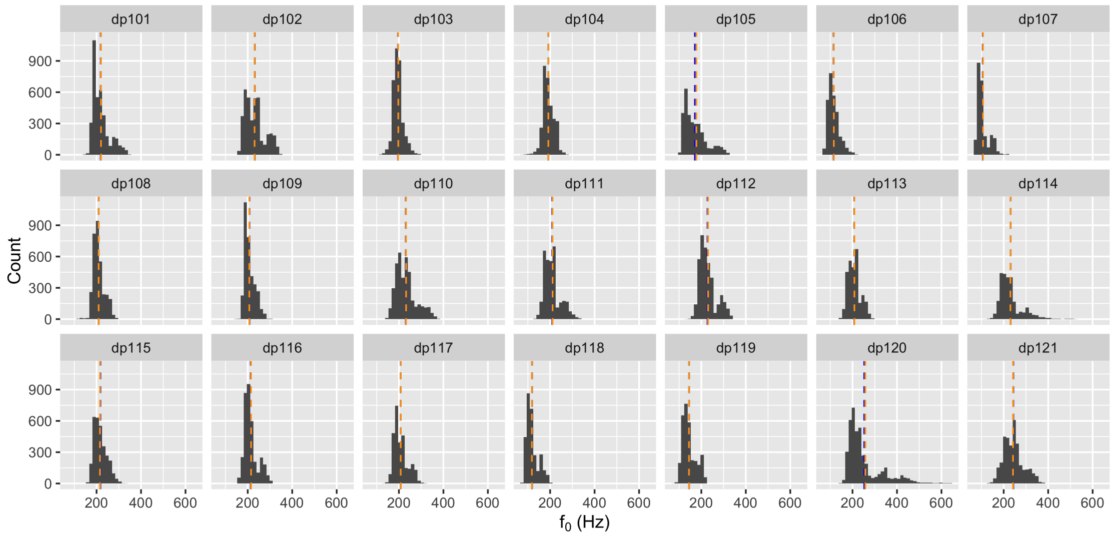


Figure 3.14: The distribution of f_0s of all speakers and comparison of two calculations of the speaker's mean. The defined $\bar{f}_s$ is marked in orange and the simple average is in blue.

3.4.2.2 Time Normalisation

The normalisation of time facilitates the inter-token comparison. In order to prepare data for functional modelling using Legendre polynomials, the corresponding time measure t_i of each f_0 of a token, originally measured at 10 millisecond intervals, was linearly scaled to the range of -1 to 1:

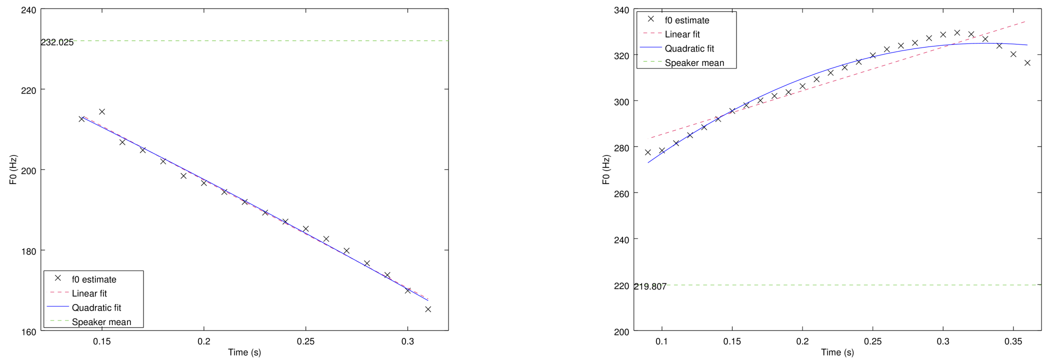
$$t_i^* = \frac{2(t_i - t_1)}{t_m - t_1} - 1 \tag{3.5}$$

Here $i \in [1, m]$ is the index of all sample times of a token, and t_m represents the final time measure, which is also the maximum time value. The normalised time axis $[-1, 1]$ was required because Legendre polynomials were defined over the interval $[-1, 1]$.

An f_0 curve, frequencies in Hz in the course of time $f_0(t)$, is now recast as the function $f_0^*(t^*)$ in the normalised domain.

3.4.3 Dimension reduction

Polynomials have been employed to model, smooth, categorise, and reconstruct f_0 contours in speech. S. H. Chen and Wang (1990) adopted the first four orthogonal



(a) [lə] ‘PERF’ by speaker 110 **(b)** [u] ‘fog’ by speaker 101

Figure 3.15: Examples of best-fit polynomial models for Plastic Mandarin utterances.

Legendre polynomials to transform, encode, and reconstruct the pitch contour of Mandarin segments. In Andruski and Costello (2004), both linear and quadratic polynomial models of pitch contours achieved satisfying results in accounting for the differences among the three low falling tones in Green Mong. Grabe et al. (2003) found that the first two orthogonal polynomial coefficients effectively differentiated utterance types such as declaratives, *wh* and yes/no questions, and declarative questions in seven urban English dialects, and that the contribution of the coefficients of higher terms was marginal. Their later paper (Grabe et al., 2007) adopted the first four coefficients of Legendre polynomials to model and reconstruct seven nuclear accents in English intonation phonology. Aston et al. (2010)’s analysis of Luobuzhai Qiang using non-parametric curve estimation FPCA suggested that a polynomial basis would be a good representation of the f_0 data, which was determined from the eigenfunctions post processing.

In a similar way that Hadjipantelis et al. (2012) modelled a f_0 curve as realisation of a stochastic Gaussian process and projected them in the lower dimensional space where the FPC coefficients served as the axis system, this study projected f_0 curves in the lower dimensional space where the Legendre polynomial coefficients served as the axis system. Following the approach in Grabe et al. (2007), quadratic polynomials of the form $a_2x^2 + a_1x + a_0$ were first motivated for modelling the monosyllables. As shown in Figure 3.15 (a), the linear and quadratic models are very similar for the neutral tone syllable [lə] ‘PERF’, but the quadratic model in (b) gives a better

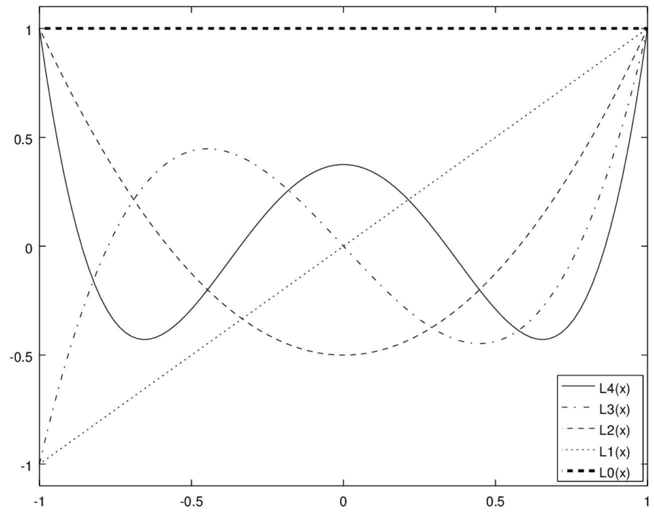


Figure 3.16: The first five Legendre Polynomials $L_0(x)$ to $L_4(x)$.

fit for the Tone 4 syllable [u] than the linear model by capturing the curvature. Higher order polynomials relate to faster varying components of the utterance and may model the data points with increasing fidelity, but such improvement may not be significant. Higher order polynomials are also more likely to lead to over-fitting, achieving poor generalisation. In this study it is desirable to keep the degree of polynomials low so that all coefficients are straightforward to interpret and understand, and potentially "filter" unstructured information out of original dataset.

Coefficients a_2, a_1, and a_0 are, however, hardly independent of each other. The use of orthogonal polynomials ensures the minimal correlations among the coefficients that describe the shape of the contours. In line with S. H. Chen and Wang (1990), Kochanski *et al.* (2005) and Grabe et al. (2007), Legendre polynomials were used, given that their coefficients are uniformly sensitive throughout the utterance. Since Legendre polynomials are defined over the interval $[-1, 1]$, the time axis of f_0 curves was normalised to cover the range of the same interval. The first five Legendre polynomials are shown in Figure 3.16.

The normalised f_0 data were then represented as a best-fit sum of Legendre polynomials with unit variance, specified by a set of c_i coefficients. These coefficients essentially quantify the weight of each Legendre polynomial in approximating an f_0 curve.

$$M(x) = \sum_{i=0}^{2} a_i x^i = \sum_{i=0}^{2} c_i L_i(x) \tag{3.6}$$

In Equation 3.6, $M(x)$ represents the best-fit polynomial model of an f_0 trajectory, $M(x) \approx f_0^*(t^*)$, and x is the normalised time domain ($x = t^*$). The a_i coefficients were obtained the *lm* function in R (R Core Team, 2021). The c_i coefficients can be derived through the a_i coefficients, as in Equations 3.7 to 3.9 for quadratic polynomials. The relevant Legendre polynomials are $L_0(x) = 1$, $L_1(x) = x$, and $L_2(x) = \frac{1}{2}(3x^2 - 1)$ respectively.

$$c_0 = a_0 + \frac{1}{3}a_2 \tag{3.7}$$

$$c_1 = a_1 \tag{3.8}$$

$$c_2 = \frac{2}{3}a_2 \tag{3.9}$$

The procedure is demonstrated with the Legendre polynomials model for the previous example in Figure 3.15 (b) by speaker 101, whose average f_0 was 219.8 Hz. Following Grabe et al. (2007), the first three c_i coefficients have the following interpretations:

1) The first coefficient c_0 is the average f_0 of this monosyllabic utterance after normalisation, predicted by the best-fit model. For this token, $c_0 = 5.898$, the mean f_0 of this utterance is about half an octave (5.9 semitones) higher than the speaker's average f_0.

2) The second coefficient c_1 is half the best-fitting slope of the utterance. So $c_1 = 1.455$ corresponds to about 2.91 semitones ascent in f_0 over time in this utterance. The $L_1(x)$ in Figure 3.17 is the same as the linear fit in Figure 3.15 (b).

3) The third coefficient c_2 indicates the curvature. More precisely, it corresponds to the trend of a parabola shape. In this case, $c_2 = -0.645$ reflects a broad rise or a concave shape in the centre of the utterance.

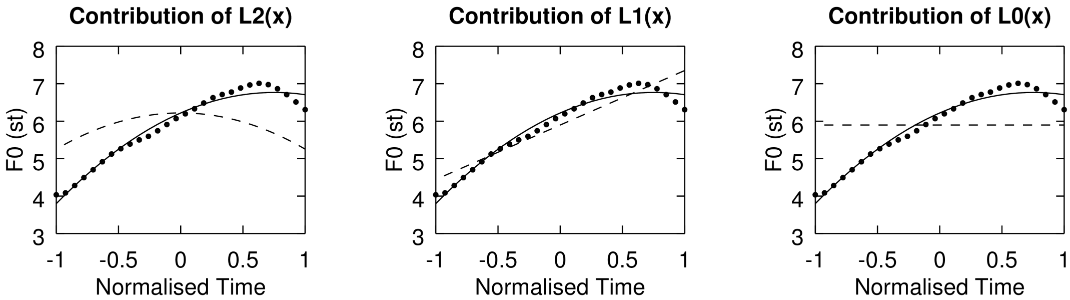


Figure 3.17: Contribution of the first three Legendre polynomials in characterising the pitch contour of the syllable [u] 'fog'. The f_0 estimates in semitone are presented in black dots. The weighted Legendre polynomials are drawn in dashed lines. The best-fit quadratic polynomial is in a solid black line.

How each Legendre polynomial characterises an aspect of the shape of an f_0 contour is illustrated in Figure 3.17, where the weighted Legendre polynomials are drawn in dashed lines. To better illustrate the relation between the $L_i(x)$ and the original contour, normalised f_0 data in black dots and the best-fit quadratic polynomial in a solid black line are also provided as a reference. The weighted $L_2(x)$ and $L_1(x)$ are shifted upwards by c_0 to enable more straightforward visual comparison.

As each c_i coefficient transforms some aspect of the **shape** of a pitch contour into an interpretable quantitative estimate, each f_0 trajectory, originally a series of correlated discrete observations, is now represented by only three coefficients. The c_i coefficients are low-dimensional encoding of the curves and can be used as a surrogate dataset for the raw f_0 contours. The f_0 contours can, therefore, be quantitatively compared across different conditions. These coefficients were then further used as the dependent variables in a series of LME models, through which meaningful categorical prototypes are built.

3.4.4 Linear Mixed Effects regression

Linear Mixed Effects models, with a combination of fixed and random effects, are extremely flexible for modelling continuous outcomes where data are collected with some sort of dependency structure between observations. It has been increasingly used in various fields including phonetic research on f_0. For instance, Evans et al. (2010) incorporated a large amount of potential linguistic and human factors in LME models to test their effects on the vowel median f_0 in Qiang. Aston et al. (2010) built LME models not on a single value but on the whole time series of f_0 data. LME analysis is capable of extracting the contribution of different effects and their interactions and characterising the inherent random variation of the data.

In this chapter, LME models were employed for the analysis of both syllable duration and f_0 contour. The reason for using LME models in this thesis is not ad hoc, but based on their adequacy for repeated-measures design and uneven sample sizes (Baayen et al., 2008). In the field of experimental phonetics, it is widely accepted that subjects and lexical items impose idiosyncratic influence in the realisation of an utterance (Winter & Grawunder, 2012). Clustering our data according to these categories was therefore beneficial and reasonable, and the differences within a category were modelled as random effects. Random effects allow us to assume a different baseline response value for each independent factor.

The formal specification of a linear mixed effects model can be:

$$y = X\beta + Zb + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma^2 I),\ b \sim \mathcal{N}(0, \sigma^2 \Sigma),\ b \perp \epsilon \tag{3.10}$$

where X is the design matrix of the fixed effects, representing factor contrasts and covariates and β is the population coefficients; Z is the design matrix of the random effects and b specifies the adjustment made to calibrate the expected values; $\mathcal{N}$ denotes multivariate Gaussian (normal) distribution; $\perp$ indicates independence of random variables; Σ is the variance-covariance matrix for the random effects, relative to σ^2, the scalar variance of the per-observation noise term ϵ (Baayen et al., 2008, p. 393).

The assumptions for an LME model include the linearity of the relationship between predictors and response, normality of the residuals, homogeneity of residual variance, and absence of collinearlity (Winter, 2013). From the formal specification (3.10), the core assumptions are that the residuals and random effect coefficients are independent and identically distributed (Schielzeth et al., 2020). While I have not conducted a formal survey, my impression is that while an increasing number of studies rely on linear regression methods to generate statistical inference, many of them did not mention assumptions check in their publication (e.g. Aylett and Turk, 2006; Tang et al., 2017; Y. Wu et al., 2020), which is crucial to the validity of the inferences. In the model design and selection phase in this thesis, I have indeed encountered situations when assumptions were violated and I had to explore alternative models or other robust methods to tackle the issues. Such discussion is included in the relevant section describing specific models in chapter 5 (see also Appendix C.2).

The LME models in this thesis were implemented in R (R Core Team, 2021) with the R packages `lme4` (Bates, Mächler, et al., 2015), `lmerTest` (Kuznetsova et al., 2017), and `emmeans` (Lenth, 2021). An example of the model formula expressed in the `lme4` syntax is:

$$y \sim 1 + X + (1 + X|Z) \tag{3.11}$$

where the random effects components are in brackets $(1 + X|Z)$, and the remaining 1 and X denote the intercept and the fixed effect of X. The $1 + X$ preceding the single vertical bar in brackets indicates correlated random intercept and slope. The "1"s can be omitted in the formula. For more model formulas, see Bates, Mächler, et al. (2015). The specific models in the analyses will be presented in the form of `lme4` syntax.

3.5 Duration of Monosyllables

This section reports on the duration of monosyllables in the three varieties, with an emphasis on the differences between the two Mandarin varieties and on the neutral tone syllables. We answer the research questions: *how long is a syllable in isolation in Plastic Mandarin, Standard Mandarin, and Changsha?* and *does tone category and syllable structure have an impact on syllable duration?*

3.5.1 Statistical method

Three groups of linear mixed effects (LME) models were used in the analysis with the aid of R (R Core Team, 2021), *lme4* (Bates, Mächler, et al., 2015), *lmerTest*

(Kuznetsova et al., 2017) and emmeans (Lenth, 2021) to explore sources of variability in the data, given the unequal sample sizes in different categories and repeated samples from the same individuals. The full script and results of the statistical analysis are available in the RMarkdown file *Monosyllables Duration Analysis* in the online supplementary materials.

The first group of LME models were based on all lexical tone syllables in a language variety as in Tables 3.6 and 3.8, to analyse the effect of syllable type and tone category on the duration of a monosyllable. Neutral tone syllables were not involved because they comprised a completely different set of syllables, whose inclusion would make the datasets extremely imbalanced. In this group, *Duration* was the dependent variable, *Syllable* ([i], [ti], [u], and [ma]), *Tone* (T1, T2, T3, and T4 for Mandarin subsets; T1, T2, T3, T4, T5, and T6 for the Changsha subset) and their interaction term were included as fixed effects. Random intercepts for speakers and by-speaker random slopes for both *Syllable* and *Tone* were included. For the Changsha dataset, the [ma] set was dropped given that it did not cover all the tones.

The second group of LME models analysed the relationship between the syllable type and duration of a monosyllable for the neutral tone syllables and the neutral tone subset of a Mandarin variety was used. In the model, *Duration* was the dependent variable and *Syllable* ([ma], [də], [lə], [t̪s̪ə], and [mən]) was the fixed effect. As random effects, I had intercepts for speakers and by-speaker random slopes for the effect of *Syllable*.

For these two groups of models, visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Package *lmerTest* was used to provide the Kenward-Roger approximation of degrees of freedom to compute *p*-values for *F* tests. The *emmeans()* function from package *emmeans* was used to calculate estimated marginal means and conduct Tukey adjusted pairwise post-hoc comparisons.

Thirdly, in order to analyse the relationship between the language variety and duration of a monosyllable, five subsets of Mandarin data based on the five tone categories were created, and for each tone an LME model was used to determine whether the variety of Mandarin significantly affect the duration and to compare the mean duration of the two Mandarin varieties. Each subset contains the duration data consisting of the same set of syllables in both Plastic Mandarin and Standard Mandarin. For example, the Tone 3 Mandarin subset contains duration data of speech tokens [i] 蚁 'ant', [ti] 底 'bottom', [u] 武 'martial art', and [ma] 马 'horse' in both Plastic Mandarin and Standard Mandarin. In the main model of each subset, *Duration* was the dependent variable, *Variety* (Plastic Mandarin, Standard Mandarin) and *Syllable* ([i], [ti], [u], and [ma] for lexical tone subsets; [ma], [də], [lə], [ʈʂə], and [mən] for neutral tone subset) were the fixed-effect factors, and *Speaker* was the random-effect factor. Random intercepts for speakers and by-speaker random slopes for *Syllable* were included. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. Likelihood Ratio Test was used as means to attain *p* values (Winter, 2013) for the fixed effect of *Variety*. In each case, the main model that included the fixed effect *Variety*, model$_{variety}$ illustrated below in R syntax (package *lme4*), was compared against a reduced model without it, model$_{base}$.

model$_{base}$: $Duration \sim Syllable + (1 + Syllable|Speaker)$

model$_{variety}$: $Duration \sim Variety + Syllable + (1 + Syllable|Speaker)$

The effect of *Variety* is significant if the difference between the likelihood of the two models is statistically significant. In most models, the interaction term *Variety***Syllable* was not significant and not included.

3.5.2 Duration and tone

The average duration of monosyllables in the Plastic Mandarin, Standard Mandarin, and Changsha dataset is 378 ms ($\sigma = 79.5$ ms), 413 ms ($\sigma = 114$ ms), and 380 ms ($\sigma = 70.5$ ms) respectively. The variation in duration of monosyllables is larger in Standard Mandarin than that in Changsha and Plastic Mandarin. It should be noted that (only) the Changsha dataset does not include neutral tone syllables.

Figure 3.18 presents the average duration and standard deviation of each tone in Changsha. On average, the shortest tone in Changsha is Tone 5 ($\mu = 350$ ms, $\sigma = 65$

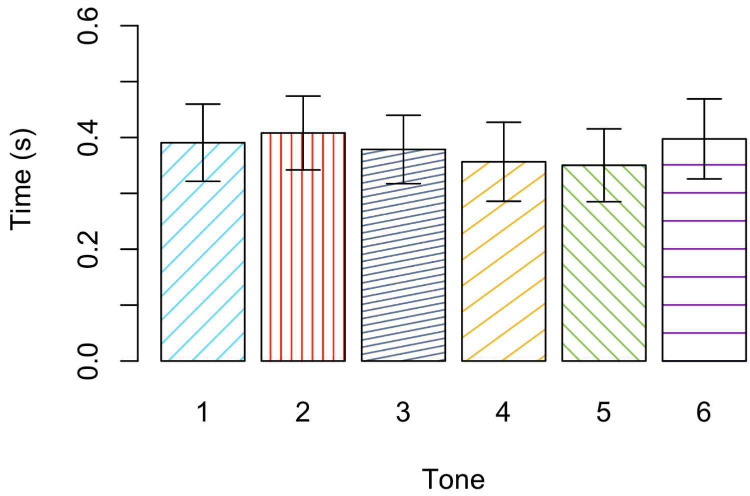


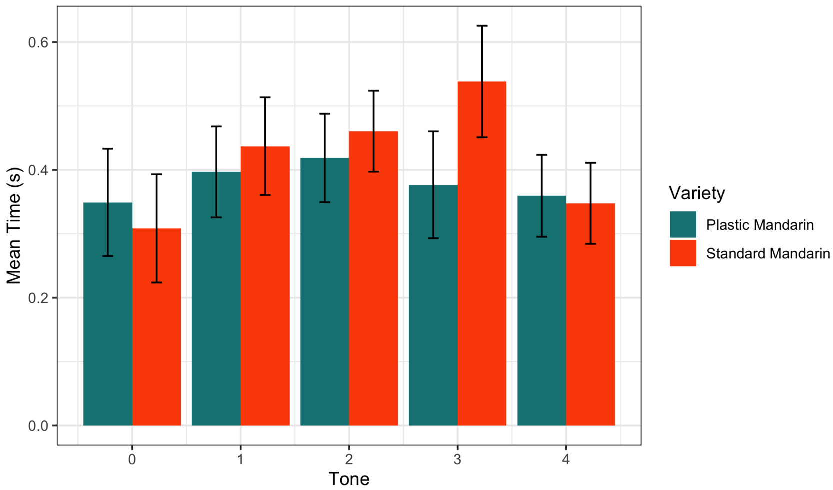
Figure 3.18: Duration of monosyllables by tone categories in the Changsha. The error bars indicate ±1 standard deviation.

ms) and the longest tone is Tone 2 ($\mu = 408$ ms, $\sigma = 66$ ms). The Tukey adjusted pairwise post-hoc comparisons of the least-square mean of each tone estimated from the linear mixed effects models indicate that both Tone 4 and Tone 5 of Changsha are statistically significantly shorter than the other tones and that the difference between Tone 4 and Tone 5 is not statistically significant. Both Tone 4 and Tone 5 of Changsha evolved from the same qù tone category (III) in Middle Chinese.

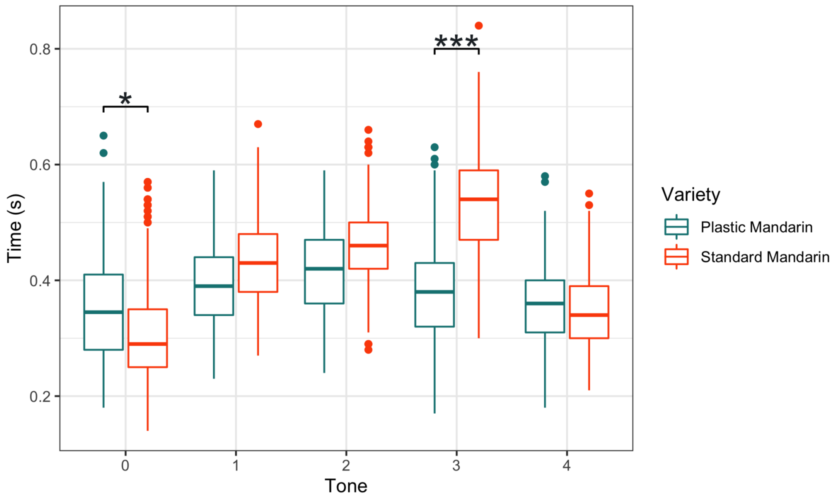
The average duration and standard deviation of each tone in the two Mandarin varieties are displayed in Figure 3.19a. On average, the shortest tone is neutral tone in both Mandarin varieties, about 349 ms ($\sigma = 84$ ms) in Plastic Mandarin and 308 ms ($\sigma = 85$ ms) in Standard Mandarin. The longest tone in Plastic Mandarin is Tone 2 ($\mu = 419$ ms, $\sigma = 69$ ms), while in Standard Mandarin is Tone 3 ($\mu = 538$ ms, $\sigma = 87$ ms). Tone 3 being the longest tone in Standard Mandarin has been consistently reported in previous studies (e.g. Blicher et al., 1990; Y. Xu, 1997; Liu and Samuel, 2004; F. Wu and Kenstowicz, 2015). The long duration has been identified as an exclusive secondary cue in the perception of Tone 3 of Standard Mandarin (Q. J. Fu and Zeng, 2000; Blicher et al., 1990; Liu and Samuel, 2004).

The shortest lexical tone is Tone 4 in both Mandarin varieties, about 359 ms ($\sigma = 64$ ms) in Plastic Mandarin and 348 ms ($\sigma = 63$ ms) in Standard Mandarin. In fact, this study replicates the duration hierarchy Tone 3 > Tone 2 >Tone 1 > Tone 4 in the earlier studies of Standard Mandarin published about 50 years ago (A. T. Ho, 1976; Dreher and Lee, 1968). The diachronically consistent durational findings of Standard Mandarin are mostly in line with cross-linguistically systematic differences in duration among different types of tones - one of the universal phonetic tendencies is that syllables with falling tones tend to be shorter than those with rising tones with other things being equal (Gandour, 1977; Blicher et al., 1990). Tone 4 in Standard Mandarin is a falling tone. There is considerable evidence of a physiological basis for the articulation asymmetry that it takes longer to implement a rising pitch than a falling pitch of the same extent (Ohala, 1978; J. Zhang, 2004). The physiological account, nevertheless, cannot explain Tone 4, the high rising tone (˦˥), being the shortest tone in Plastic Mandarin. Interestingly, the duration of Tone 4 in Plastic Mandarin is similar to Tone 4 in Standard Mandarin (˥˩) despite the tone pattern differs dramatically. The pitch patterns of tones will be introduced in the following sections. The occurrence frequency account seems promising. In Y. Wu et al. (2020)'s corpus study of Standard Mandarin continuous speech, Tone 4 is the most frequently used tone (34.9%), followed by Tone 1 (24.8%) and Tone 2 (23.9%), and Tone 3 is the least frequent tone (16.4%). More specifically, Tone 4 is the most frequent tone in monosyllabic words, as well as almost all syllable positions in disyllabic and trisyllabic words (F. Wu & Kenstowicz, 2015). For monosyllabic words, the occurrence frequency hierarchy is Tone 4 > Tone 1 > Tone 2 > Tone 3 in their corpus. This suggests an inverse relationship between frequency and duration.

Figure 3.19b further shows the distribution of the duration data by tone categories and the comparison of the means of duration between the two Mandarin varieties. The analysis of linear mixed effects models shows that the addition of the factor of *Variety* enabled statistically detectable difference between the two models (i.e. $\text{model}_{variety}$ and model_{base}) in the neutral tone Mandarin data subset ($\chi^2(1) = 9.66, p = 0.0019^{**}$) and Tone 3 Mandarin data subset ($\chi^2(1) = 36.08, p < .001^{***}$).



(a)



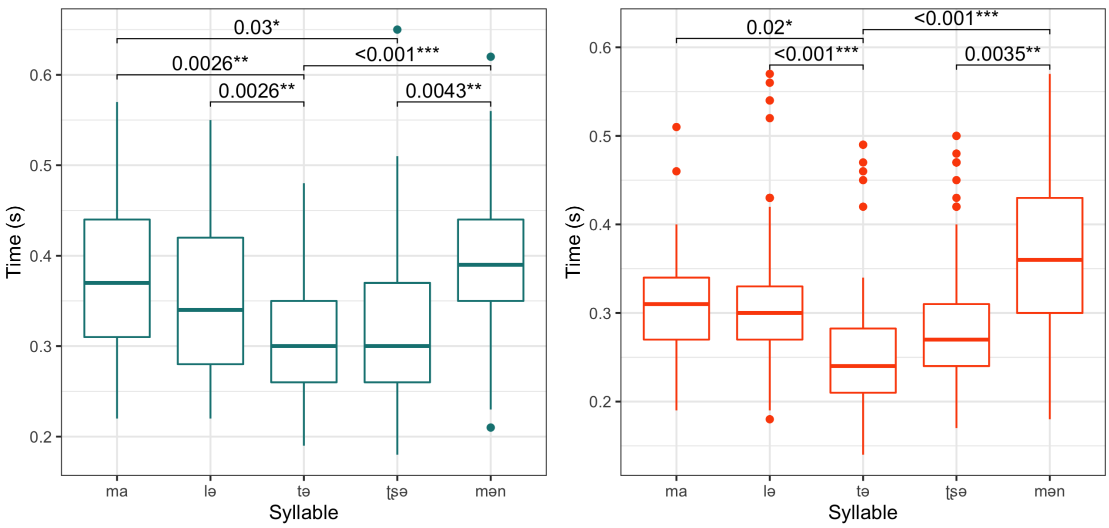
(b)

Figure 3.19: Duration of monosyllables in the two Mandarin varieties. (a) The mean duration by tone categories. The error bars indicate ±1 standard deviation. (b) The distributions of the duration data through their quartiles by tone categories.

In other words, there is statistically significant difference in duration between the two levels of *Variety*, Plastic Mandarin and Standard Mandarin, with regard to neutral tone and Tone 3. For neutral tone, a Standard Mandarin monosyllable is about 58.5 ±18.5 ms (standard errors) shorter than a Plastic Mandarin counterpart. For Tone 3, a Standard Mandarin monosyllable is about 160 ±21.9 ms (standard errors) longer than a Plastic Mandarin counterpart.

3.5.3 Duration and syllable structure

Since the standard deviation for neutral tone syllables is relatively high in both Mandarin varieties, I first examined the segmental pattern of a syllable (represented by *Syllable*) as a source of variability in the citation form of neutral tone syllables. *Syllable* affects the duration of the citation form of a neutral tone syllable in both Plastic Mandarin ($F(4) = 8.38, p < .001^{***}$) and Standard Mandarin ($F(4) = 19.52, p < .001^{***}$), in a similar way. Figure 3.20 presents the duration of neutral tone monosyllables by segmental patterns in the two Mandarin varieties and the p values of the statistically significant pairs from Tukey adjusted pairwise post-hoc comparisons of the levels of *Syllable* effect in the linear mixed effects models.



(a) Plastic Mandarin **(b)** Standard Mandarin

Figure 3.20: Duration of neutral tone monosyllables by segmental patterns in the two Mandarin varieties. The numbers on the top indicate p values from Tukey adjusted pairwise post-hoc comparisons. p values larger than .05 are not shown.

In both Mandarin varieties, [tə] is the shortest syllable, and statistically significantly shorter than [lə], [ma], and [mən]. Meanwhile, [mən] is the longest syllable, and statistically significantly longer than [tə] and [ʈʂə], but not [ma] and [lə]. In terms of syllable structure, [mən] is the only closed syllable with a nasal coda, CVN, while the others are in CV structure. The results, thus, suggest that the presence of a coda is not alone in leading to statistically significant difference in duration.

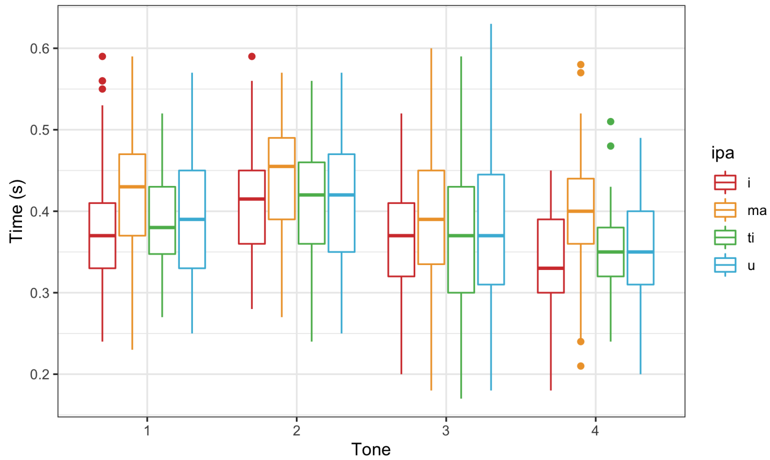
For these neutral tone syllables, a concatenative view of segmental structure tends to capture the general trend of syllable duration — [mən] is the longest given three segments; sonorants such as [m] and [l] are typically longer than pre-vocalic stops; [tə] is the shortest in that [t̪s̪ə] has a fricative part that [tə] does not.

The fact that /tə/ is the shortest may also be related to the effect of lexical frequency since it is the most frequent item (R. Xiao, 2009). A wealth of recent studies report that frequent or predictable words are shorter and exhibit lenited characteristics (e.g. Jurafsky et al., 2000; Aylett and Turk, 2006; Gahl, 2008; Sherr-Ziarko, 2015). Duration measurement may affect the findings too. The shorter tokens /tə/ and [t̪s̪ə] have obstruent onsets, and the silent closure phase during their production is hard to measure when they follow a silent pause. The closure phase might not have been included fully in the token interval.

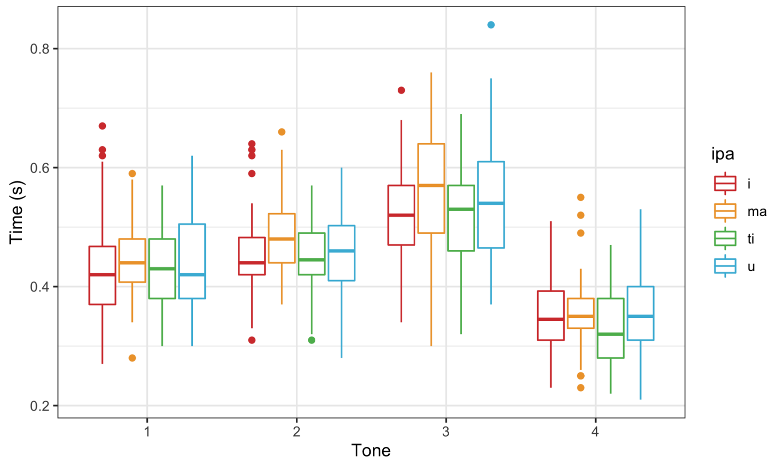
Figure 3.21 presents the duration of lexical tone monosyllables by tone categories and segmental patterns in the two Mandarin varieties. The Tukey adjusted pairwise post-hoc comparisons of the least squares means of different syllables estimated from the linear mixed effects models indicate that [ma] is statistically significantly longer than [i], [u], and [ti] in Plastic Mandarin about 38 ±8 ms ($t(20) = 4.77, p < .001^{***}$), 27 ±5.5 ms ($t(20) = 4.94, p < .001^{***}$), and 32 ±6.8 ms ($t(20) = 4.77, p < .001^{***}$) respectively. In Standard Mandarin, [ma] is statistically significantly longer than only [ti] about 30 ±7.5 ms ($t(13) = 3.98, p = .0075^{**}$). When the incomplete [ma] set was dropped in the Changsha dataset, the fixed effect of *Syllable* was not statistically significant.

3.6 Pitch Contour of Citation Tones

This section presents the results of the analysis of f_0 contours of citation tones in the three varieties, with an emphasis on the scarcely studied Plastic Mandarin. We answer the research questions: *what is the prototypical pitch contour of each tone in Plastic Mandarin and how does it vary?* and *what is the prototypical pitch contour of each tone in Standard Mandarin and Changsha?*



(a) Plastic Mandarin



(b) Standard Mandarin

Figure 3.21: Duration of lexical tone monosyllables by tone categories and segmental patterns in the two Mandarin varieties.

3.6.1 Statistical method

Data in Set B were used to explore the distribution of f_0 among speakers throughout the production of a tone. Data in Set A were prepared for statistical modelling. Following the method introduced in §3.4.2 and §3.4.3, each f_0 contour in Set A was normalised and transformed into three c_i coefficients, which were taken as the dependent variables of a series of LME models. In this way, statistical models of the shape of the normalised f_0 contours as a function of tone categories and other contextual factors were built, to investigate the way speakers manipulate f_0 to convey citation tones in the three varieties.

The main independent fixed-effect predictors were the *Tone* categories (5 levels: T0, T1, T2, T3, and T4 for Mandarins; 6 levels: T1 to T6 for Changsha; T0 represents neutral tone), with *Speaker* being a random-effect factor. Since syllables with different vowels and structures were included in the datasets and repetitions of the same syllable were sequential temporally, these contextual factors were also incorporated in the models as *Syllable* and *Position* respectively to explore and control their influence on the f_0 contours. The analysis was carried out in R (R Core Team, 2021), using the packages *lme4* (Bates, Mächler, et al., 2015) and *lmerTest* (Kuznetsova et al., 2017). With regard to the random-effect structure of LME models, after an initial attempt to fit a 'maximal' model (Barr et al., 2013) which failed to converge, only random intercepts were included to ensure that the models were properly supported by the data and to avoid uninterpretable overparameterised models in the absence of convergence (Bates, Kliegl, et al., 2015). More specifically, the analysis for the Mandarin varieties was conducted in two steps, summarised in the following *lme4* model formulae:

1) $c_i \sim Tone + Position + (1|\ Speaker)$

   The first set of LME models mainly investigate the effect of tone category on the aspects of f_0 curves. *Position* was included as a fixed factor in the model, which contained 6 levels represented by integers from 1 to 6 corresponding to the first token to the last in a list of repeated sequences. It can account for some potential f_0 variance attributed to the serial position of utterances, similar to *list intonation*, or f_0 patterns that can occur on enumerations. It is, however, hard to uniformly control the levels of *Position* across speakers. Participants repeated the syllables in their own pace with longer or shorter pauses. Therefore, the results of each level of the effect of *Position* would hardly be theoretically generalisable. But the omnibus position effect can be controlled to extract the contribution from *Tone*. Given that most neutral tone syllables in the production materials do not share the same segments as other syllables, we do not include a factor of syllable type at this stage.

Alternative models with an interaction term of *Tone* and *Position* added were also examined, but the interaction term was not significant in the models of c_0 and c_2. The interaction term is thus not considered further. More discussion of the interaction term is provided in §3.6.2.2.

2) $c_i \sim Tone + Syllable + Position + Tone : Syllable + (1| \ Speaker)$

The second set of LME models further investigate the effect of the segmental pattern of a syllable represented by *Syllable* on the aspects of f_0 curves. A subset of each Mandarin dataset excluding the neutral tone syllables was used, so that each syllable type covers all four lexical tone categories (T1, T2, T3, and T4). Most interaction terms were not significant and not included. The two-way interaction term $Tone{\times}Syllable$ was found to be significant, and was examined and discussed in §3.6.2.3.

The LME model formulae of Changsha data is summarised as below:

$c_i \sim Tone + Syllable + Position + (1| \ Speaker)$

The effect of *Syllable* and *Position* was included so as to be controlled when the effect of *Tone* was focused. Given that not all syllables were combined with all tone categories in the data, the interaction term $Tone{\times}Syllable$ was not considered.

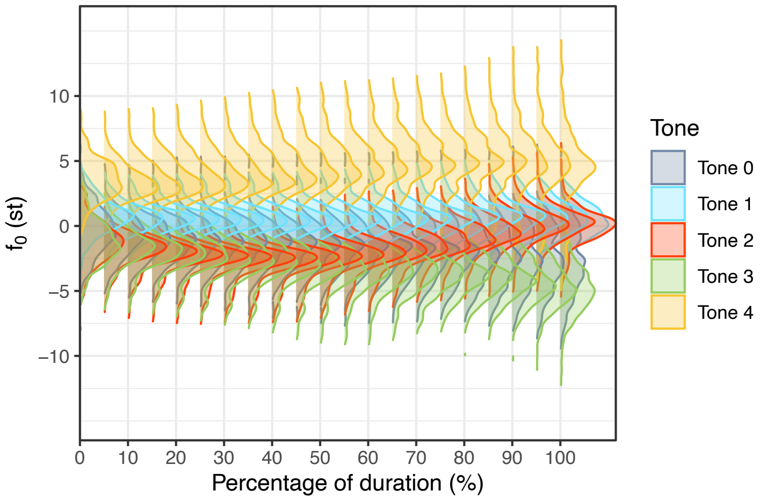
Prior to the LME analysis, assumptions of the models, i.e. linearity, absence of collinearity, homoscedasticity, normality of the residuals, and influential data points, were carefully checked. No violations of the assumptions were found and influential data points were marked for further examination. The significance of the fixed effects was estimated using the ANOVA function in the *lmerTest* package (Kuznetsova et al., 2017) and the Kenward-Roger Approximation from the *pbkrtest* package (Halekoh & Højsgaard, 2014) was used to derive the denominator degrees of freedom. For significant fixed effects, Bonferroni adjusted post-hoc comparisons were performed on the multilevel factors (or interactions), and marginal means in least-square predictions from the LME models for all levels were obtained, using the *emmeans* package (Lenth, 2021). Least squares means generalise the average for unbalanced data or groups in complex design with covariates (Lenth, 2016).

3.6.2 Prototypical citation tones in Plastic Mandarin

We hypothesize that some invariant features or prototype of a f_0 contour mark each tone category to differentiate lexical meanings. In this section, LME models of f_0 curve parameters were employed to extract the effect of tone categories.

Prior to the statistical analysis, Set B data (f_0 contours at 5% intervals) were helpful for us to visualise and compare the extent of variation exhibited in the extracted f_0 data throughout the production of a tone. The probability density distribution of all normalised f_0 data at every 5% of the voicing portion of each syllable grouped by tone categories is displayed in Figure 3.22, where the upper panel offers a comparative angle with superimposed f_0 distributions of all the tones and the lower panel presents the f_0 distribution for each tone. In Figure 3.22, most distributions have a single peak. The ridge or concatenation of the peaks of the density distributions signalling the mode at all times portrays the most likely f_0 contour for each tone. Both Tone 0 and Tone 3 have a downward f_0 trend from the onset of an utterance, but the f_0 distribution is much wider for Tone 0, becoming bimodal towards the end of an utterance. This means that neutral tone syllables do not have a homogeneous citation tone shape. Tone 1 and Tone 2 have similar f_0 distribution at the end of an utterance, but Tone 2 has a distinct overall curve shape. The f_0 distribution of Tone 4 largely occupies a higher region of f_0 (0 - 10 semitones) without much overlap with other tones. The distribution is much wider towards the end of a Tone 4 than the other non-neutral tones.

The variation of f_0 contours in the same tone category revealed in Figure 3.22 may be attributed to many factors, as suggested by previous literature, such as vowel intrinsic pitch (e.g. Shi and Zhang (1986), Whalen and Levitt (1995)), consonant voicing (e.g. House and Fairbanks (1953)), the location of a syllable in speech (e.g. Selting (2007)), word frequency and ambient noise (e.g. Y. Zhao and Jurafsky (2009)), and differences in physiological, social, emotional factors and so on that jointly expressed as speaker idiosyncrasies (Baayen et al., 2008). In this study we incorporated the factors of syllable segmental pattern, location of a syllable in a list, and speaker ID in the LME models.



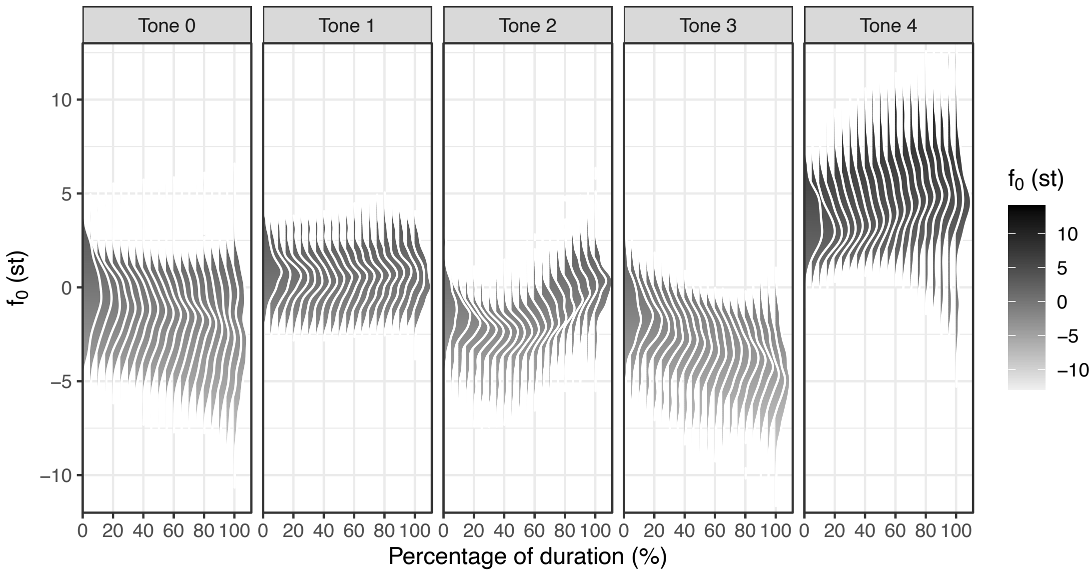


Figure 3.22: The distribution of f_0 in semitone of five tones by all speakers across time.

The f_0 curve coefficients derived from Set A data (10 ms sampled data) were used in the LME models. The fixed effects results of *Tone* and *Position* in the first set of LME models are presented in Table 3.9 and Table 3.10 respectively. For transparency, the dependent variables are referred to as *Average* (c_0), *Slope* (c_1), and *Curvature* (c_2) in the analysis.

3.6.2.1 Tone category

Table 3.9 shows that tone category has highly significant effect on the average, slope, and curvature trend of an f_0 contour. Post hoc tests depict the main effect of *Tone*

		Tone Category				
		Tone 0	Tone 1	Tone 2	Tone 3	Tone 4
Average c_0	Main Effect	$F(4, 2606.7) = 2331.807, p$ **<.001*****				
	Post-hoc Tests	−1.728	0.447	−1.785	−3.268	4.439
Slope c_1	Main Effect	$F(4, 2606.1) = 624.0823, p$ **<.001*****				
	Post-hoc Tests	−0.9832	0.0748	1.2633	−1.5763	0.7428
Curvature c_2	Main Effect	$F(4, 2606.1) = 291.0027, p$ **<.001*****				
	Post-hoc Tests	0.165	0.144	1.254	0.233	−0.652

Table 3.9: Results of the fixed effect *Tone* in LME models of Legendre coefficients. Post-hoc tests show the Least Square Means for each level. Items in bold indicate significant findings.

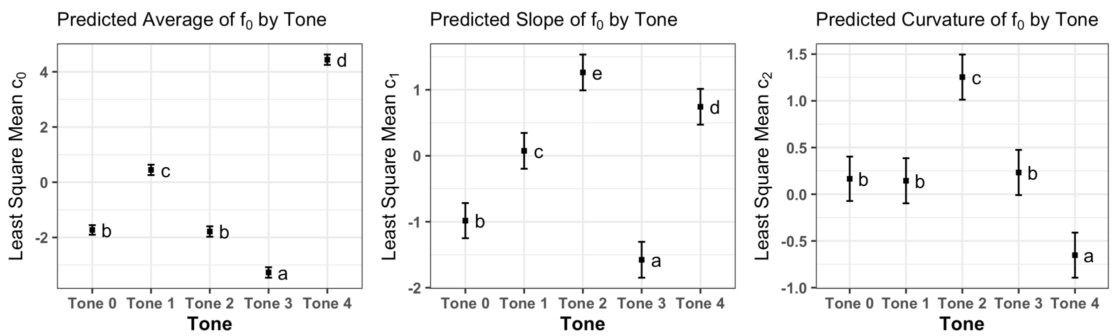


Figure 3.23: Predictions and pairwise comparisons of c_i for all tones in Plastic Mandarin. Results are average over repetitions. Error bars indicate the 95% confidence interval of the least squares mean. Means sharing a letter are not significantly different (Bonferroni-adjusted comparisons).

in Figure 3.23, where the least-square mean of each tone and its 95% confidence interval is shown and the results of pairwise comparison are summarised in letters.

From Figure 3.23, lexical tones T1, T2, T3, and T4 are significantly different from each other in the first two coefficients. In other words, the average of an f_0 contour distinguishes the four tones, so does the slope. Tone 4 is the highest tone, on average about 4.4 semitones higher than the speaker mean, whereas Tone 3 is the lowest tone, on average about 3.3 semitones lower than the speaker mean. Tone 1 tends to be in the centre of the tonal space comprising the five tones, while Tones 2 and 3 are lower. As expected, the slope of Tone 1 is very close to 0, indicating its flat linear trend. Both Tone 2 and Tone 4 are rising tones, and the average f_0 increase over the course of a syllable is about 2.5 and 1.5 semitones respectively.

		Position					
		1	2	3	4	5	6
Average	Main Effect	$F(5, 2607.5) = 10.587, p$ **<.001***					
	Post-hoc Tests	−0.045	−0.254	−0.355	−0.442	−0.502	−0.674
Slope	Main Effect	$F(5, 2606.2) = 1.4318, p = 0.2095$					
	Post-hoc Tests	−0.0091	−0.0597	−0.0819	−0.1142	−0.1207	−0.1888
Curvature	Main Effect	$F(5, 2606.2) = 0.8764, p = 0.496$					
	Post-hoc Tests	0.294	0.219	0.214	0.182	0.212	0.254

Table 3.10: Results of the fixed effect *Position* in LME models of Legendre coefficients. Post-hoc tests show the Least Square Means for each level. Items in bold indicate significant findings.

Neutral tone and Tone 3 have falling slopes and the average f_0 drop is approximately 2 and 3.2 semitones respectively. Tone 2 has the largest positive c_2 least-square mean, indicating a curvy concave upwards shape in the centre of the utterance. In contrast Tone 4, another rising tone, has a concave downwards curve. Other tones are generally much less curvy than Tones 2 and 4.

3.6.2.2 Position

The effect of serial *Position* is only significant on the average of an f_0 contour (see Table 3.10), but not on the slope nor the curvature. We do not expect the f_0 contours of reduplicated utterances to be very different. The average c_0 least-square means of each repetition across tones are presented in Figure 3.24. The overall pitch declines as repetition accumulates but the magnitude of change is small compared to that in Figure 3.23. Such f_0 declination is similar to the gradual downtrend in f_0 over utterances of list reported in literature such as the sequence of step accents in the berry list in Liberman and Pierrehumbert (1984, p. 171). Participants tended to treat the six repetitions of each syllable as a sequentially organised list. Some even produced all repetitions of a syllable in one breath group before moving into the next. The difference of the average f_0 between an initial utterance and the corresponding final utterance is statistically significant. The finality of a exhaustive list may be marked by the falling pitch.

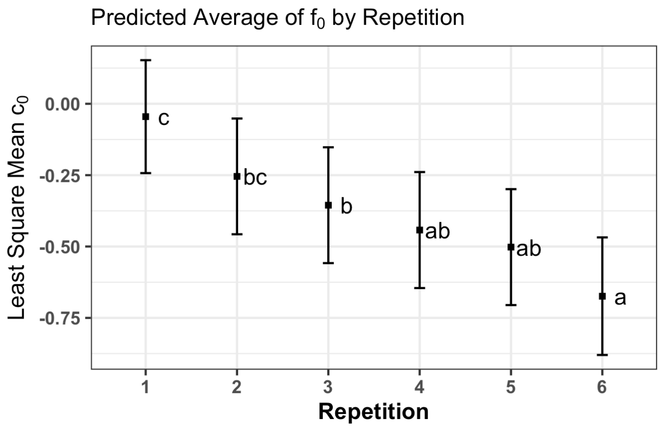


Figure 3.24: Predictions and pairwise comparisons of c_0 for all positions. Results are average over tones. Error bars indicate the 95% confidence interval of the least squares mean. Means sharing a letter are not significantly different (Bonferroni-adjusted comparisons).

The interaction of *Tone* and *Position* was statistically significant ($F(20, 2586.1) = 3.996, p < 0.001^{***}$) in the LME model for *Slope*, that is, *Slope* ∼ *Tone* + *Position* + *Tone* : *Position* + (1| *Speaker*), although the effect of *Position* is not significant ($F(5, 2586.2) = 1.941, p = 0.0844$). The estimated marginal means of c_1 for combinations of predictors *Tone* and *Position* are presented in Figure 3.25, where adjusted predictions were generated by varying the focal variable while holding the non-focal variable constant. From the left graph in Figure 3.25, the means at various positions cluster together around the predicted means for each tone as shown in the middle graph in Figure 3.23, which reveals that the interaction effect is tiny compared to the main effect of *Tone*, despite being statistically significant.

In Figure 3.25, the initial token in general has a larger absolute value of the predicted slope (c_1) than other repetitions of the same syllable, indicating a steeper slope of the f_0 contour and therefore often a larger pitch range. In other words, this suggests that the first utterance may be hyperarticulated. Such hyper-articulation effect tends to be the largest for the highest tone Tone 4, which is consistent with Liberman and Pierrehumbert (1984)'s observation that pitch range manipulations exert larger effect on higher pitches. The position effect is attenuated for tones with less f_0 excursion over time, such as Tone 1 and neutral tone. The absolute

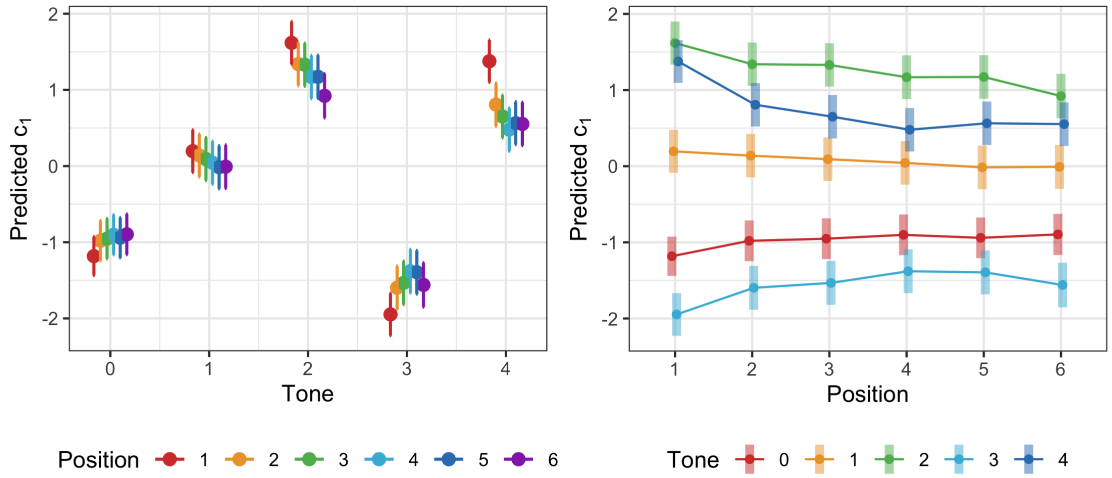


Figure 3.25: The estimated marginal means of c_1 across five tones and six repetitions. Error bars indicate the 95% Confidence interval of the Least Square mean.

value of the predicted c_1 gradually diminishes in a slight degree with each repetition, suggesting a gradually decaying pitch range.

3.6.2.3 Segmental pattern

Excluding the neutral tone syllables, we obtained a subset of the Plastic Mandarin dataset in which each syllable was pronounced in four tones (T1 to T4), as displayed in Table 3.3. This design will enable us to explore the effect of syllable segmental pattern. More specifically, by comparing the set of [iː], [uː], and [maː], the effect of vowel intrinsic pitch may be examined. Comparing the set of [iː] and [tiː] allows us to probe into the effect of voiceless consonant onset.

The set of LME models based on the subset data incorporated one more fixed effect syllable type represented by *Syllable*, and an interaction term *Tone* × *Syllable*. The omnibus effects of the multilevel main factors and interactions are summarised in Table 3.11. All four fixed factors significantly contributed to the variation of the *Average* and *Slope*. The main effects except *Position* significantly affect *Curvature*. Post-hoc tests showed that the effect of *Tone* remained similar to that in the previous models (see results of T1 to T4 in Figure 3.23). Likewise, the *Position* effect was much smaller in scale than *Tone* effect. When the repetition accumulated,

	Average	*Slope*	*Curvature*
Tone	$F(3, 1956.6) = 3593.815, p^{***}$	$F(3, 1956.1) = 725.789, p^{***}$	$F(3, 1956.1) = 381.860, p^{***}$
Syllable	$F(3, 1956.6) = 32.902, p^{***}$	$F(3, 1956.1) = 35.364, p^{***}$	$F(3, 1956.1) = 37.090, p^{***}$
Position	$F(5, 1957.2) = 10.036, p^{***}$	$F(5, 1956.2) = 4.242, p^{***}$	$F(5, 1956.2) = 0.845, p = 0.518$
Tone×Syllable	$F(9, 1956.5) = 5.426, p^{***}$	$F(9, 1956.0) = 6.523, p^{***}$	$F(9, 1956.1) = 2.255, p = 0.0166^{*}$

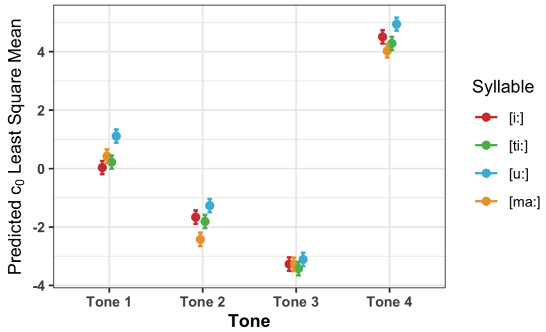
Significance codes: $^{***}p < 0.001$; $^{*}p < 0.05$

Table 3.11: Results of the main effects and interaction effect in LME models of Legendre coefficients.

the average f_0 of a contour gradually dropped (as in Figure 3.24, but with different c_0 means given the averaging results of only four tones instead of five tones), and the slope of a contour levelled off slightly.

Figure 3.26 presents predictions of *Average*, *Slope*, and *Curvature* of a f_0 contour for various syllables in different tones, and the accompanying tables show pairs distinguished significantly by the corresponding factor (in bold). Overall, the effect of *Syllable* is very small in magnitude compared to the *Tone* effect, as the dots in four colors in Figure 3.26 jitter around the estimated marginal means of each tone. Since we do not have the data of [aː] syllables that are often used as interjections in Mandarin, [maː] with a voiced nasal onset can be used as a proxy for low-vowel syllables with care (potential influence of the nasal onset), in comparison to high-vowel syllables [iː] and [uː]. Figure 3.27 shows the prototypical f_0 contour of each syllable type in each tone, consisting of the f_0 means at every 5% of the normalised duration (*Set B* data). Figure 3.27 is conducive to identifying the vowel portion (non-initial portions) in [maː] syllables. For instance, the pitch differences between [uː] and [maː] in Tone 1 and Tone 2 become more evident from about 20% of the duration onwards (a nasal onset [m] was usually about 10 to 20% in duration in the data). Thus the pitch differences in this case mainly lie in the vowel portions.

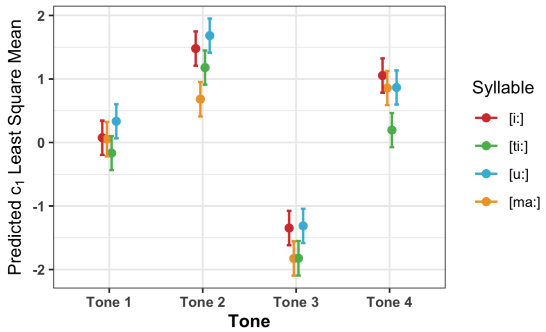
The contours of the syllables [iː] and [uː] do not significantly differ in terms of *Average*, *Slope*, and *Curvature* except that when they are in Tone 1, the average of [uː] is approximately 1 semitone higher than the average of [iː]. [uː] has the highest average f_0 in Tone 1 amongst the four syllables. Similarly in Shi and Zhang (1986)'s study on T1 monosyllables, the middle point f_0 of [u] was 12 Hz



Tukey adjusted pairwise contrasts of c_0

Pair	Tone 1	Tone 2	Tone 3	Tone 4
[i:] - [u:]	**<.0001*** **	0.4619	0.9998	0.298
[i:] - [ma:]	0.5218	**0.0003*** **	1	0.1509
[ma:] - [ti:]	0.9969	**0.0116***	1	0.963
[i:] - [ti:]	0.9985	0.9999	0.9999	0.9896
[ma:] - [u:]	**0.0016** **	**<.0001*** **	0.9996	**<.0001*** **

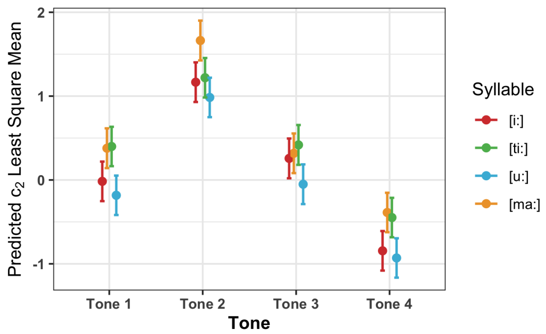
*** p < 0.001 ** p < 0.01 * p < 0.05



Tukey adjusted pairwise contrasts of c_1

Pair	Tone 1	Tone 2	Tone 3	Tone 4
[i:] - [u:]	0.8155	0.9706	1	0.9854
[i:] - [ma:]	1	**<.0001*** **	**0.0206***	0.9782
[ma:] - [ti:]	0.9455	**0.0118***	1	**<.0001*** **
[i:] - [ti:]	0.8783	0.602	**0.0241***	**<.0001*** **
[ma:] - [u:]	0.7099	**<.0001*** **	**0.0078** **	1

*** p < 0.001 ** p < 0.01 * p < 0.05



Tukey adjusted pairwise contrasts of c_2

Pair	Tone 1	Tone 2	Tone 3	Tone 4
[i:] - [u:]	0.9858	0.9678	0.3415	1
[i:] - [ma:]	**0.0461***	**0.0017** **	1	**0.0056** **
[ma:] - [ti:]	1	**0.0104***	1	1
[i:] - [ti:]	**0.0217***	1	0.9916	**0.0392***
[ma:] - [u:]	**0.0001*** **	**<.0001*** **	0.0926	**0.0002*** **

*** p < 0.001 ** p < 0.01 * p < 0.05

Results are averaged over the levels of Position.
Error bars indicate the 95% confidence interval of the Least Square Mean.
Kenward-Roger's menthod for Degrees of Freedom.

Figure 3.26: The estimated marginal means of c_i and pairwise contrasts across five tones and four syllables. The tables contain the Tukey adjusted p values of pairwise comparisons.

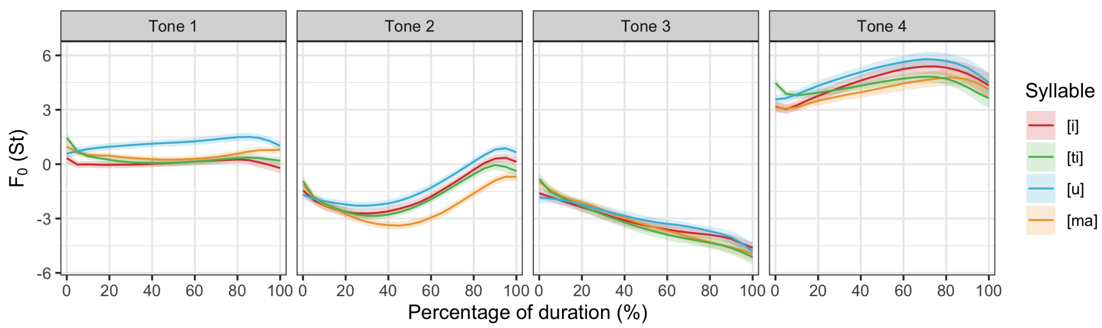


Figure 3.27: Average f_0 contour of different lexical tone syllables by all speakers.

higher than that of [i], and 30 Hz higher than that of [a] on average. The average f_0 of [uː] tends to be significantly higher than [maː] in all tones **but Tone 3**, and their estimated difference was about 0.69, 1.16, and 0.91 semitones in Tone 1, Tone 2, and Tone 4 respectively. In fact, the four types of syllables do not significantly differ in terms of the average f_0 and the curvature when they are in Tone 3, the lowest tone. Such results are consistent with Whalen and Levitt (1995)'s observation on tone languages that the low tones failed to show the intrinsic f_0 effect of vowels. In Whalen and Levitt (1995)'s review of 31 languages, the mean f_0 difference between high and low vowels was about 1.65 semitones. In this study, the results of high back vowel [uː] are consistent with their claim, yet the high front vowel [iː] exhibits a complicated pattern.

The contours of the syllables [iː] and [tiː] do not significantly differ in the average f_0 across lexical tone categories in Plastic Mandarin. [tiː] has a slightly steeper downward slope in Tone 3 but a less steep upward slope in Tone 4 than [iː]. [tiː] is more curvy, concave upwards in Tone 1 and and less curvy, concave downwards in Tone 4 compared to [iː]. Figure 3.27 better illustrates their difference. The contours of [tiː] (in green) feature a high f_0 onset followed by a small portion of steep falling. For the first 5% of the contours, [tiː] is higher than [iː], and indeed higher than all other syllables across four tonal environments. The first 5% data do not significantly change the average f_0 of the whole contour, but influence the slope and curvature trend in the polynomial fitting. The initial high f_0s occurring with the preceding voiceless plosive has been observed in other languages (Hanson, 2009) and such f_0 variation was mostly related to the attribute of voicing in the preceding obstruent (e.g. House and Fairbanks (1953), J. P. Kirby and Ladd (2016)) and aspiration (C. X. Xu & Xu, 2003). In C. X. Xu and Xu (2003)'s study of Mandarin Chinese, the onset f_0 of a tone was higher following unaspirated consonants than following aspirated counterparts.

In sum, the effect of vowel, consonant onset, and utterance sequence on aspects of f_0 contours is statistically significant. The magnitude of these effects are, however, small compared to the main effect of tone category. The robustness of tone category

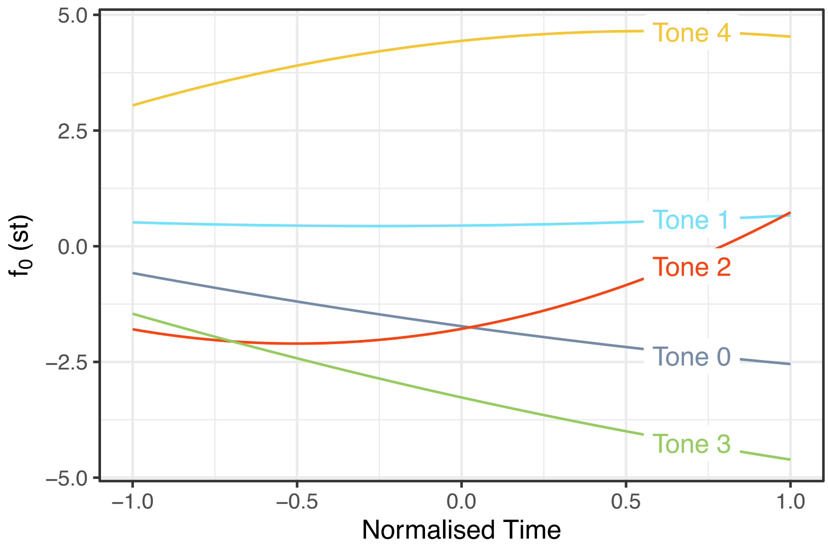


Figure 3.28: Reconstructed f_0 contour of different Plastic Mandarin tones using predicted c_is from LME models. Reconstructed Tone 0 is less reliable, see §3.6.2.4

effect is crucial in accurately conveying the meanings. The analysis extracted the effects of tone category, which determine the prototype of f_0 contour of a tone. Figure 3.28 reconstructs the f_0 contour of each tone using the predicted c_1, c_2, and c_3 from the first set of the LME models. In Plastic Mandarin, Tone 4 is a high rising tone, Tone 1 is mid level tone, Tone 2 and Tone 3 occupy the lower pitch space where Tone 2 rises to a mid pitch and Tone 3 falls to a low pitch. By employing Chao (1930)'s tone letters and numerical notation on a 5-point scale, Tone 1 to Tone 4 could be represented as ˧ 33, ˨˧ 23, ˨˩ 21, and ˦˥ 45 respectively.

Although we have reconstructed an f_0 contour for neutral tone, it may not be the most representative contour given the large variation and bimodal distribution shown in Figure 3.22. The discussion of neutral tones continues in the next section.

3.6.2.4 Neutral tone in isolation

Few studies of citation tones include neutral tones. Neutral tone was analysed as a contextual tone by many previous studies, and some very frequent syllables such as 的 *de* were considered "toneless" (X. S. Shen, 1992). Under the circumstance of citation tone without surrounding syllables, how do speakers pronounce a neutral tone?

Figure 3.29 presents the average f_0 contours of different neutral tone syllables in normalised time and the distribution of f_0 values every 5% of the duration using the *Set B* data. Figure 3.29 reveals that despite the fact that all these syllables are classified as having neutral tone, some including [ma] and [mən] have distinctly different contours while others [lə], [tə], and [ʈʂə] have similar patterns. Comparing Figure 3.29 with Figure 3.27 and Figure 3.28, we find that [ma]'s f_0 contours resemble Tone 1's, [mən]'s resemble Tone 2's, while [lə], [tə], and [ʈʂə]'s resemble Tone 3's.

Neutral tone syllable [mən] has another dictionary pronunciation in Tone 2, which seems to have been evoked during the elicitation session. [ma] is mostly used as an interrogative particle at the end of an utterance. X. S. Shen (1989) observed that yes/no questions, with or without the *ma* particle, ended with a high to mid-high mean f_0 value, different from declaratives which had a low mean f_0 ending point. The sentence-final particle *ma* is often produced with a high tone, which was analysed by S. Peng et al. (2005) as a high boundary tone in the ToBI framework and typically associated with interrogative intonation. One possible explanation for a neutral tone syllable [ma] higher pitched than other neutral tone syllables is that its frequent occurrence in boundary high pitch enabled an association with a high tone, which is Tone 1 in Standard Mandarin. As the tone categories of Standard Mandarin are inherited by Plastic Mandarin, citation form of *ma* in Plastic Mandarin was mainly patterned with Tone 1. A high tone about 5 semitones higher than the speaker mean was also evoked by some participants, as seen from the upper tail of the distribution of f_0 values of [ma] shown in Figure 3.29.

Particles [lə], [tə], and [ʈʂə] are claimed to have lost their tonal specification and been inherently unstressed in Standard Mandarin (S. Peng et al., 2005). Here their citation tones pattern together as a falling tone, start below the speaker mean and end with a low pitch. This might suggest that these neutral tones be not completely toneless but have an underlying low tone shape.

In order to compare such low falling neutral tone contours with contours of Tone 3, a subset data containing these syllables was created. Linear mixed effects models $c_i \sim \mathit{Tone} + \mathit{Position} + (\mathit{Tone} \mid \mathit{Speaker})$ were then built to test if the shape of

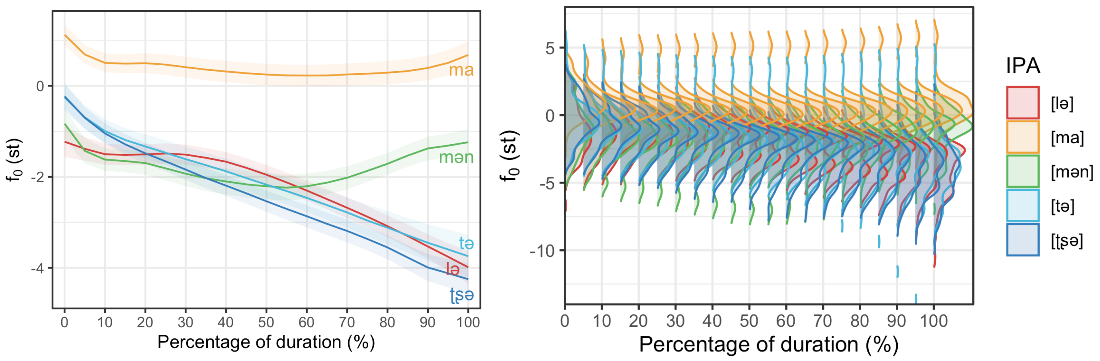


Figure 3.29: Average f_0 contour of different neutral tone syllables by all speakers. Colour bands indicate 95% confident intervals.

	Tone	*Position*
Average (c_0)	F(1, 20.00) = 19.45, p <.001***	F(5, 828.19) = 10.33, p <.001***
Slope (c_1)	F(1, 20.00) = 0.01, p = 0.93	F(5, 828.55) = 10.20, p <.001***
Curvature (c_2)	F(1, 20.00) = 2.11, p = 0.16	F(5, 828.48) = 1.46, p = 0.20

Table 3.12: Results of fixed effects of tone category and repetition in the LME models of Legendre polynomial coefficients for neutral tone syllables and Tone 3 syllables in Plastic Mandarin.

Tone 3 and low falling neutral tone in Plastic Mandarin is significantly different. The dependent variable is the c_i coefficients while the main independent variable is tone category, represented by *Tone* (Tone 3 and Tone 0) in the formulae. Same as the formulae in the previous section, *Position* in a repeated sequence was also included as a fixed effect, and was controlled when we extracted the effect of *Tone*. *Speaker* was included as a random effect, with random intercept and slope design.

Table 3.12 exhibits the analysis of these LME models. Tone 3 and the neutral tone do not significantly differentiate in terms of their slope and curvature. They differ, however, in average f_0. The least-square means of Tone 3 and neutral tone are -3.26 and -2.35 semitones respectively, averaged over the levels of *Position*. In other words, Tone 3 is on average about 0.91 semitones lower than these neutral tones.

3.6.3 Citation tones in Changsha

In the literature, Changsha has six lexical tones. Firth and Rogers (1937) described them as a low mid level tone, a low falling-rising tone, a mid falling tone; a mid high rising tone, a low falling tone, and a low mid rising tone, although they grouped the mid high rising tone and the low falling tone under group 4. In R. Li (1998), the six tones are represented in Chao (1930)'s system as: 1) ˧ 33, 2) ˩˧ 13, 3) ˦˩ 41, 4) ˦˥ 45 or ˥ 55, 5) ˨˩ 21 or ˩ 11, 6) ˨˦ 24. This section presents the analysis of f_0 data of citation tones in Changsha produced by a subset of the same group of participants who contributed to the Plastic Mandarin data. These thirteen teenage Changsha natives speak both Changsha and Plastic Mandarin (as introduced in subsection 3.2.4).

Similarly, Set B data of Changsha was employed to visualise the overall variation in the course of producing a tone. Figure 3.30 shows the density distribution of normalised f_0 data at every 5% of the duration grouped by the six tone categories. For Tone 1, 2, 3, 4, and 6, most distributions are unimodal and the ridge portrays the most likely f_0 contour. Both Tone 1 and Tone 3 have a relatively flat shape, but Tone 3 is generally higher than Tone 1, with the f_0 distribution being wider towards the end. There are three rising tones in Changsha: Tone 2, 4, and 6 have an upward f_0 trend since the onset of an utterance. Compared to Tone 2, Tone 6 in general has a larger rise, especially in the later half of an utterance. f_0 values of Tone 4 are mainly distributed above 0, although there are some deviations given the faint strips below 0. The f_0 distribution of Tone 5 bifurcates and there are two different f_0 patterns: one resembles Tone 4; the other largely occupies a lower f_0 region below 0. The variations in Changsha tones shown in Figure 3.30 reflect that Changsha is not as strictly prescribed as Standard Mandarin and that perhaps Changsha is undergoing a change.

It was found that the bimodality exhibited in their production of Tone 5 is not random but mostly attributed to the alternative pronunciation of the Tone 5 syllable [uː] 雾 'fog'. Figure 3.31 demonstrates the two pronunciations of [uː] 'fog' by two speakers. Speaker 110 has consistent f_0 contours for all Tone 5 syllables, which match the description of a low falling tone in R. Li (1998). Speaker 112, together with all other speakers, adopted a high tone instead, similar to their Tone 4 f_0 contours.

Both pronunciations are acceptable and this is related to a more ubiquitous phenomenon among Chinese varieties that some characters have differentiated

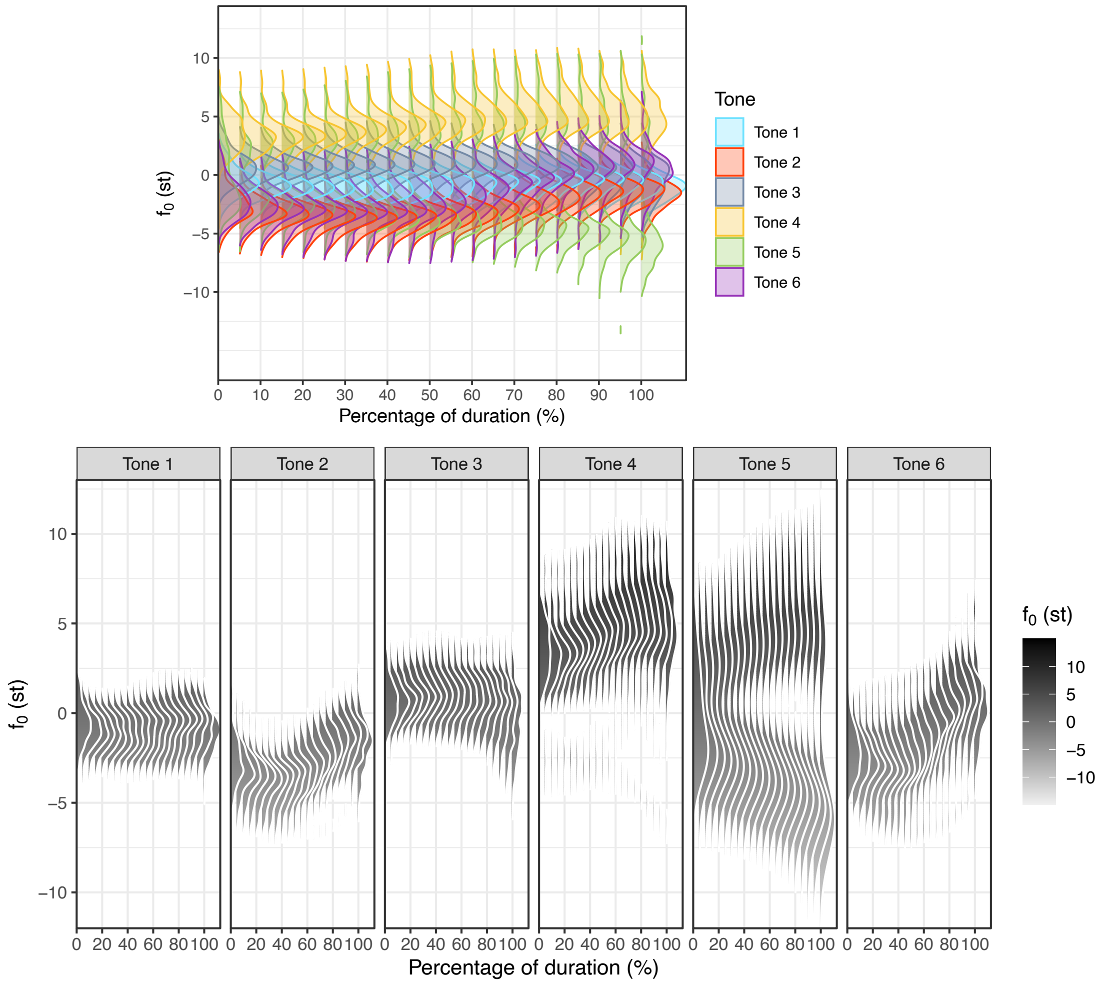


Figure 3.30: The distribution of f_0 in semitone of six tones by all speakers across time. The upper panel offers a comparative angle with superimposed f_0 distributions. The lower panel separates the f_0 distributions for each tone.

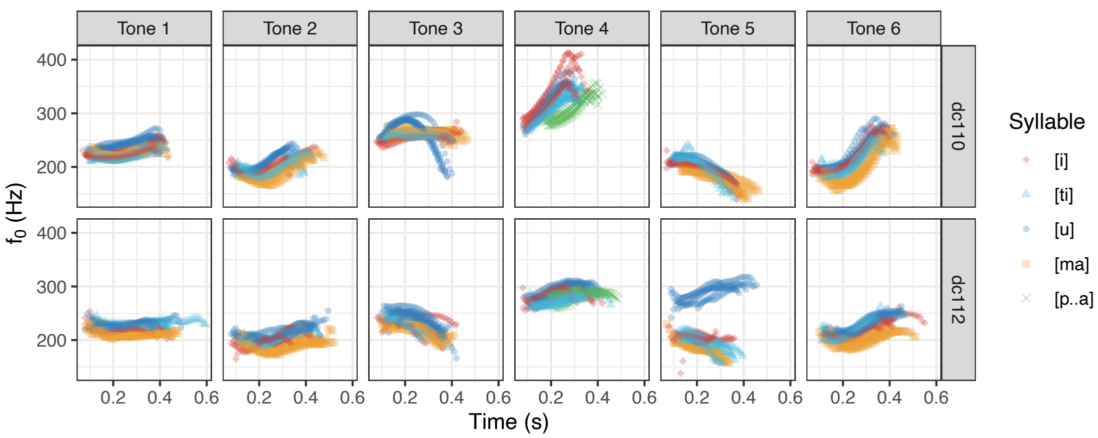


Figure 3.31: All f_0 (Hz) contours in Changsha by speaker 110 (upper) and 112 (bottom).

literary and vernacular readings (D.-a. Ho, 2015). R. Li (1998) stated that for characters with two readings in Changsha, their vernacular reading is a low falling tone ˩ (tone pattern of Tone 5, historical *yáng qù* IIIb tone), while their literary reading is a high rising tone ˥ (tone pattern of Tone 4, historical *yīn qù* IIIa tone). The doublets of readings result from language contact and reflect lexical stratification under the influence of the standard language of education and literary instruction at various periods (Norman, 1979), especially for non-Mandarin regions such as Changsha. The discourse context, hence, matters. The morpheme [uː] 雾 'fog' may have activated disyllabic words such as 雾霾 'smog' (wù mái in Standard Mandarin), which are relatively formal and academic vocabulary and may in turn evoke its literary reading.

J. Peng (2006) claimed that many Changsha Tone 5 syllables with a vernacular form of a low tone having their literary reading of a high tone was due to the influence of neighbouring Southwestern Mandarin, in which the historical qù Tone (III) has been maintained as a single tone category (modern Mandarin Tone 4) while it was split into two registers that became modern Tone 4 and 5 in Changsha. J. Peng (2006) thus predicted the trend of merging Tone 4 and Tone 5 in Changsha, considering the levelling effect that Changsha is becoming more similar to Mandarin. In this case, both morphemes [uː] 雾 'fog' and [uː] 务 'task' in the speech materials are Tone 4 in Mandarin, and participants may have thought them to be Tone 4 in Changsha too. While it is hard to rule out the hypothesis that it is the Mandarin group that has been exerting influence over the tone categories of Changsha spoken by the young, there is not sufficient evidence to ascertain whether the source of influence is the neighbouring Southwestern Mandarin or the national language Standard Mandarin or both.

Figure 3.32 presents a simple mean model of f_0 contours of Changsha tones. Informed by the bimodal distribution in Figure 3.30, Figure 3.32 labels the two variants of Tone 5 (Tone 5A and 5B) and represents their average contours separately. There were two variants of the Tone 4 morpheme [uː] 务 'task' in the elicitation, but they were not consistent among the group and each was spoken by only one

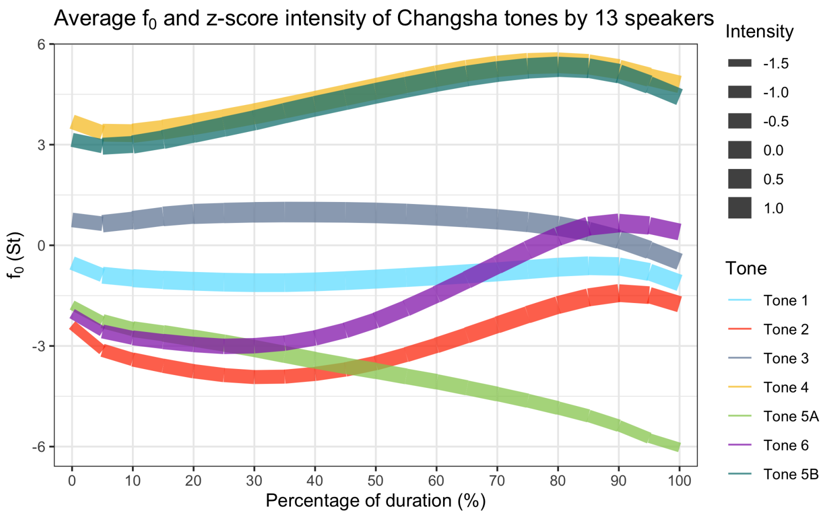


Figure 3.32: Average f_0 and z-score intensity of the Changsha tones by 13 speakers. f_0 in semitone is referenced on the speaker mean. Intensity z-score normalised to the speaker mean is represented by the thickness of lines.

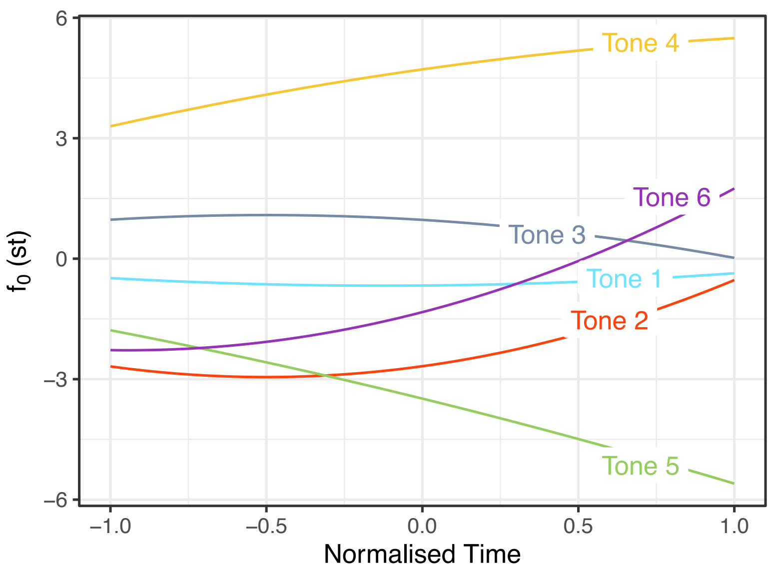


Figure 3.33: Reconstructed f_0 contour of different Changsha tones using predicted c_is from LME models.

participant. These variants (the faint strips below the main ridge in Figure 3.30) might be some sandhi form. However, without consistent data, they were not considered and were removed in the modelling. The average f_0 contours of Tone 4 and Tone 5 variant (Tone 5B) largely overlap as shown in Figure 3.32. This suggests that some morphemes in Tone 5 (Tone 5B) are merging with morphemes in Tone 4 in Changsha.

Figure 3.33 reconstructs the f_0 contours of Changsha tones using the least-square means prediction of c_i coefficients from the LME models, having averaged the effects of other factors. All minority variants including the Tone 5 variant (5B) and Tone 4 variants were not included in this model given they were extremely imbalanced (e.g. Tone 5B consists of mostly [uː] syllables). According to Figure 3.33, Changsha tones produced by these young Changsha natives can be represented as ˧ 33, ˨˧ 23, ˦˧ 43, ˦˥ 45, ˨˩ 21, ˨˦ 24, with a Tone 5 variant being also ˦˥ 45. It worth noting that the 5-point system may not be the most accurate in representing the f_0 height. Although both Tone 3 and Tone 4 start with a “4”, the starting pitch for Tone 4 is much higher than Tone 3. The choice of starting with a “4” for Tone 3 is to differentiate it from Tone 1.

3.6.4 Citation tones in Standard Mandarin

Standard Mandarin tones have been well established in extensive prior research (e.g. Dreher and Lee, 1968; Lee and Zee, 2003). Our data mostly tally with the prescriptive patterns in the textbooks and dictionaries: a high level Tone 1, a mid rising Tone 2, a low dipping Tone 3, and a high falling Tone 4. Allophonic creaky voice (Kuang, 2017) was extensively observed in the data, and the majority (64%) of Tone 3 syllables have a creaky portion in an utterance or entirely creaky.

Figure 3.34 illustrates three variants of the same syllable in Tone 3 by the same speaker, in which two features partial non-modal phonation: in (b) the vowel shows modal-creaky-modal pattern, and the middle of (b) is characterised by low rate of vocal fold vibration and irregular f_0, while in (c) there is a full closure in the middle. These two cases demonstrate variable acoustic realisations of the low tone in Standard Mandarin. Yet the glottalisation is not simply a consequence

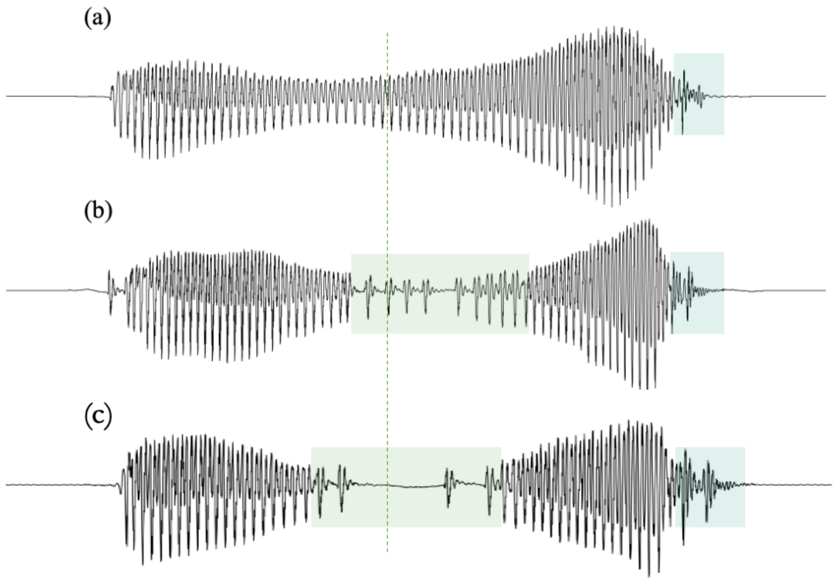


Figure 3.34: Three tokens of [iː] 'ant' in Tone 3 by speaker 201. Creaky portions highlighted. (a) modal voice with regular pulses throughout the syllable; (b) creaky voice in the middle of the syllable. (c) a short 'pause' (glottalisation) in the middle of the utterance.

of low tone. In contrast, in Yucatec Maya vowels with similar phonation pattern ('rearticulated vowels') were analysed as long vowels marked for high tone (Frazier, 2013). Full glottal stop and creaky voice tend to occur as two common instantiations of phonologically specified laryngealisation in many other languages such as the Danish stød and Hawaiian glottal stops (Davidson, 2021). At the end of all three utterances, there are a few highly damped glottal pulses. The f_0 measurements of these speech portions are very low, although Kuang and Liberman (2015) suggests that the perceived pitch is much higher than the actual f_0 in such cases of creaky phonation. In the present analysis, such low f_0 points and their associated time were removed, and a continuous function was fitted to the remaining modal voice f_0 points in the process of orthogonal polynomial modelling. Therefore, tokens with creak in the middle of the vowel had a predicted modal voice trough, different from the actual low f_0 measurement of the creaky pulses.

Figure 3.35 visualises the f_0 contours of Standard Mandarin neutral tones syllables using *geom_smooth* command from the *ggplot2* package. It is a similar case to Plastic Mandarin that speakers did not have a uniform shape for neutral tone

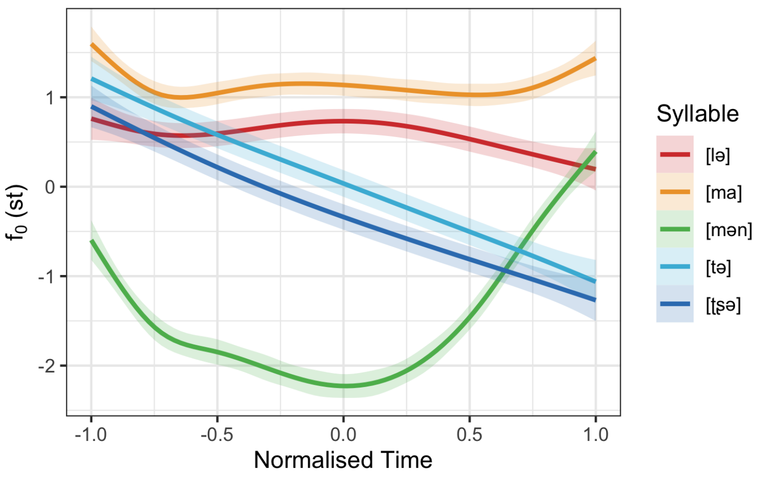


Figure 3.35: Smoothed conditional means using generalised additive models of different neutral tones of Standard Mandarin; formula $y \sim s(x, bs = \text{“}cs\text{”})$, repeated measurements from the same speaker were not handled in these models. Shades indicate the 95 % confidence intervals.

syllables when reading them out in isolation. By comparing Figure 3.35 and Figure 3.29 (a), we can find some common patterns: 1) Syllables [tə] and [tʂə] in Standard Mandarin have a falling pitch, similar to their Plastic Mandarin counterparts. Yet the pitch fall is about 2 semitones in Standard Mandarin and about 4 semitones in Plastic Mandarin. 2) Syllable [ma] has the highest pitch among the other neutral tone syllables and a relatively flat shape. 3) Syllable [mən] has a dipping shape in both Mandarin varieties, although the magnitude of pitch rise is different: about 2 semitones in Standard Mandarin and about 1 semitone in Plastic Mandarin. We also notice the differences between the citation form of these selected neutral tone syllables in the two Mandarin accents. The mean f_0 of the five neutral tone syllables in the course of a tone in Standard Mandarin is in the positive domain, higher than speaker mean, while in Plastic Mandarin it stays in the negative domain. Syllable [lə] patterns with [tə] and [tʂə] in Plastic Mandarin, but not in Standard Mandarin.

Likewise the analysis of Plastic Mandarin tones, Figure 3.36 reconstructs the prototypical f_0 contours of citation tones in Standard Mandarin, using the least-square mean prediction of c_i coefficients from the LME models, having averaged the effects of other fixed factors. Neutral tone was excluded given its heterogeneous

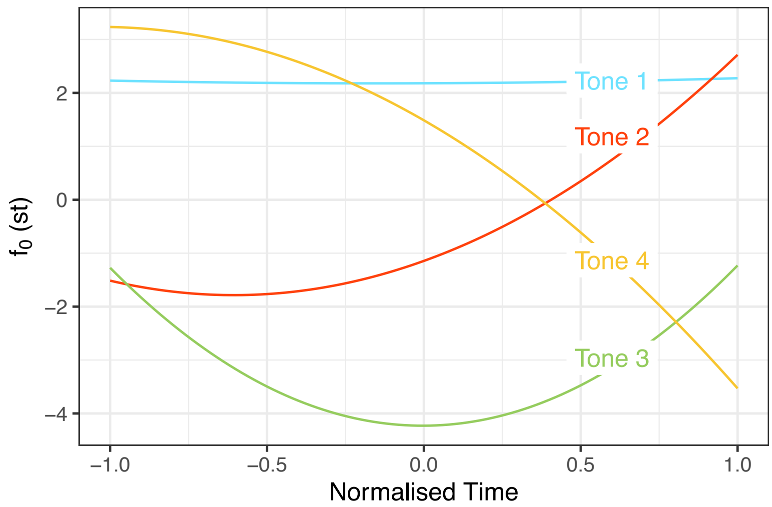


Figure 3.36: Reconstructed f_0 contour of different Standard Mandarin lexical tones using predicted c_is from LME models.

within-category manifestations as shown in 3.35. The reconstructed Tone 3 was based on the Tone 3 syllables with modal voice or modal-creak-modal voice in the dataset.

3.7 The Origin of Plastic Mandarin Tones

Having established the prototypes of Plastic Mandarin tones and presented an analysis of f_0 data of the corresponding speech materials in Standard Mandarin and Changsha, this section further compares the f_0 contours of these varieties, so as to explore the development of the tones in Plastic Mandarin. Given the visual inspection of contours presented in §3.6 and the coefficient clusters in the following §3.7.1, our initial hypothesis is: Plastic Mandarin has borrowed some tone patterns of the corresponding Changsha tones.

In order to attest the origin of Plastic Mandarin tones based on our hypothesis, one has to at least demonstrate that the similarities of the shape of f_0 contours between Plastic Mandarin and Changsha are strongly present in the speech signal. Our parallel datasets of Plastic Mandarin and Changsha from the same group of multilingual participants served as a perfectly-controlled research ground where we can track both group behaviours and individual performances. What's more, we will also show that the Plastic Mandarin tone patterns are significantly different from the patterns of corresponding tones in Standard Mandarin.

3.7.1 Three varieties in the functional parameter space

By comparing the prototypical tone patterns in Plastic Mandarin (Figure 3.28) and Changsha (Figure 3.33), we can see that several tone patterns are similar. In both tone inventories, there is a tone pattern, of which the f_0 starts higher than all other tones and continues to ascend: Tone 4, corresponding to historical tone category *yīn qù* IIIa in both varieties. There is also a level tone that lies in the middle of the tonal space: Tone 1, which corresponds to historical *yīn píng* Ia in both varieties. In addition, there are two tones that start low; one rises to a mid pitch towards the end of an utterance: Tone 2, corresponding to historical *yáng píng* Ib in both varieties, and the other continues to decline reaching the lowest pitch in the tonal space: Tone 3 (*yīn shǎng* IIa) in Plastic Mandarin and Tone 5 (*yáng qù* IIIb) in Changsha. Note that the former three cases have matching historical tone categories.

All f_0 contours were modelled by a sum of weighted Legendre polynomials with three coefficients, which provided a quantitative description of the shape the contours. If we consider the three coefficients as a vector $\vec{c} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}$, contours of the three varieties can be visualised as points in a 3D space. In Figure 3.37 (a), data points of Tone 1 in all 3 varieties spread along the c_0 dimension, Standard Mandarin data points mostly cluster towards the higher end of c_0 to as high as 5 semitones, while Changsha data points mostly cluster towards the lower end of c_0. In both (a) and (b), Plastic Mandarin data points tend to lie somewhere in the middle ground between Changsha and Standard Mandarin, with more overlap with Changsha in (a). In (c) we can see three distinct clusters of data points with little overlap, which indicates that the shapes of Tone 3 differ among the three varieties in the data. Plastic Mandarin and Standard Mandarin points tend to be similar on the c_0 dimension. In (d), Standard Mandarin data points of Tone 4 form a distinct cluster with negative c_1, while Changsha and Plastic Mandarin tend to

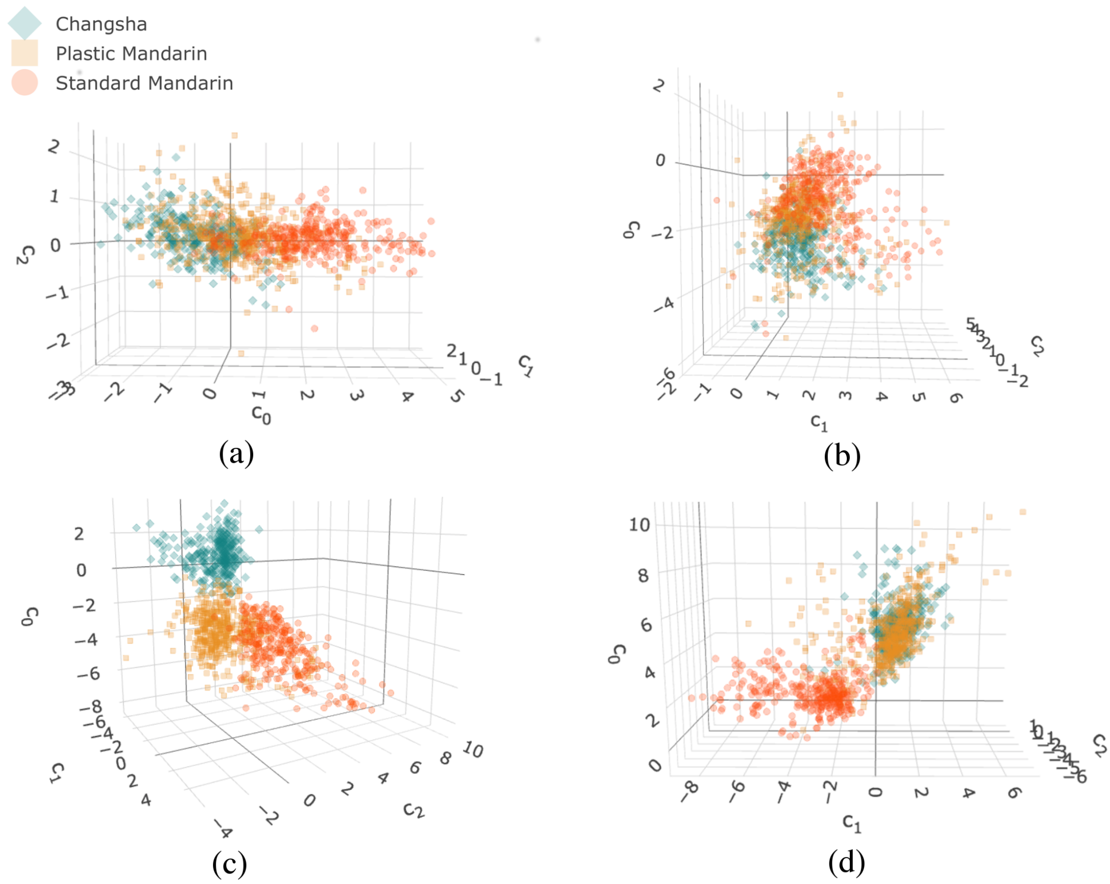


Figure 3.37: Distribution of the coefficient vectors of Legendre polynomial models by tone categories: (a) Tone 1; (b) Tone 2; (c) Tone 3; (d) Tone 4. c_0, c_1, and c_2 indicate the average, slope, and curvature of a contour. Each graph is presented from a viewpoint that shows the wide spread of data points (see 3D interactive visualisation online).

cluster together with positive c_1. Standard Mandarin data points in general also has lower c_0 compared to Plastic Mandarin and Changsha in (d).

Visual comparison of the three varieties shows more similarities of f_0 contours (a greater degree of overlap of the clusters) between Plastic Mandarin and Changsha, especially in terms of Tone 4. Given such observations, the hypothesis is worthy of exploring: is Plastic Mandarin borrowing some tone patterns of the corresponding tones in Changsha? If so, Plastic Mandarin would be a hybrid of Standard Mandarin tone categories and some Changsha tone specification.

This following sub-sections assess the plausibility of this hypothesis by comparing the similar tone patterns in Changsha and Plastic Mandarin, produced by Changsha-Plastic Mandarin bilingual speakers under the same circumstance. These Plastic

Mandarin tones are also compared to corresponding Standard Mandarin tones to check if the differences are statistically significant.

3.7.2 Statistical method

Since thirteen participants contributed to the recordings of both Changsha and Plastic Mandarin monosyllables, their data were selected in this comparative analysis. The analysis concentrates on pairs of Tones 1, 2, and 4 of Plastic Mandarin and Changsha, since Tone 3 in these two varieties differs greatly from visual inspection. Subsets of the combined *Plastic Mandarin–Changsha* dataset on c_i coefficients of the 13 speakers (within-speaker design) were created by tone categories: subsets of all Tone 1, Tone 2, and Tone 4 syllables. Meanwhile the same subsets of another combined *Plastic–Standard Mandarin* dataset of all speakers (cross-speaker design) were also created.

Linear mixed effects models were built to test if the similar tone pairs in Plastic Mandarin and Changsha (i.e. PM Tone 1 and CS Tone 1, PM Tone 2 and CS Tone 2, PM Tone 4 and CS Tone 4) are statistically significantly different. In the same manner, the differences in these tone pairs in Plastic Mandarin and Standard Mandarin (i.e. PM Tone 1 and SM Tone 1, PM Tone 2 and SM Tone 2, PM Tone 4 and SM Tone 4) are also tested. The main independent variable is language variety, represented by *Variety* (2 levels: Plastic Mandarin and Changsha; or Plastic Mandarin and Standard Mandarin) in the following formulae a) and b), while the dependent variable is a c_i coefficient. For each tone pair in each dataset, there were three models, one for each c_i coefficient.

a) $c_i \sim \textit{Variety} + \textit{Syllable} + \textit{Position} + (\textit{Variety} \mid \textit{Speaker})$

b) $c_i \sim \textit{Variety} * \textit{Syllable} + \textit{Position} + (\textit{Variety} \mid \textit{Speaker})$

For comparison of Tone 4's (formula a), *Syllable* ([i], [u], [ti], and [ma]) and *Position*(six levels indicated by 1 to 6) were also included as fixed effects. Interaction between the two fixed effects were not considered given the imbalance in data structure, where only [p^{h}a] in Tone 4 was involved. For comparisons of Tone 1's and Tone 2's (formula b), interaction term *Variety* × *Syllable* was considered given

Speaker	$\bar{f_s}$ (PM)	$\bar{f_s}$ (CS)	adjusted $\bar{f_s}$
102	230.86	231.96	234.11
103	197.00	208.75	207.23
104	191.95	201.03	198.35
105	177.60	174.31	177.77
106	115.80	131.97	127.34
107	106.78	103.98	107.04
110	232.03	236.33	237.53
111	210.88	227.87	221.90
112	230.49	224.97	229.63
113	208.38	216.03	213.53
114	230.34	240.98	239.24
115	215.53	227.40	222.72
117	207.95	221.56	216.92

Table 3.13: The speaker means (Hz) of the thirteen subjects. $\bar{f_s}$ (PM) is based on five Plastic Mandarin tones including neutral tones. $\bar{f_s}$ (CS) is based on six Changsha lexical tones. The adjusted $\bar{f_s}$ is based on 10 distinct lexical tone patterns: four Plastic Mandarin lexical tones plus six Changsha lexical tones, in which variant Tone 5B is grouped with Tone 4.

the balanced data structure. *Speaker* was included as a random effect, with random intercept for all models. For the *Plastic Mandarin–Changsha* dataset, by-speaker random slopes for the main effect of *Language* were also included to capture speaker-specific patterns in the language contrast. Random slopes for the effects of *Syllable* and *Position* were dropped due to the convergence failure of the models.

The c_i coefficients obtained from the previous analyses were based on the speaker means that were defined over a full set of tones in one variety. The speaker means calculated from datasets of different language varieties for the same speaker were, therefore, different. A **variety-independent** speaker mean may offer some insights into the pitch levels of tones in the two varieties. The speaker mean that is **uniform across varieties for the same speaker** $\bar{f_s}$ was then motivated and based on 10 distinct lexical tone patterns: four Plastic Mandarin lexical tones plus six Changsha lexical tones, in which variant Tone 5B is grouped with Tone 4[6]. Table 3.13 presents the speaker means for the 13 Plastic Mandarin-Changsha speakers. With the adjusted $\bar{f}_s$, the c_i coefficients had adjustments accordingly. Only c_0 were affected since we modified the relative height of f_0 contours. The fixed effects of *Syllable* and *Position* were included here to absorb the variance associated with them, and our main discussion focuses on the effects of *Variety* and *Variety* × *Syllable* in the following sections.

3.7.3 Comparison of tone pairs

The following sections compare the f_0 contours of Tone 4's, Tone 1's and Tone 2's in the *Plastic Mandarin–Changsha* dataset and *Plastic–Standard Mandarin* dataset.

3.7.3.1 High rising Tone 4: Signature Xiang tone

The high rising tone patterns include the Tone 4 pattern of Plastic Mandarin and the Tone 4 pattern of Changsha. Before we probe into the similar tone pairs of Tone 4 in Plastic Mandarin and Changsha, we first test if the difference in Tone 4 between Plastic Mandarin and Standard Mandarin is statistically significant. The LME models of all tone 4's in the combined *Plastic–Standard Mandarin* dataset show that *Variety* (2 levels: Standard Mandarin and Plastic Mandarin) is a statistically significant effect in the c_0 ($F(1, 33) = 80.34, p < .001$), c_1 ($F(1, 33) = 77.23, p < .001$), and c_2 ($F(1, 33) = 11.09, p < .01$) of Tone 4. It indicates that Plastic Mandarin Tone 4 is statistically significantly different from Standard Mandarin Tone 4 in terms of all three aspects of a contour, i.e. average f_0, slope, and curvature. This is not surprising in that Tone 4 in Standard Mandarin is a high falling tone.

Figure 3.38 displays the distribution and central tendency of the coefficients of the Tone 4 contours in the *Plastic Mandarin-Changsha* dataset, with jittered

[6] As discussed in §3.6.3, some morphemes in the Tone 5 category in Changsha has a literary pronunciation (5B), which has the same tone pattern as Tone 4, since Tones 4 and 5 are evolved from the *yin* and *yang* registers of the same historical tone, *qù* (III) tone respectively.

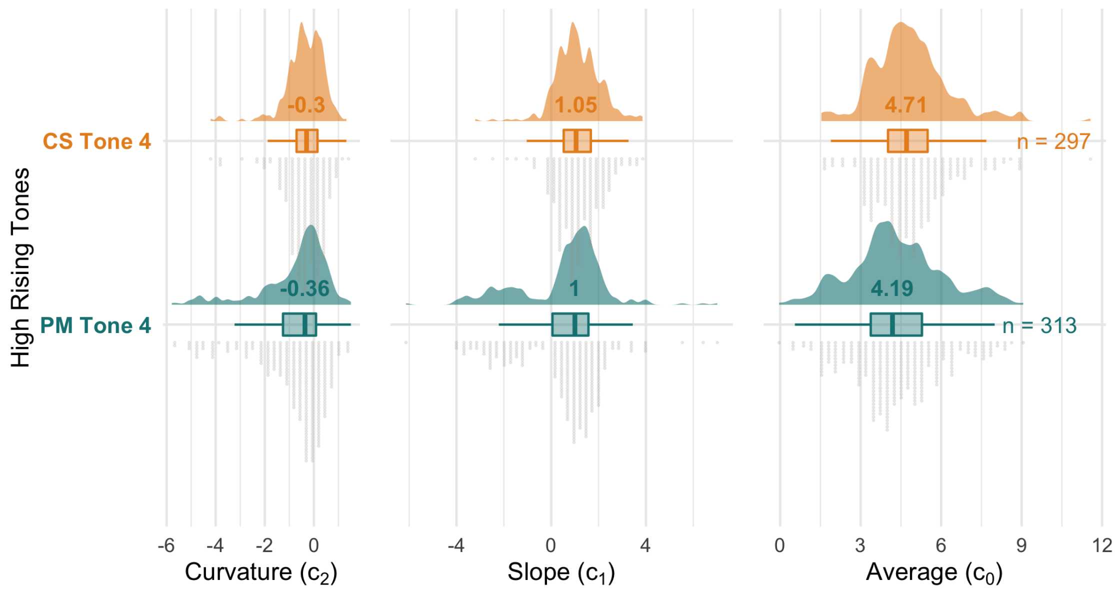


Figure 3.38: Comparison of coefficients of high rising Tone 4's in Plastic Mandarin (PM) and Changsha (CS). The numbers indicate medians. The speaker mean f_0 in the Changsha dataset is different from that in the Plastic Mandarin dataset for a speaker.

	Variety	*Syllable*	*Position*
Average (c_0)	$F(1, 12.68) = 4.64, p = 0.051$	$F(4, 575.36) = 21.36, p^{***}$	$F(5, 575.11) = 10.19, p^{***}$
Slope (c_1)	$F(1, 12.29) = 3.13, p = 0.10$	$F(4, 575.62) = 14.83, p^{***}$	$F(5, 575.18) = 6.61, p^{***}$
Curvature (c_2)	$F(1, 12.36) = 1.49, p = 0.24$	$F(4, 576.54) = 35.78, p^{***}$	$F(5, 575.45) = 4.44, p = 0.52$

Significance codes: $^{***}p < 0.001$

Table 3.14: Results of fixed effects in the LME models of Legendre polynomial coefficients for all Tone 4 syllables in Plastic Mandarin and Changsha.

raw data. For each coefficient, the medians and the interquartile ranges of the distributions are similar among the Tone 4's.

Table 3.14 exhibits the analysis of the LME models of Tone 4's in the *Plastic Mandarin-Changsha* dataset and *Language* is not a significant fixed effect in any aspect of the shape of a Tone 4 contour. This suggests that the shape of a Tone 4 contour is not significantly different between Plastic Mandarin and Changsha and that the average relative height of a Plastic Mandarin Tone 4 in the tonal space of Plastic Mandarin is similar to that of the Changsha counterpart in the tonal space of Changsha. The segmental pattern of a syllable and the position in a repeated sequence, instead, significantly contributed to the variability of the shape of a contour. Figure 3.39 presents the distribution of coefficients grouped by the language and syllables.

Figure 3.40 shows the distribution of the Legendre coefficients based on f_0 values normalised to the variety-independent speaker mean. By comparing Figure 3.38 and Figure 3.40, we find that the median c_0 of Plastic Mandarin Tone 4 is lower in the latter, while the median c_0 of Changsha Tone 4 is slightly higher. Having re-run the LME model on the adjusted c_0, this time the fixed effect of *Language* was statistically significant ($F(1, 12.27) = 7.86, p = 0.0156*$). The least-square mean c_0 of Changsha and Plastic Mandarin Tone 4's estimated in the LME model are 4.94 and 3.74 semitones respectively. Changsha Tone 4 is on average 1.2 semitones higher than Plastic Mandarin Tone 4. It indicates that although the shape of Tone 4 is similar between Plastic Mandarin and Changsha, the average pitch of Tone 4 in Changsha is slightly higher than that in Plastic Mandarin.

In summary, Plastic Mandarin might have borrowed the high rising tone pattern of Changsha Tone 4 and adapted it in the Plastic Mandarin tone system, so that the shape of Tone 4 and its relative pitch level in the tonal space of Plastic Mandarin resembles that of Changsha Tone 4.

Such high rising tones can indeed be a signature of Xiang varieties. A majority of Xiang vernaculars retain a tone category evolved from the high (*yīn*) register of qù tone (IIIa) (H. X. Bao & Chen, 2005) and typically manifested as a high rising tone or mid to high rising tone, such as Changsha, Xiangtan, Xiangxiang, Yueyang, Shaoyang, Shuangfeng, Loudi, and so on. Rather rarely we find such high rising qù tone[7] in other most spoken Chinese varieties. For instance, there is no such high rising tone pattern in the tone inventories of Guangzhou Cantonese, Xiamen Min, Suzhou Wu, and Meixian Hakka (Beijing University, 1995). Tone 4 in Changsha is also reported to be the most frequently occurred among all Changsha tones (Firth & Rogers, 1937) [8]. The high rising Tone 4 is hence one of the most salient tones of Changsha, given its high frequency of occurrence and that its status as a common

[7] Nanchang Gan and Xi'an Mandarin have the high rising qù tone.

[8] In Firth and Rogers (1937)'s work, Tone 4 and Tone 5 were considered as one tone group with two pitch patterns.

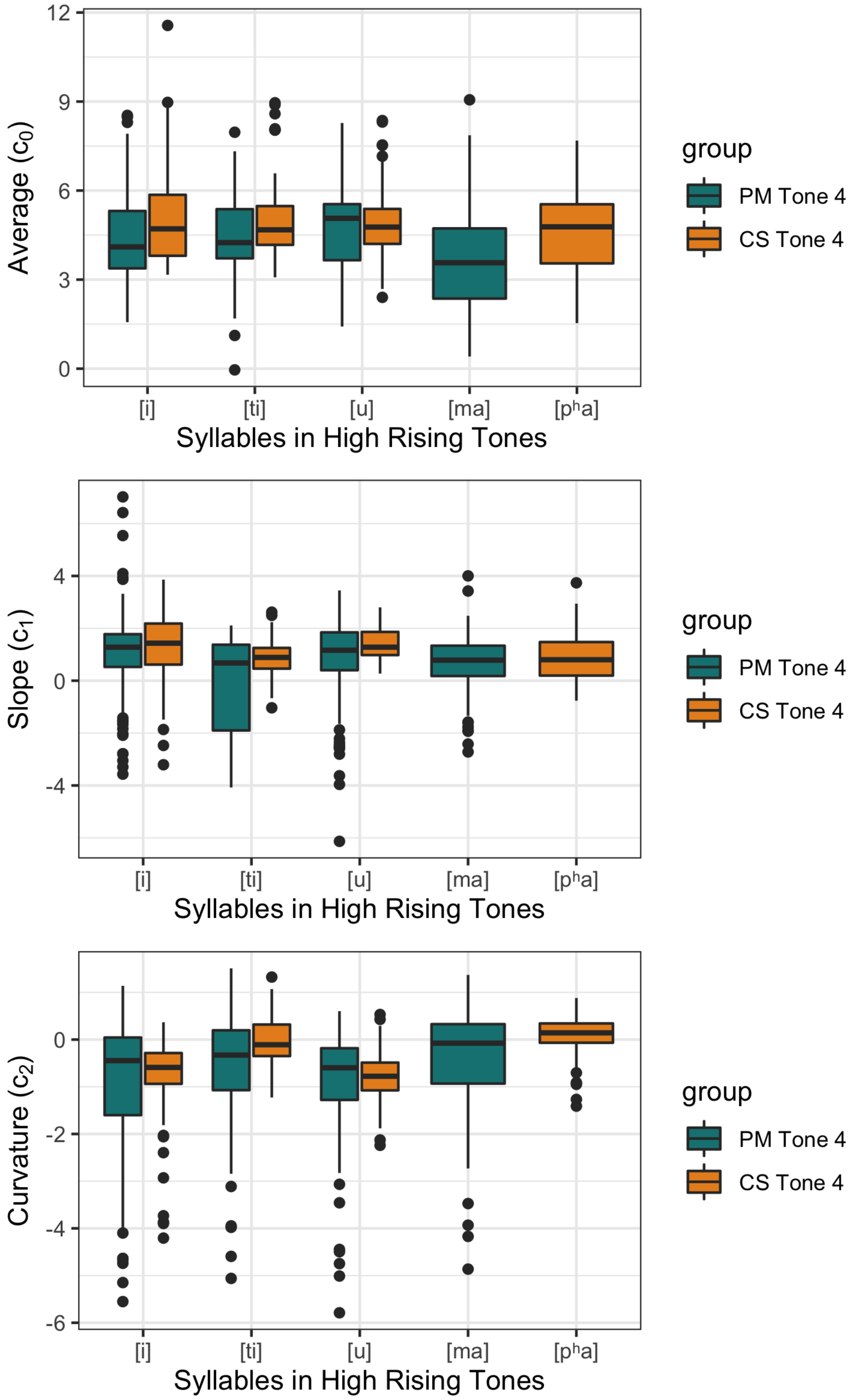


Figure 3.39: Comparison of coefficients of high rising tones by syllable types in Plastic Mandarin (PM) and Changsha (CS).

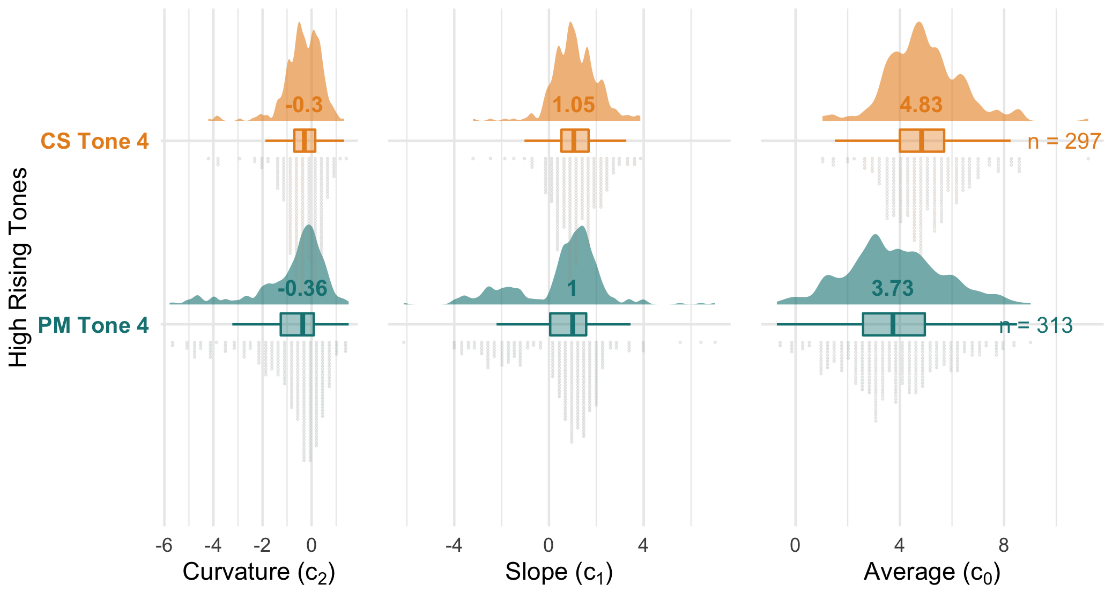


Figure 3.40: Comparison of coefficients of high rising tones in Plastic Mandarin and Changsha. The numbers indicate medians. The speaker mean f_0 in the Changsha dataset is the same as that in the Plastic Mandarin dataset for a speaker.

and consistent feature among many Xiang varieties. The adaptation of the high rising Tone 4 in Plastic Mandarin brings a spoonful of local flavour into the accent.

3.7.3.2 Tone 1 and Tone 2

In line with the analysis for Tone 4, Figures 3.41 and 3.43 show the distribution and central tendency of the Legendre polynomial coefficients obtained in the analysis of f_0 contours of Tone 1's and Tone 2's respectively, with f_0 normalised to the variety-independent mean. Tables 3.15 and 3.16 provide the analysis of all fixed effects in the LME models of these coefficients for Tone 1's and Tone 2's in the *Plastic Mandarin-Changsha* dataset respectively.

For Tone 1, the slope and curvature are not significantly different between Plastic Mandarin and Changsha, nor between Plastic Mandarin and Standard Mandarin. This is expected since Tone 1 in the three varieties is roughly level tones. The interaction term *Variety* × *Syllable* is significant for the c_0 model in the *Plastic Mandarin-Changsha* dataset, so that the difference in c_0 is syllable-dependent. The least-square means predicted by the model are shown in Figure 3.42. The differences of the c_0 means between Plastic Mandarin and Changsha

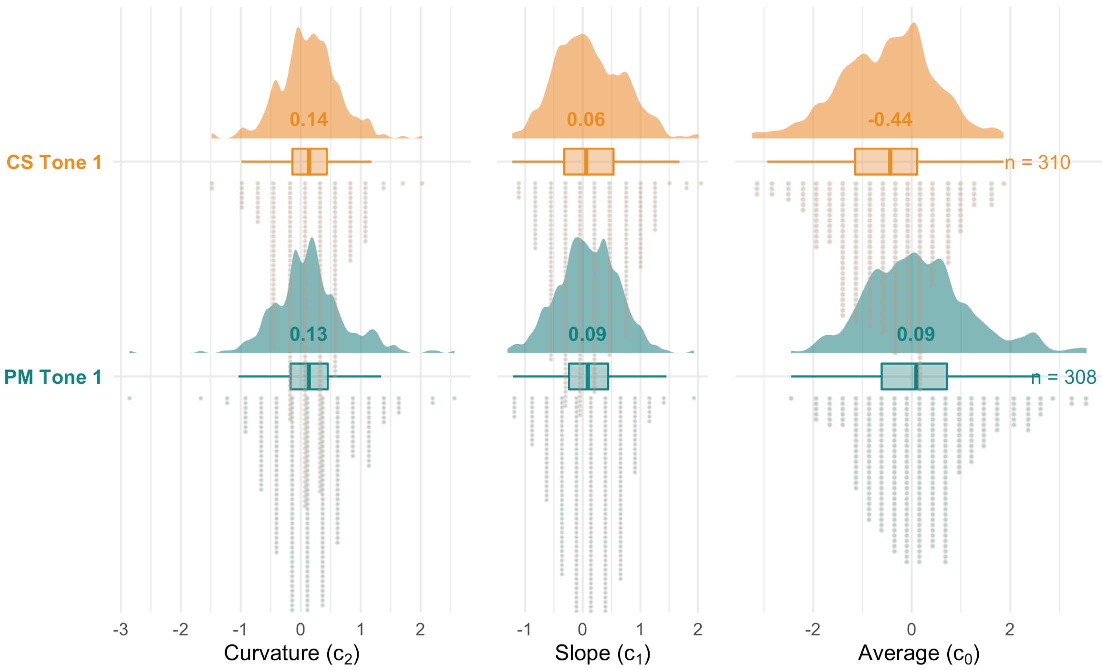


Figure 3.41: Comparison of coefficients of Tone 1 in Plastic Mandarin and Changsha. The numbers indicate medians. The speaker mean f_0 in the Changsha dataset is the same as that in the Plastic Mandarin dataset for a speaker.

	Average (c_0)	*Slope* (c_1)	*Curvature* (c_2)
Variety	$F(1, 12.00) = 6.12, p = 0.03$*	$F(1, 12.00) = 0.12, p = 0.74$	$F(1, 12.00) = 0.05, p = 0.83$
Syllable	$F(3, 581.02) = 23.12, p$***	$F(3, 581.04) = 46.31, p$***	$F(3, 581.15) = 65.42, p$***
Position	$F(5, 581.10) = 5.66, p$***	$F(5, 581.13) = 5.43, p$***	$F(5, 581.64) = 1.28, p = 0.27$
Variety × *Syllable*	$F(3, 581.02) = 10.98, p$***	$F(3, 581.03) = 2.07, p = 0.10$	$F(3, 581.11) = 1.46, p = 0.22$

Significance codes: $^{***}p < 0.001$; $^{*}p < 0.05$

Table 3.15: Results of fixed effects in the LME models of adjusted Legendre polynomial coefficients for all Tone 1 syllables in Plastic Mandarin and Changsha, with Kenward-Roger's method.

for Tone 1 syllables [ma] and [u] are statistically significant, but not for syllables [i] and [ti]. For the Plastic-Standard comparison, the difference of the c_0 means between Plastic Mandarin Tone 1 and Standard Mandarin Tone 1 is statistically significant ($F(1, 33) = 46.27, p < .001$) and the average c_0 of Plastic Mandarin Tone 1 is 1.74 semitones lower than that of Standard Mandarin Tone 1.

For Tone 2, the interaction effects are significant in the c_0, c_1, c_2 models. The statistical difference between the least-square means of c_0 of the two varieties

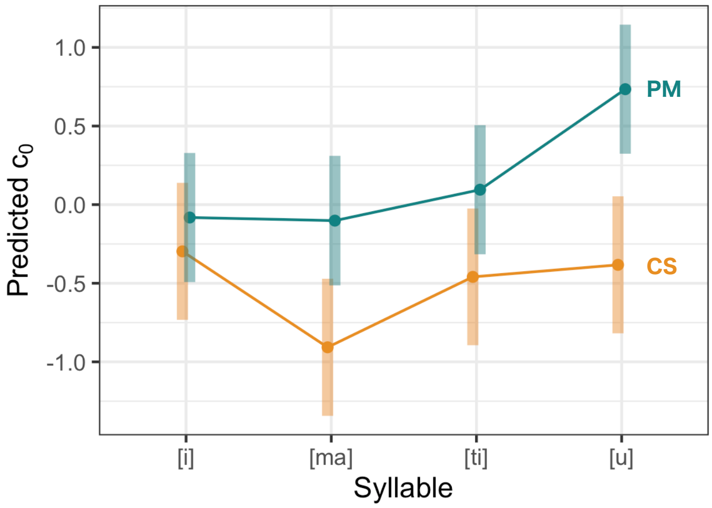


Figure 3.42: Comparison of predicted average f_0 of Tone 1 in Plastic Mandarin and Changsha.

is (marginally) significant. Figure 3.44 illustrates the interaction effect between language variety and segmental pattern of a syllable. In fact, the difference of the c_0 means between Plastic Mandarin and Changsha Tone 2 is only statistically significant for syllables [i] (i.e. between CS [i] T2 and PM [i] T2), but not for syllables [ma], [ti], or [u]. The slope is not significantly different between Plastic Mandarin and Changsha. For the Plastic-Standard comparison, the difference of the c_0, c_1, c_2 means between Plastic Mandarin Tone 2 and Standard Mandarin Tone 2 is statistically significant. The interaction term in these models is not statistically significant. The average c_0 of Plastic Mandarin Tone 2 is 0.65 semitones lower than that of Standard Mandarin Tone 2 ($F(1, 33) = 4.72, p = 0.037$), the average c_1 of Plastic Mandarin Tone 2 is 0.86 lower (less steep rise) than that of Standard Mandarin Tone 2 ($F(1, 33) = 9.65, p = 0.004$), and the average c_2 of Plastic Mandarin Tone 2 is 0.48 lower (less curvy) than that of Standard Mandarin Tone 2 ($F(1, 33) = 5.42, p = 0.03$)

Despite the resemblance to corresponding Changsha tones, Plastic Mandarin tones are not a simple copy of them. There are some statistically significant differences in the f_0 contours of similar tone patterns, especially in the average pitch. The analysis reveals that the c_0 means of Tone 4 are not statistically different between Plastic Mandarin and Changsha when f_0 is normalised to the

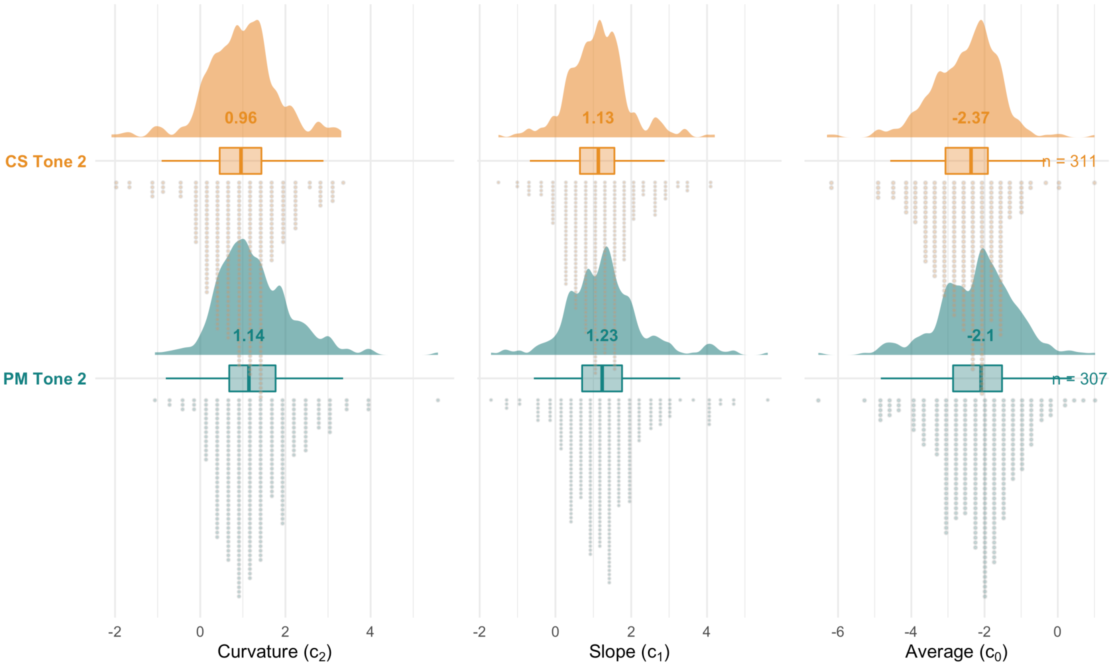


Figure 3.43: Comparison of coefficients of Tone 2 in Plastic Mandarin and Changsha. The numbers indicate medians. The speaker mean f_0 in the Changsha dataset is the same as that in the Plastic Mandarin dataset for a speaker.

	Average (c_0)	*Slope* (c_1)	*Curvature* (c_2)
Variety	$F(1, 12.00) = 4.96, p = 0.046^{*}$	$F(1, 12.00) = 1.41, p = 0.26$	$F(1, 12.00) = 6.33, p = 0.03^{*}$
Syllable	$F(3, 581.04) = 64.75, p^{***}$	$F(3, 581.05) = 71.31, p^{***}$	$F(3, 581.05) = 54.14, p^{***}$
Position	$F(5, 581.13) = 9.50, p^{***}$	$F(5, 581.22) = 10.39, p^{***}$	$F(5, 581.26) = 10.58, p^{***}$
Variety × *Syllable*	$F(3, 581.04) = 5.75, p^{***}$	$F(3, 581.05) = 3.43, p = 0.02^{*}$	$F(3, 581.05) = 4.17, p = 0.006^{**}$

Significance codes: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3.16: Results of fixed effects in the LME models of adjusted Legendre polynomial coefficients for all Tone 2 syllables in Plastic Mandarin and Changsha, with Kenward-Roger's method.

speaker mean defined in the tonal space of each variety (similar relative pitch height in a tone system), despite the significant differences in the c_0 means when f_0 is normalised to the speaker mean that is uniform across language varieties (different absolute pitch height). This suggests that Tone 4 is adjusted in pitch level to fit the whole tone system of a different variety. If Plastic Mandarin has borrowed the tone patterns of Changsha, another possible interpretation of the average pitch differences between Plastic Mandarin and Changsha tones is the

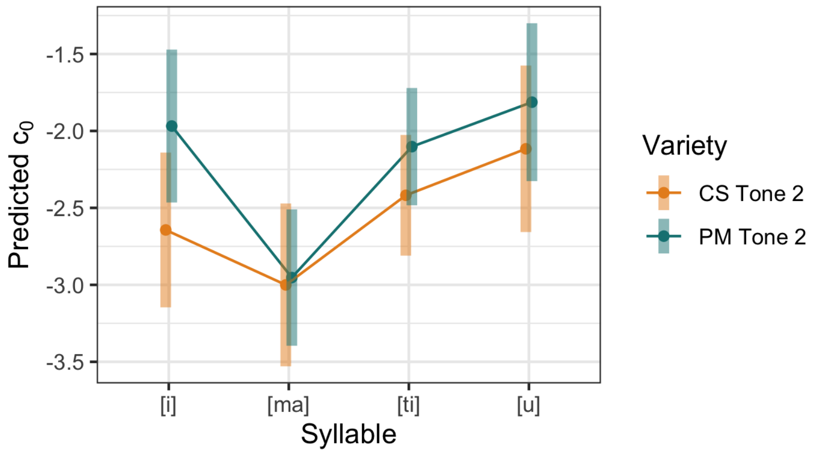


Figure 3.44: Comparison of predicted average f_0 of Tone 2 in Plastic Mandarin and Changsha.

converging influence from Standard Mandarin. Tone 1 in Standard Mandarin is a high level tone, and it might have influenced Plastic Mandarin Tone 1, which is slightly higher than Changsha Tone 1, a mid level tone.

3.7.4 Variations in tones of individuals

The previous section examines and compares the tone shapes of Tones 1, 2, and 4 on a group level, this section probes into the tone patterns of individuals. Figure 3.45 shows the Plastic Mandarin and Changsha tones produced by speaker 110. His Tone 4's of Plastic Mandarin greatly overlap with his Tone 4's of Changsha, while his Plastic Mandarin Tone 1's and Tone 2's have similar shapes with his Changsha counterparts but are slightly higher.

Figure 3.46 presents the f_0 contours of Tones 1, 2, 3, and 4 in Changsha and Plastic Mandarin by all thirteen speakers. Some speakers such as 103, 111, 113, and 117 have highly overlapping Tone 1s and Tone 2's of Plastic Mandarin and Changsha, while some others such as 102, 107, 110, and 112 have the Plastic Mandarin Tone 1's and Tone 2's higher than the Changsha counterparts. The Plastic Mandarin Tone 3's and Tone 4's have larger within-tone category variations than Tone 1's and Tone 2's: speakers 105, 112, and 113 have a dipping (falling-rising) shape for Tone 3's, while the majority have a falling shape; speakers 103, 104, 105, 106 have large

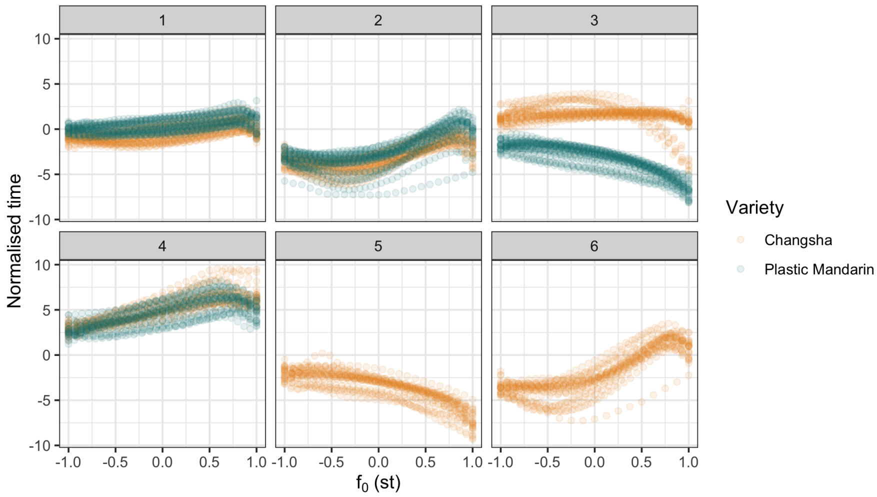


Figure 3.45: Normalised f_0 data of Changsha (yellow) and Plastic Mandarin (teal) by tone categories from speaker 110. The variety-independent speaker mean was used as the reference frequency in the semitone conversion.

falling tails for some of their Tone 4's while the majority maintain a rising shape. The dipping shape of Tone 3 and large falling shape of Tone 4 are characteristic of Standard Mandarin. These variants might be a manifestation of the overshadowing influence of Standard Mandarin or an indicator of ongoing tonal change for Plastic Mandarin. Speaker 105's Plastic Mandarin accent, for instance, tends to orient more towards Standard Mandarin than others. Appendix C.1 presents the influential data analysis of the LME models of c_i coefficients of Plastic Mandarin, and speaker 105 stands out, especially in terms of the curvature and slope of his citation tones. This in turn suggests that methods that are commonly used to estimate the influence of a data point can be used to profile individual accent.

3.7.5 Potential tonal change mechanism

Figure 3.47 illustrates the potential mechanism of forming Plastic Mandarin lexical tones. Plastic Mandarin inherits the tone categories of Standard Mandarin; the tone specification, however, largely resembles the corresponding Changsha tone (except for Tone 3). Plastic Mandarin hence acquired elements from both varieties. The reason why Tone 3 behaves differently is still unknown. One possibility is that Tone 3 adopts the the low falling form of Tone 3 in Standard Mandarin instead of the

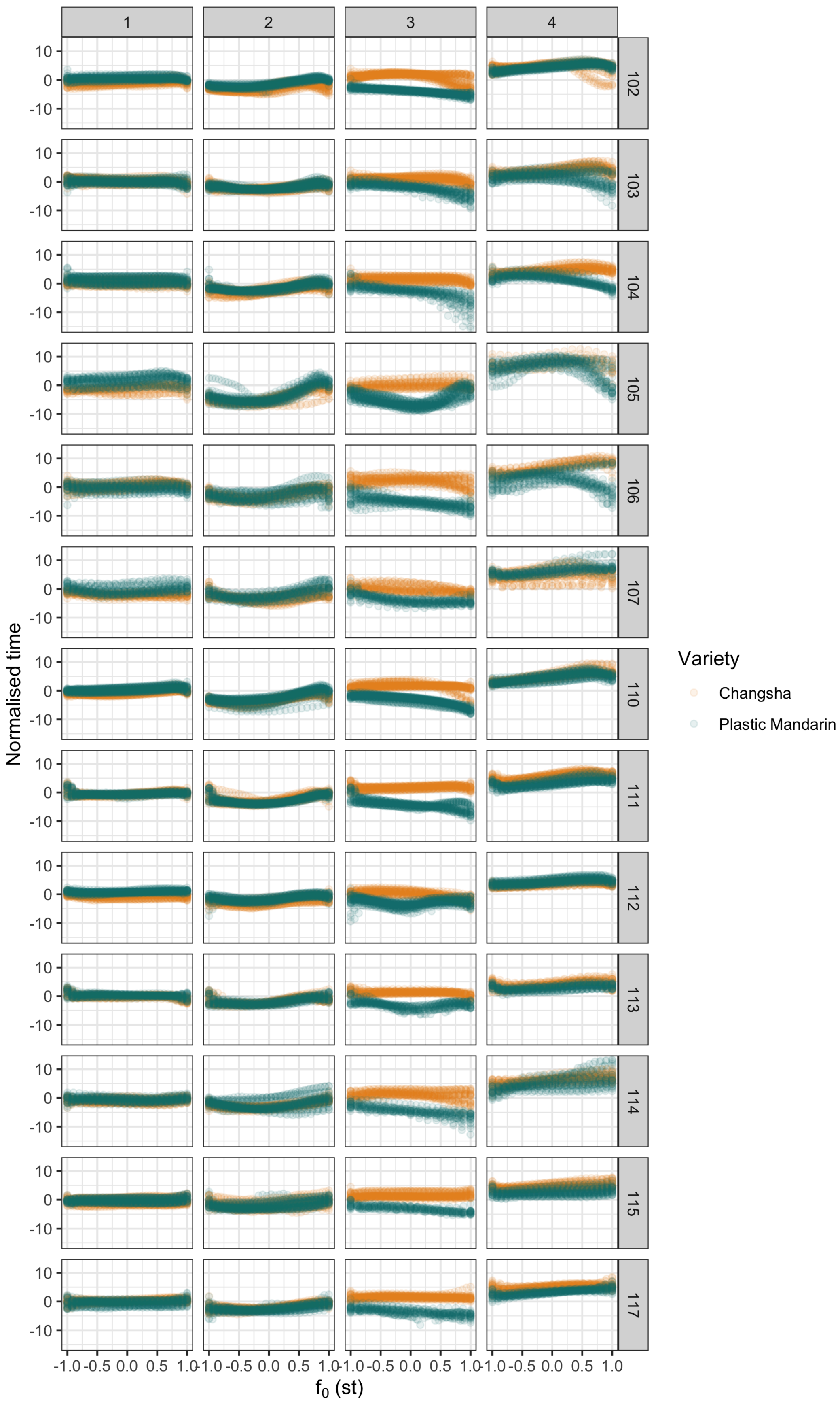


Figure 3.46: Normalised f_0 data of Changsha (yellow) and Plastic Mandarin (teal) by tone categories and speakers. Adjusted uniform speaker mean was used as the reference frequency in the semitone conversion.

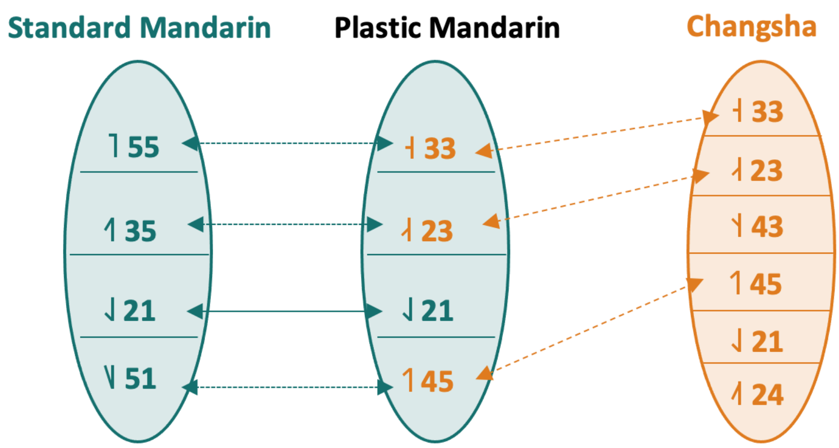


Figure 3.47: Illustration of the potential formation of Plastic Mandarin tones.

dipping form ˨˩˦. Standard Mandarin Tone 3 in connected speech is typically a low falling tone but has a rising tail in, for example, stressed utterance-final positions. Hartman (1944) indeed chose the low falling form as the basic form. Shih (1988) had the same choice, though intended for *guóyŭ*, Taiwanese Mandarin.

The historical Middle Chinese tone categories still play a role in such "tonal borrowing" in that a borrowed tone specification in Plastic Mandarin was from the corresponding historical tone category in Changsha. For example, the Tone 4 in Plastic Mandarin is qù tone (III), which acquired the tone value of the high register qù tone (IIIa) of Changsha. The analysis also suggests that Plastic Mandarin speakers did not simply copy Changsha tones, but adjust mostly their pitch levels in constructing a new tone system.

3.8 Summary

This chapter reports on the duration and f_0 contours of monosyllables produced in isolation in three varieties: Plastic Mandarin, Standard Mandarin, and Changsha. The duration of Plastic Mandarin tones is similar to the corresponding Standard Mandarin tones except for neutral tone and Tone 3. In general, the dipping Tone 3's in Standard Mandarin are longer than all other tones, while the neutral tones in Standard Mandarin are shorter than all other tones. The variation in the duration of citation tones in Standard Mandarin is much larger than that in Plastic Mandarin and in Changsha. Tone categories evolved from the historical qù (III) tone including Tone 4 in both Mandarins and Tone 4 and 5 in Changsha tend to have shorter duration. Syllable structure has some influence in the duration of a monosyllable.

Plastic Mandarin tones reveal a *Mix and Match* mechanism in contact-induced tonal variation and change. The emergence of Plastic Mandarin is closely related to the popularisation of Standard Mandarin in the Changsha city where the vernacular Changsha used to dominate. Auer (1998) indicates that levelling on the standard-dialect dimension often leads to dialects converging towards the standard variety. Plastic Mandarin can be considered as a product of levelling in language contact, which was much closer to Standard Mandarin than Changsha. The segmental inventory of Plastic Mandarin is highly similar to that of Standard Mandarin, which make it intelligible to all Mandarin speakers. Local varieties such as Changsha with distinct phoneme inventory are hardly intelligible to Mandarin speakers. In terms of citation tones, Plastic Mandarin blends elements from both Standard Mandarin and Changsha. It has the same tone categories as Standard Mandarin. The phonetic specification of each tone, however, tends to acquire and adapt the pattern of corresponding Changsha tone (except for Tone 3). The four lexical tones of Plastic Mandarin are: a mid level tone, a low-to-mid rising tone, a low falling tone, and a high rising tone. The locally tuned tone patterns enable Plastic Mandarin to be a recognisable Xiang (Hunan) accent.

As many young Plastic Mandarin speakers in Changsha are 'trilingual' in Standard Mandarin, Plastic Mandarin and Changsha, it is worthwhile in a future study to examine how they produce Standard Mandarin, so as to further investigate the origin of Plastic Mandarin tones from a language acquisition perspective.

4

Neutral Tones in News Broadcast Speech Corpus

Contents

4.1 Introduction

Neutral tone in Mandarin Chinese has been an object of studies where opinions diverge—discussion has been centred around topics such as whether neutral tone is a fifth lexical tone, whether it is underlyingly toneless or underspecified, whether there is a pitch target, whether it is some tone sandhi, and so on. Chapter 2 §2.5 has given a detailed introduction on Mandarin neutral tone and laid out the context for this chapter. It is established that neutral tone syllables are weak syllables exhibiting contextually conditioned pitch realisations. Its surface variability tends to have an impact on its acquisition, which was reported to be a protracted process not completed until the age of five (Tang et al., 2019).

In light of the background of inconclusive research on Mandarin neutral tone, this chapter lays the foundation for our understanding of the realisation of neutral tone through a corpus-based study, as most previous studies on neutral tone were either impressionistic or experimental investigations based on very limited read speech data. This chapter examines the acoustics of neutral tones in a corpus of broadcast news speech in Standard Mandarin, where the speech of broadcasters is usually considered the most standard. The corpus approach has the advantage of acquiring a large amount of data from a wide range of contexts in connected natural speech, which enables us to check the validity of prior generalisations of the variable phonetic realisations of neutral tone.

This chapter addresses the following specific questions:

1) Is neutral tone syllable shorter than its preceding lexical tone syllable?
2) How does the preceding lexical tone affect the f_0 contour of a neutral tone or a sequence of neutral tones?
3) Is there a pitch target for neutral tone syllables in Standard Mandarin?

Unfortunately no corpus of equivalent or larger size in Plastic Mandarin was available or openly accessible. The analysis of neutral tone in Standard Mandarin broadcast news speech established a benchmark for comparison with the Standard and Plastic Mandarin fieldwork data in Chapter 5. This chapter also suggests a probabilistic target-approaching view of tone contours and contextual tonal variation.

4.2 Method

4.2.1 The Hub4-NE Mandarin Corpus

The 1997 Mandarin Broadcast News Speech (HUB4-NE, LDC98S73) corpus (S. Huang et al., 1998b), distributed and hosted on the Linguistic Data Consortium, was chosen as the primary site to conduct exploratory analysis on Mandarin speech based on two criteria: accessibility and availability of corresponding transcripts. This corpus consists of 30 hours of broadcast news recordings of a single channel and 16,000 Hz sample frequency, from radio sources including Voice of America (VOA), China Central TV (CCTV), and KAZN-AM in Los Angeles, California. Corresponding transcripts (S. Huang et al., 1998a) are also released on the Linguistic Data Consortium. Moreover, the speech data were automatically time aligned by syllables with corresponding transcripts and tone category labels, which were obtained from Jiahong Yuan (Ryant et al., 2014; Yuan and Liberman, 2015). The majority of the speech data were from news broadcasters whose speech of Standard Mandarin can be considered the most standard and less susceptible to change.

4.2.2 Corpus Queries

This study utilised the HUB4 Mandarin corpus that was readily accessible in the Phonetics Laboratory of the University of Oxford to explore speech sequences with neutral tone syllables. The corpus contains recordings from 27 speakers, but some speakers were excluded in this study: one speaker was excluded because most of his recordings have background noise and music; another four speakers with a non-Standard Mandarin accent [1] and one speaker whose recordings mainly involve interviewing other accented speakers were also excluded. In total, recordings were used from 21 speakers, 13 male and 8 female.

The study mined all speech intervals containing a neutral tone syllable or multiple consecutive neutral tone syllables from the corpus, inspired by Y. Chen and Xu (2006). In this corpus, utterances containing three or more consecutive

[1] Judgements of the non-standard accents were based on the author's native speaker experience. Yuan and Liberman (2015) also pointed out the non-standard accents in this corpus.

Rank	Count	Neutral tone syllable		
1	7137	*de*	的	/tə/
2	1441	*le*	了	/lə/
3	691	*men*	们	/mən/
4	348	*ge*	个	/kə/
5	345	*ne*	呢	/nə/
6	217	*me*	么	/mə/
7	172	*de*	地	/tə/
8	147	*zhe*	着	/ʈʂə/
9	128	*hou*	候	/xoʊ/
10	111	*zi*	子	/tsz̩/

Table 4.1: Top 10 neutral tone syllables in the HUB4 corpus. Ones shaded in grey were selected for analysis.

neutral tone syllables were not found. The study, hence, concentrates on two groups of speech intervals: disyllabic occurrences notated as [X-N], in which a neutral tone syllable represented by N follows a non-neutral tone syllable X, which can be any of the four lexical tones of Mandarin; and trisyllabic occurrences [X-N_1-N_2], in which two consecutive neutral tone syllables follow a lexical tone syllable.

Due to the occasional errors of the automatic tone labels in the corpus (Ryant et al., 2014) and the inconsistencies in the use of neutral tone in many expressions, not all of the audio clips were analysed. [X-N] occurrences that end with five frequently-used functional morphemes *de*, *le*, *men*, *zhe*, and *zi* were selected for analysis, and they accounted for about 80% of all [X-N] occurrences. Table 4.1 listed the 10 most frequent neutral tone syllables in the corpus. Among them *ne* and *me* are question particles and they were not included in order to avoid the effect of question intonation. These selected morphemes except *men* occurred in various [X-N] sequences in which X covers instances of all four lexical tones. They are considered 'toneless' (X. S. Shen, 1992), although they have their corresponding full tone variants or homographs, which usually occur much rarely and in different contexts. Furthermore, only [X-N] utterances with at least two occurrences in the corpus were included in the analysis. Examples of disyllabic occurrences in the corpus are shown in Table 4.2.

[X-N_1-N_2] occurrences were further divided into two types depending on the middle syllable N_1: in type I occurrences, N_1 was a lexical morpheme, while in type II it was a functional morpheme. These are illustrated with examples in Table 4.3. The last neutral tone syllable N_2 was always a functional morpheme and in most cases it was one of the five selected morphemes.

Short audio clips of the [X-N] and [X-N_1-N_2] occurrences were extracted from the relevant audio files of the HUB4 speech corpus given the syllable start and end times in the time-aligned transcripts using SoX (Bagwell et al., 2015). A separate set of audio clips containing only the neutral tone syllable in the disyllabic utterances and the [N_1-N_2] part of the trisyllabic utterances was also extracted. All sound clips were listened to and those with background music or noise that led to unreliable speech-text alignment were discarded. Extremely short clips that generate no pitch track or clips that are almost voiceless throughout the utterance with no more than two f_0 points were excluded from the pitch analysis, but these were included in the duration analysis.

In summary, the duration analysis included 8417 disyllabic tokens and 199 trisyllabic tokens. The final data for pitch analysis comprised of 7208 tokens of neutral tone syllables from the [X-N] utterances and 180 tokens of two consecutive neutral tone syllables from the [X-N_1-N_2] trisyllabic utterances. The number of the sound tokens for pitch analysis by morphemes and X tone categories were presented in Tables 4.2 and 4.3. There is an uneven number of tokens of each phrase in the data.

In the following analysis, Standard Mandarin lexical tones T1, T2, T3, and T4 will be represented in Duanmu (2000a)'s tone specification: H, LH, L, and HL, instead of the arbitrary number index, so that we can easily have access to the prototypical shape of the preceding tone.

Particle		[T1-N]	[T2-N]	[T3-N]	[T4-N]	Count
de 的	possessive	[tʰa tə] *tā de* his/her	[ʐən tə] *rén de* people's	[ni tə] *nǐ de* your	[ta tə] *dà de* big	5250
le 了	perfective	[tʂʰu lə] *chū le* excited	[laɪ lə] *lái le* came	[tsoʊ lə] *zǒu le* went	[tɑʊ lə] *dào le* arrived	1168
zhe 着	durative	[tʂʰwan tʂə] *chuān zhe* wearing	[jɛn tʂə] *yán zhe* following	[joʊ tʂə] *yǒu zhe* having	[kʰɑʊ tʂə] *kào zhe* leaning	627
zi 子	diminutive	[tʂʰɤ tsz̩] *chē zi* car	[fɑŋ tsz̩] *fáng zi* house	[kɑʊ tsz̩] *gǎo zi* script	[ɻ̩ tsz̩] *rì zi* days	93
men 们	plural	[tʰa mən] *tā men* they	[ʐən mən] *rén men* people	[wo mən] *wǒ men* we		70
Count		1654	1750	1210	2594	7208

Table 4.2: Disyllabic occurrences [X-N] with a neutral tone.

	[T1-N_1-N_2]	[T2-N_1-N_2]	[T3-N_1-N_2]	[T4-N_1-N_2]	Count
I	[ɕiɛn ʂəŋ tə] *xiān sheng de* gentleman's	[pʰɤŋ joʊ mən] *péng you men* friends	[naɪ naɪ tə] *nǎi nai de* grandma's	[ʂɻ̩ tɕʰiŋ lə] *shì qing le* matter	60
II	[tʰa mən tə] *tā men de* their	[xaɪ tsz̩ tə] *hái zi de* child's	[wo mən tə] *wǒ men de* our	[jɑŋ tsz̩ tə] *yàng zi de* look's	120
Count	49	43	56	32	180

Table 4.3: Trisyllabic occurrences [X-N_1-N_2] with two consecutive neutral tones.

4.2.3 Acoustic measurement

For all the recordings of each speaker, f_0 estimates (in Hertz) in 10 millisecond intervals of voiced regions were obtained using the `get_f0` program (Talkin, 1995) from the ESPS package[2] (Entropic Research Laboratory, 2006), for the purpose of calculating the average f_0 of each speaker. Then, the same f_0 measurement was generated for all neutral tone tokens from the disyllabic and trisyllabic utterances introduced in §4.2.2. The duration of each syllable can be calculated using the temporal information in the time-aligned transcripts.

4.2.4 Data analysis

The methodology for analysis in this chapter is consistent with that used in Chapter 3, in which the f_0 contour analysis started with data normalisation, followed by orthogonal polynomial modelling, and completed with statistical inference using LME models. In this section, the procedure of data analysis is outlined; for more details, refer to §3.4. Minor discrepancies in the implementation between the two chapters will be addressed.

4.2.4.1 Data normalisation

The f_0 estimates and their corresponding times were normalised in order to facilitate inter-speaker and inter-token comparison. Using the same approach in §3.4.2.2, the time of each f_0 of a token was linearly scaled to the range of $[-1, 1]$. Each f_0 in Hz were normalised to the corresponding speaker mean $\bar{f}_s$ and centred around zero:

$$f_0^* = \frac{f_0}{\bar{f}_s} - 1 \tag{4.1}$$

In Equation 4.1, a normalised f_0 of zero ($f_0^* = 0$) equates to the defined speaker mean. This normalisation method was also used by Maidment and Lecumberri (1996), Grabe et al. (2007), and others. Here the speaker mean f_0 was defined

[2]Entropic Signal Processing System (ESPS) is released by the Phonetics Laboratory, University of Oxford, and can be downloaded from http://www.phon.ox.ac.uk/releases.

over the average of **all** f_0 measurements of the voiced regions in all recordings of a speaker, as in Equations 4.2.

$$\bar{f}_s = \frac{1}{n}\sum_{j=1}^{n} f_{0j} \tag{4.2}$$

where j is the index of each f_0 measurement of a given speaker, and n is the total number of f_0 measurements in the recordings of a speaker.

4.2.4.2 Orthogonal polynomial modelling

The f_0 contours were modelled using the best-fit sum of weighted Legendre polynomials. The selection of the number of the coefficients or the degree of the polynomials is an analytical decision depending on how complex the shape is and what components we are interested in. While three coefficients encoded the informative modes of variations in the pitch patterns of monosyllables, it is likely that more coefficients are needed for longer utterances. For instance, Figure 4.1 illustrates that a cubic polynomial with four coefficients is indispensable to the major trend of pitch change in the trisyllabic utterance /ny mən tə/, in that it captures the valley at the end of the first syllable *nǚ* (Tone 3, a low tone in Standard Mandarin) and the peak at the end of the second syllable *men* (neutral tone). The likelihood ratio tests showed that the cubic model significantly improves the fit over the model with a quadratic term ($\chi^2(1) = 8674.5, p < .001$), while the model with a quartic term is not significantly different from this cubic model ($\chi^2(1) = 14.96, p = 0.37$). The cubic model in this case, though, tends not to be sensitive to microprosodic variations such as the potential obstruent-related perturbation represented as a small spike at the beginning of the third syllable *de* in Figure 4.1.

Figure 4.1 also shows that the f_0 contour of the neutral tone sequence, the part after the dashed vertical line at 14 cs, is not complicated—a rise before a fall. Having examined such graphs for all trisyllabic tokens in the dataset, it was decided to fit a quadratic model to the contour of the [N_1-N_2] part of each token. The c_i coefficients of the polynomial models were obtained using the *polyfit* function in GNU Octave (Eaton et al., 2020), an open-source version of Matlab, and converted to a_i coefficients. The details of Legendre polynomials are presented in §3.4.3.

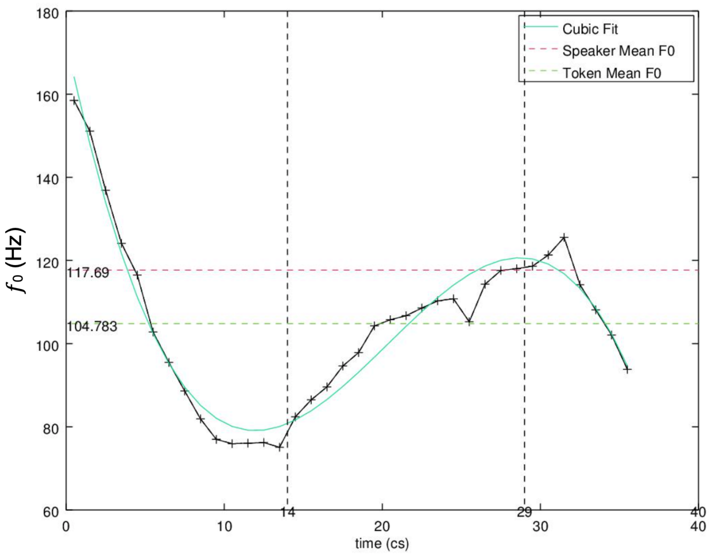


Figure 4.1: The best-fit cubic polynomial model (in cyan) for the trisyllabic token *nǚ men de* /ny mən tə/ 'girls" in Standard Mandarin from the HUB4 corpus. f_0 measurements are marked in crosses and interpolated in black line. Dashed vertical black lines demarcate the syllables. The mean f_0 for this token and the speaker is marked for reference.

4.2.4.3 Statistical technique

The first research question concerns the comparison of syllable duration between the neutral tone syllable(s) and the lexical tone syllable in the same token. Paired-samples t-tests were used and the null hypothesis for statistical analysis in this case is that there is no significant difference in syllable duration between the neutral tone syllable(s) and the lexical tone syllable in the same token. The second research question investigates whether and how the tone category of the preceding lexical tone syllable predicts aspects of the tone shape of the neutral tone or neutral tone sequence, using LME analysis. The specific statistical models are presented in §4.4.1.

4.3 Neutral Tone Duration

4.3.1 Results of syllable duration

Table 4.4 lists the mean and standard deviation of the duration of both syllables of the disyllabic utterances by the neutral tone suffix, and Figure 4.2 draws the corresponding kernel density function, a smoothed alternative to histogram for continuous data. The boxplots show the interquartile range and the median of the duration of all lexical tone and neutral tone syllables. From 4.4, the overall mean of the neutral tone is about 72% of the mean duration of full lexical tones. In order to examine whether the mean difference in syllable duration is significant, a paired-samples upper-tailed t-test was conducted to compare the duration of the neutral tone with its preceding lexical tone in the same utterance. There was a significant difference in the duration of lexical tones ($\mu = 0.158, \sigma = 0.046$) and neutral tones ($\mu = 0.114, \sigma = 0.049$); $t(8416) = 68.169, p < .001$. This suggests that neutral tone syllables in general are shorter than their preceding lexical tone syllables in these disyllabic utterances.

Suffix	Lexical tone X $\mu(\sigma)$	Neutral tone N $\mu(\sigma)$	Count
de /tə/ 的	0.161 (0.045)	0.109 (0.047)	6378
le /lə/ 了	0.157 (0.040)	0.122 (0.049)	1295
men /mən/ 们	0.109 (0.033)	0.141 (0.052)	542
zhe /ʈʂə/ 着	0.167 (0.045)	0.129 (0.055)	122
zi /tsɿ/ 子	0.234 (0.057)	0.185 (0.061)	80
Overall	**0.158** (0.046)	**0.114** (0.049)	8417

Table 4.4: Mean and standard deviation of the duration (s) of syllables in the [X-N] disyllabic utterances.

Table 4.4 also suggests that the shorter duration of neutral tone syllable may relate to the syllable structure. Among the five neutral tone syllables, suffix *men* /mən/ (highlighted in grey), the only one with a nasal coda in the syllable structure, has the reversed result where the neutral tone syllable is on average significantly longer ($t(541) = -15.471, p < .001$) than its preceding lexical tone syllable, which consist of mostly (82.3%) CV syllables such as /t^{h}a/(217/542) and /wo/ (229/542). This is illustrated in Figure 5 through the rightward shift of the yellow peak relative to the blue peak in the *men* graph compared to other graphs, where the neutral tone syllables have a CV structure. Although F. Wu and Kenstowicz (2015) reported that hardly any significant difference between the duration of CVN syllables and CV

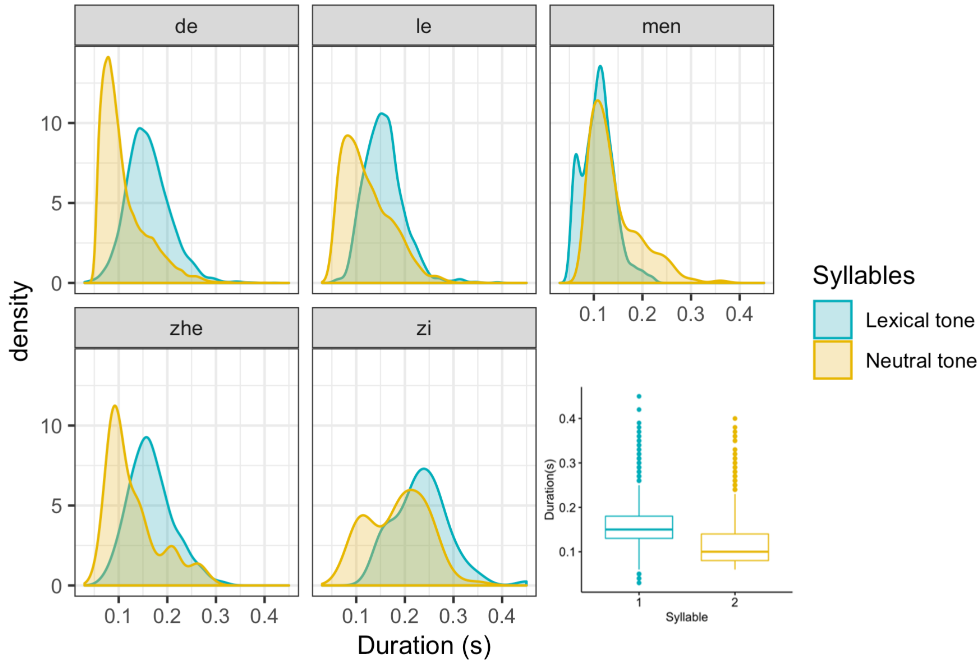


Figure 4.2: Kernel density functions and boxplots for the duration (s) of syllables in [X-N] disyllabic utterances.

Lexical tone X $\mu(\sigma)$	Neutral tone 1 N_1 $\mu(\sigma)$	Neutral tone 2 N_2 $\mu(\sigma)$
0.160 (0.061)	0.122 (0.033)	0.108 (0.051)
$t(198) = 8.584, p < .001$		
	$t(198) = 3.665, p < .001$	

Table 4.5: Mean and Standard Deviation of the duration (s) of syllables in the trisyllabic utterances and two paired-samples upper-tailed t-test results.

syllables was found, we cannot yet hastily conclude that it is the neutral tone status instead of simpler syllable structure that leads to the relatively shorter duration.

Table 4.5 and Figure 4.3 present the durational information of syllables in trisyllabic utterances. From the initial syllable to the last syllable in an utterance, the average duration of each syllable decreases, as shown in Table 4.5. The downward shifting boxplots capture the same trend. In Figure 4.3 the graph on the left with three non-overlapping peaks shows the distribution of the duration of syllables in different positions. A paired-samples upper-tailed t-test was conducted to compare

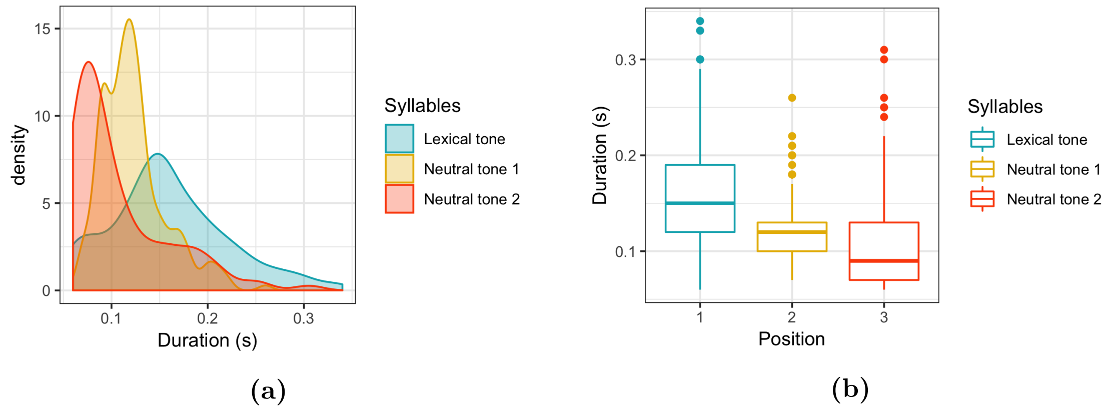


Figure 4.3: Kernel density functions and boxplots for the duration (s) of syllables in the [X-N_1-N_2] trisyllabic utterances.

the duration of the middle neutral tone with its preceding lexical tone and the duration of the two neutral tones in the same utterance. The p values are less than .001, which indicates that in these utterances the last neutral tone on average is significantly shorter than the middle neutral tone and the middle neutral tone on average is significantly shorter than the preceding lexical tone.

The majority (76.9%) of the last neutral tone syllables in these utterances are de /tə/, and about half (47.7%) of the middle neutral tone syllables are men /mən/, the relatively long neutral tone in the disyllabic sample. This may explain why the second neutral tone on average is shorter than the first neutral tone in the trisyllabic dataset when there are two consecutive neutral tones.

The above results of the duration of syllables suggest that neutral tone syllables tend to be shorter than their preceding lexical tones, especially when the neutral tone syllables have CV structure.

4.3.2 Discussion

Consistent with most previous studies, the duration of a neutral tone syllable is in general significantly shorter than its preceding lexical tone syllable. But in Table 5, we also find that neutral tone syllable *men* /mən/ is significantly longer than its preceding syllable. Similarly in K. Huang (2018)'s study of Taiwan Mandarin, no evidence was found that neutral tone *men* is reduced in duration while neutral tone syllables *zi*, *zhe*, *de*, *le*, *ge* are reduced in duration. Such findings suggest that syllable structure may play a role in the duration of the neutral tone because one major difference between *men* and the other neutral tone syllables is that *men* has a slightly more complex structure of a closed syllable with a nasal coda.

Both Duanmu (1994) and F. Wu and Kenstowicz (2015) suggested that no evidence was found that monosyllables with a nasal coda [CVN] are longer than monosyllables with only a single vowel [CV]. Relative to CV syllables, the nuclear vowel is shorter in CVN structure (Duanmu, 2007; F. Wu and Kenstowicz, 2015). Their studies, however, were based on Mandarin full lexical tone syllables instead of neutral tone syllables. Similar vowel shortening in closed syllables is also observed in many languages, analysed as a phonological process that neutralises the vowel length contrasts as in Turkish, or as a phonetic process when the vowel duration decreases in closed syllables across vowel length distinction as in Finnish (Flemming, 2001).

One explanation proposed here is that when a syllable is unstressed, the duration of the syllable is susceptible to its syllable structure (the number and type of segments), and that syllables with a coda is highly likely to be longer than CV syllables if the nasal coda is not completely coarticulated with the preceding vowel. In other words, unstressed neutral tone provides a context for shorter vowels in CV syllables, restraining open-vowel lengthening observed in stressed syllable, similarly to closed syllables. The most frequently used functional morphemes such as *de*, *le*, *zi*, *zhe*, and *ge*, which are the most typical examples of neutral tone, are all CV syllables. These unstressed neutral tone syllables, mostly with a mid central vowel, are shorter in duration. This has led to the claim in some previous studies that neutral tone syllables were shorter than other full lexical tone syllables. A more promising conclusion, therefore, is that the duration of a neutral tone syllable is sensitive to the syllable structure, and that neutral tone syllables with CV structure are generally shorter in duration than their preceding lexical tone syllables.

Neutral tones have also been reported to be correlated with lexical frequency independent of other factors. More frequent neutral tones tend to be shorter in duration, lower in pitch, and weaker in intensity (Kong & Wu, 2019). Productive functional morphemes, particularly *de* and *le*, the most frequently-used morphemes, are thus shorter than most other syllables.

4.4 Neutral Tone Pitch Contours

4.4.1 Statistical method

Using the same approach as in Chapter 3, the f_0 contours of neutral tone and neutral tone sequences are modelled using three coefficients, i.e. c_0, c_1, and c_2, characterising the Average Pitch Height, Slope, and Curvature of a contour. The detailed procedure of obtaining these coefficients is introduced in §3.4.2 and §3.4.3 in Chapter 3.

Linear Mixed-Effects models were built to investigate whether and how the preceding tone influence the contour of neutral tone. The dependent variable is a c_i coefficient, the independent fixed-effect predictor was the preceding tone category (4 levels: H, LH, L, and HL), and the random effect was speakers. The five different neutral tone particles in disyllabic phrases were also included as a random effect. Only random intercepts were included in the models because random slopes were not properly supported by the data and the models failed to converge if adding them. The significance of fixed effects was estimated using the ANOVA function from `lmerTest` package, which reports omnibus effects for multilevel factors using F-tests. The Kenward-Roger Approximation for the denominator degrees of freedom in the `pbkrtest` package (Halekoh & Højsgaard, 2021) was used to derive p values. This is a slightly more conservative method than the alternative Satterthwaite Approximation and can work well with unbalanced data or small sample (Halekoh & Højsgaard, 2014). For the significant effect, Bonferroni adjusted post-hoc pairwise comparisons were conducted using `emmeans` package (Lenth, 2021), and the estimated Least-Squares mean outperforms the regular mean in unbalanced designs.

4.4.2 f_0 patterns of monosyllabic neutral tones

Table 4.6 summarises the significance level of aspects of shape predicted by the preceding lexical tone category in the linear mixed effects models. The p values

Shape	Predictor	df1	df2	F	*p*
HEIGHT	Preceding tone	3	7112.5	59.011	< .001
SLOPE	Preceding tone	3	7153.7	371.06	< .001
CURVATURE	Preceding tone	3	7195.9	402.14	< .001

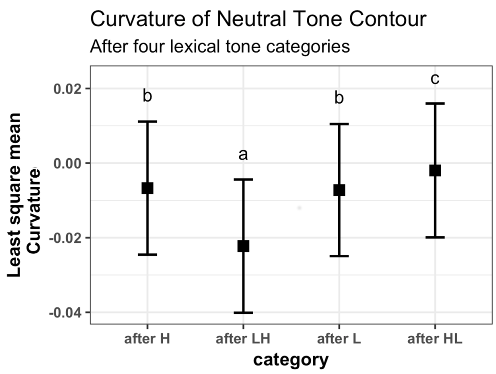
Model: Shape ∼ Preceding tone + (1 | Speaker) + (1 | Particle)

Table 4.6: Fixed effects in the linear mixed-effect models for aspects of pitch contour of monosyllabic neutral tone particles.

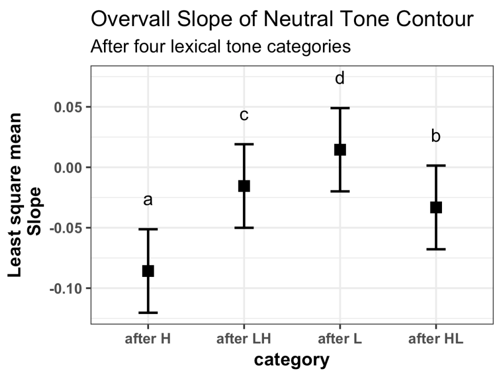
are all less than .001, which suggest the preceding tone exerts some statistically significant influence on the shape of f_0 contour of the neutral tone.

Bonferroni post-hoc tests were then conducted on these models. Figure 4.4 further presents the Least Squares means calculated from the linear mixed effect models. In Figure 4.4, most of the least squares means of c_i coefficients across different tonal contexts in a graph do not share a letter, which indicate that the preceding tone has a statistically significant impact on the pitch height, overall slope, and curvature of the contour of neutral tone. For example, only the neutral tone after L tone has a positive c_1 Least Squares mean (labelled d in graph (b)), a rising linear trend, though it still has a relatively low average pitch (labelled a in graph (c)) compared to neutral tones in other tonal contexts. The neutral tone after H tone has a distinctively low negative c_1 (labelled a in graph (b)), indicating a relatively large decline in f_0. The neutral tone after LH tone has the lowest negative c_2 (labelled a in graph (c)), representing a more convex shape, while the neutral tone after HL tone with a c_2 close to 0 (labelled c in graph (c)) has a flatter shape.

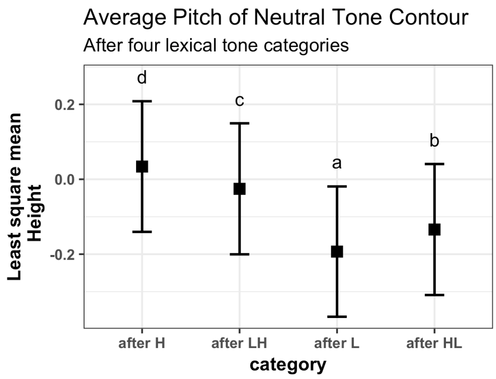
Figure 4.5 illustrates the aspects of shape that are represented by these Least Squares means by reproducing the modelled contours of neutral tone. These simulated neutral tone contours are slightly different from some of the descriptions in the previous studies summarised in §2.5. Instead of assigning a pitch level impressionistically, Figure 4.5 offers an acoustically based detailed prototype of neutral tone contours in different tonal contexts. The contours start at various pitch heights due to the different preceding tones, and end within a narrower pitch range.



(a) c_2 predictions



(b) c_1 predictions



(c) c_0 predictions

Figure 4.4: Least Square Means and the pairwise significance of c_i coefficients of the neutral tone contours in the [X-N] disyllabic utterances. The square dots indicate the Least Squares mean. Error bars indicate the 95% confidence interval of the Least Squares mean. Means sharing a letter are not significantly different (Bonferroni-adjusted comparisons).

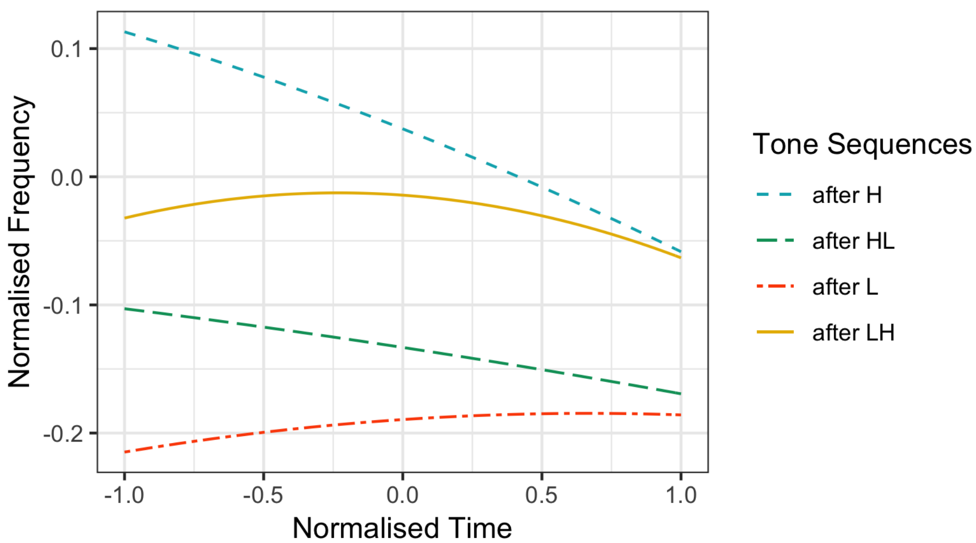
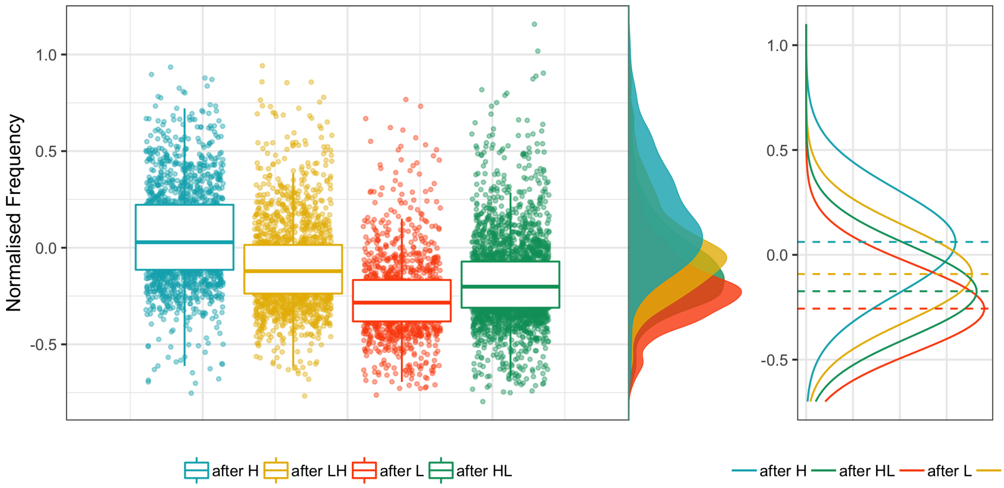
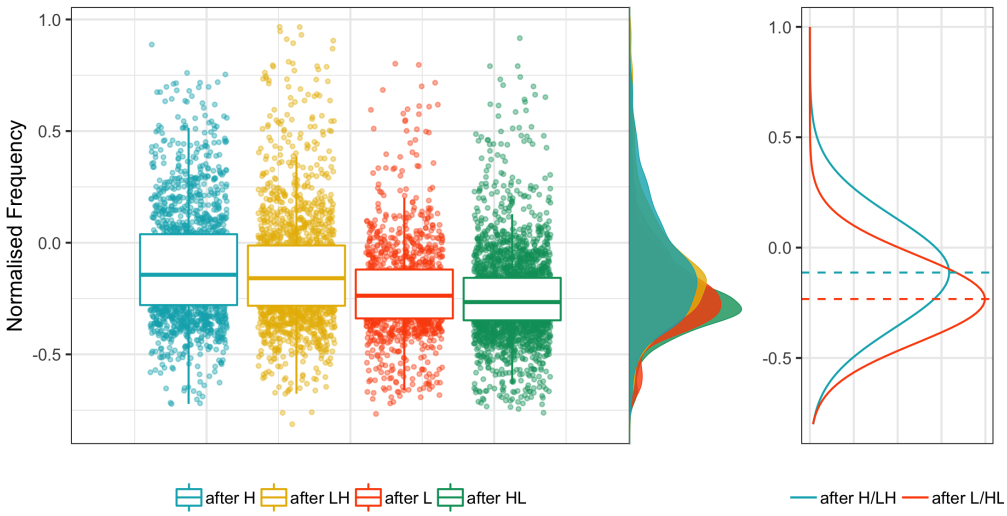


Figure 4.5: Simulation of single neutral tone contours when preceded by different tones.



After H	$\mu = 0.06$	$\sigma = 0.25$
After LH	$\mu = -0.09$	$\sigma = 0.23$
After L	$\mu = -0.26$	$\sigma = 0.21$
After HL	$\mu = -0.17$	$\sigma = 0.22$

Figure 4.6: Distribution of modelled starting pitch of neutral tones in [X-N] disyllabic utterances.



After H	$\mu = -0.11$	$\sigma = 0.25$
After LH	$\mu = -0.23$	$\sigma = 0.20$

Figure 4.7: Distribution of modelled ending pitch of neutral tones in [X-N] disyllabic utterances.

Shape	Predictor	df1	df2	F	p
HEIGHT	Preceding tone	3	172.98	5.04	$< .001$
SLOPE	Preceding tone	3	174.58	19.11	$< .001$
CURVATURE	Preceding tone	3	151.87	7.15	$< .01$

Model: Shape ∼ Preceding tone + (1 | Speaker)

Table 4.7: Fixed effects in the linear mixed-effect models for aspects of pitch contour of two consecutive neutral tones in [X-N_1-N_2] trisyllabic utterances.

Figure 4.6 and Figure 4.7 show the distribution of normalised f_0 at the beginning and the end of the voiced part of the neutral tone of all the utterances respectively. To the right of the boxplots, kernel density functions and single Gaussian distributions for the four contexts are shown to model the pitch variation at this time point.

The contours after H and after LH, with higher starting pitch (Figure 4.6), tend to end at a similar mid range pitch height (Figure 4.7), and the contours after L and after HL, with lower starting pitch, tend to approximate to a similar lower pitch. The end point tone specification, H or L, of the preceding tone tends to correlate with the end point of the neutral tone. In Figure 4.7, two Gaussian distributions ($\mu = -0.11, \sigma = 0.25; \mu = -0.23, \sigma = 0.20$) are shown to present the two main patterns at the end of f_0 contours. The observed means of the Gaussian distributions are slightly different from the Least Squares Means computed in the models.

4.4.3 f_0 patterns of two consecutive neutral tones

Table 4.7 summarises the results of the linear mixed effects models for the contours of two consecutive neutral tones. The preceding tone category leads to significant differences in average height, overall slope, and curvature of the contour of sequences of two neutral tones.

The prototype of the contours of two consecutive neutral tones in four different contexts is shown in Figure 4.8, based on the Least Squares Means of c_i coefficients calculated from the linear mixed effect models. The shape of contours in the normalised time range from -1 to 0.25 resembles those in Figure 4.5, where neutral

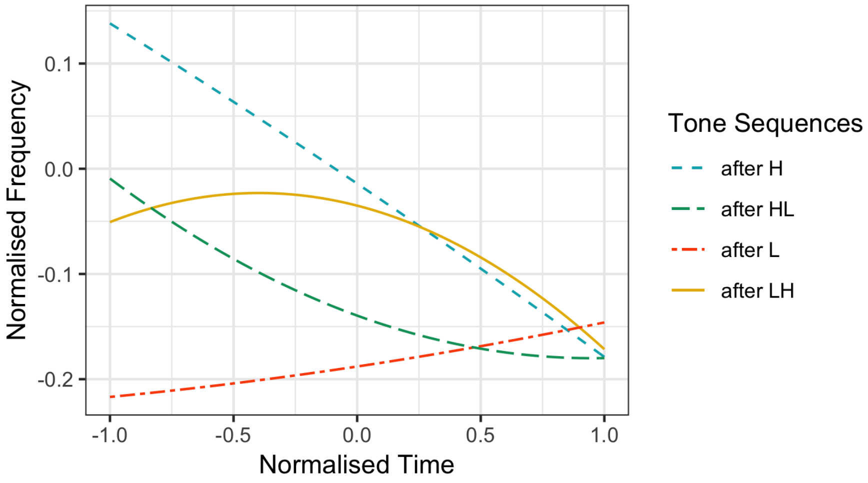
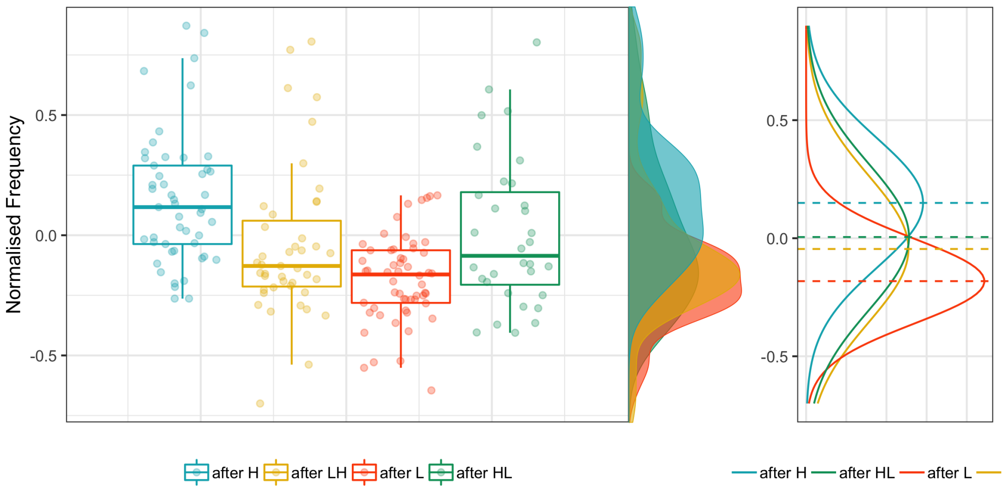


Figure 4.8: Simulation of contours of two consecutive neutral tones in different contexts.

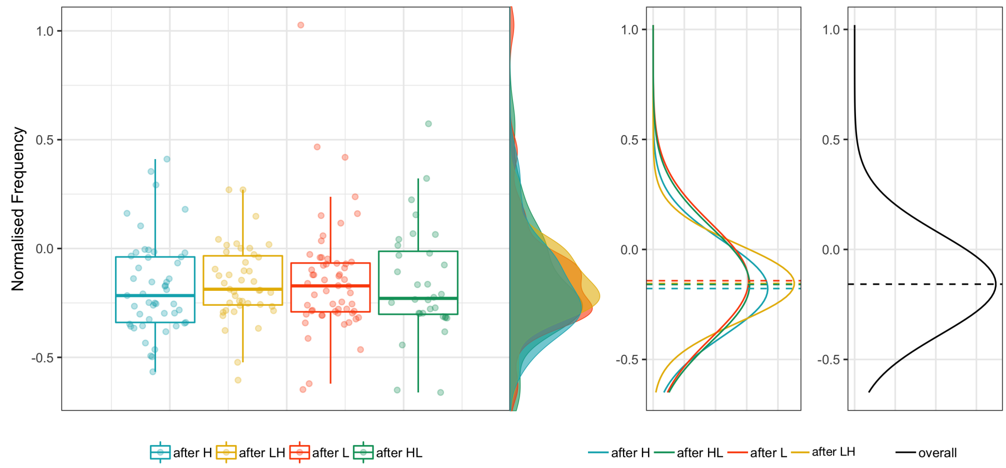


After H	$\mu = 0.15$	$\sigma = 0.27$
After LH	$\mu = -0.05$	$\sigma = 0.31$
After L	$\mu = -0.18$	$\sigma = 0.18$
After HL	$\mu = -0.005$	$\sigma = 0.31$

Figure 4.9: Distribution of modelled starting pitch of the first neutral tone in [X-N_1-N_2] trisyllabic utterances.

tone contours converge at two pitch levels: one at about -0.05, i.e. 5% lower than the average pitch of the speakers, and the other about -0.17, 17% lower than the average pitch. Figure 4.8 shows that the four contours continue to converge to a single low pitch of around -0.18. From the normalised time range 0.25 to 1, these f_0 contours continue to converge at a similar low pitch.

Figure 4.9 and Figure 4.10 show the distribution of normalised f_0 at the beginning



Overall $\mu = -0.158 \quad \sigma = 0.23$

Figure 4.10: Distribution of modelled ending pitch of the second neutral tone in [X-N_1-N_2] trisyllabic utterances.

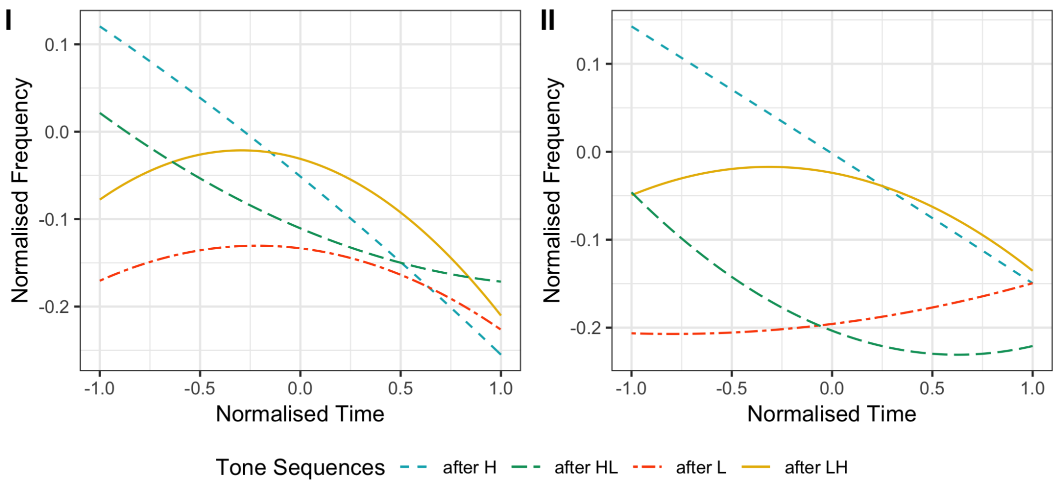


Figure 4.11: Simulation of two consecutive neutral tone contours of two types of phrases in different contexts.

and the end of the voicing part of the two consecutive neutral tones of all the utterances respectively. The Gaussian distributions across the four tonal contexts vary a lot at the beginning of f_0, but they are much more similar, as shown in Figure 13 where the distribution curves are superimposed on each other. A uniform Gaussian distribution ($\mu = -0.158, \sigma = 0.23$) was drawn based on all the end point f_0 data. Such convergence suggests that neutral tone also has a pitch target and it is fairly low, about 16% below the speaker's average f_0.

Figure 4.11 presents the simulated pitch curves for type I and type II phrases in Table 4B based on the Least Square means computed in the linear mixed effects models. These pitch contours may not be representative since the further classification into two types results in small and unbalanced samples for some tonal contexts. The curvature, overall slope and average height of the contours tend to be slightly different between the two types, but at the end of the f_0 contours, they converge towards a low pitch in both types.

4.4.4 Approaching a low target for neutral tone

The f_0 contours of two consecutive neutral tones in different tonal contexts in trisyllabic utterances show a converging trend towards a low pitch. The single Gaussian distribution of f_0 values at the end of the f_0 contours suggests that the average f_0 at the end of two consecutive neutral tones is about 16% lower than a speaker's average f_0. The contours of two consecutive neutral tones complete the patterns of a single neutral tone in disyllabic utterances that appeared to have two pitch targets. Instead of having varied pitch targets for neutral tones following different lexical tones, the conception of target approximation (Y. Chen & Xu, 2006) offers a convincing way to explain the findings. Different from the conclusion of Y. Chen and Xu (2006) that Standard Mandarin neutral tone has an underlyingly mid-level pitch target, however, our findings suggest a low level pitch target for neutral tones, although the pitch target for the converging trend at the end of two consecutive neutral tones seems slightly higher than the end pitch for lexical L tones. Our results are similar to K. Huang (2018), which identified a static pitch target in the mid-low to low range for Taiwan Mandarin.

4.4.5 Target with variance distribution

Although an underlying L target is assigned for neutral tones in our analysis, it cannot be neglected that the pitch realisation pattern of neutral tone is distinct from that of L tone where the f_0 usually reaches a low pitch by the end of the L tone. Neutral tone does not consistently trigger the classic Tone 3 sandhi phenomenon where a L tone is replaced by a LH tone when it precedes another lexical L tone.

This is demonstrated in (10a) where the first lexical L tone becomes LH tone /ny˧˥/ (in bold), when the following /tsɿ˨˩˦/ is a lexical L tone. When /tsɿ/ is a neutral tone in the utterance in (10b), the preceding L tone remains an L tone /kɑʊ˨˩/. Zhou (2018) reported that neutral tone occasionally trigger the Tone 3 sandhi, and (10c) is an example.

(10) a. *nǚ zǐ* 女子 /ny˧˥ tsɿ˨˩˦/ 'woman'

b. *gǎo zi* 稿子 /kɑʊ˨˩ tsɿ/ 'script'

c. *gú tou* 骨头 /ku˧˥ tʰəʊ/ 'bone'

The differentiation between lexical L tone and neutral tone, that both have L pitch targets can be accounted for by the concept of articulation strength (Kochanski et al., 2003) in the soft template model (Kochanski and Shih, 2000, 2001) or implementation strength in the PENTA model (Y. Chen and Xu, 2006; Y. Xu and Prom-on, 2014). In an intuitive sense, neutral tone tends to have much weaker strength and heavier carry-over influence from the preceding syllable.

In the target-approximation conception of tone, such weak strength can also be realised by a wider variance distribution in a probabilistic model (Blackburn and Young, 2000; Coleman et al., 2016) so that more variation can be tolerated. Keating (1990) proposes the window model for phonetic variation, in which a wider window permits more variable articulation and thus is more unconstrained and highly underspecified. The probabilistic extension of this model by Blackburn and Young (2000)'s work on articulator movements and Coleman et al. (2016)'s on F2 frequency of nasals employs Gaussian functions, encoding a central tendency. Here it is extended to model fundamental frequency contours using Gaussian models. As shown in the local distribution curves in Figure 4.9 and Figure 4.10 and the σ values given above, the variance at the end of a lexical L tone (Figure 4.9, $\sigma = 0.18$) is smaller than the variance at the end of the second neutral tone (Figure 4.10, $\sigma = 0.23$). This shows that variance distribution differentiates a lexical L tone from a neutral L tone, and that a lexical L tone has more specified and less variable f_0 realisation than a neutral L tone.

4.5 Summary

This chapter examined the acoustic properties of Standard Mandarin neutral tones in a News Speech corpus. The timimg data show that the duration of neutral tone is susceptible to syllable structure, and that open neutral tone syllables are in general shorter than their preceding lexical tone syllables. The f_0 patterns suggest a low pitch target with larger variance distribution for neutral tone syllables. A probabilistic target approximation model offers a better account than target interpolation in the phonetic implementation of a neutral tone. Chapter 5 will report on more linguistically controlled fieldwork data of neutral tone in both Standard Mandarin and Plastic Mandarin, in order to further investigate its contextual variation.

5

Neutral Tones in Two Mandarin Varieties

Contents

5.1 Introduction

Neutral tone in Standard Mandarin displays an immense amount of variability with regard to pitch patterns (review §2.5 for an introduction to neutral tone). Unlike other lexical tones with a canonical tone pattern, Chapter 3 has shown that there is no homogeneous citation form for different neutral tone syllables in Standard or Plastic Mandarin. Through an analysis of neutral tone syllables in broadcast news speech in Standard Mandarin, Chapter 4 revealed that the preceding tone category influenced the shape of f_0 contour of a neutral tone syllable, and that, despite different shapes, f_0 contours converged at the end of two consecutive neutral tones. This chapter extends the investigation by examining phrases containing neutral tone syllables in semi-spontaneous speech in the two Mandarin varieties. The main research questions in this chapter are:

1) How is a neutral tone or neutral tone sequence realised in connected speech of Plastic Mandarin?
2) How does the pitch contour and duration of a neutral tone syllable vary in contexts?
3) What are the differences in neutral tone between Plastic Mandarin and Standard Mandarin?

The first two questions are probed into with a cross-variety comparative perspective. If the emergence of Plastic Mandarin involves changing the pitch targets for the four lexical tones in Standard Mandarin, how does neutral tone with heterogeneous tone patterns in Standard Mandarin behave in Plastic Mandarin?

This chapter is organised as follows. Methods of speech data collection, acoustic possessing, and statistical techniques are firstly described and justified: §5.2.1 presents the rationale and details of the production experiment, §5.2.2 sets out the phrasal data with neutral tone syllables that were the main research targets, §5.2.3 illustrates the acoustic processing procedure, and §5.2.4 outlines the statistical analysis employing the Generalised Additive Mixed Models. Then the analyses and findings are organised in the areas of duration and pitch contours separately in the following sections. Under each section, the specific statistical models are introduced prior to the presentation of results and discussion.

5.2 Methods

The analyses in this chapter are based on recordings of the second task (B), the *word guessing game* embedded with carefully designed speech materials, in two Mandarin varieties. This section describes the design and implementation of the game, followed by a depiction of the acoustic processing and an overview of statistical technique.

Most details of the data collection have been introduced in Chapter 3 including the participants in §3.2.4, the premises in §3.2.2, the equipment in §3.2.5, and the general procedure of a recording session in §3.2.6.

5.2.1 Experiment design

The experiment design was primarily guided by the consideration of two key aspects: first, providing a communication or discourse context in which Plastic Mandarin could be elicited; second, creating a linguistic context where various tonal sequences were considered while maximally controlling for other linguistic factors.

5.2.1.1 Eliciting a non-standard accent: the disguised word-guessing game

Inspired by the map task (Anderson et al., 1991) and the quiz-questionnaire (Sangster, 2002) that are conducted with pairs of subjects, the word guessing game allows for phonetic specification of tokens as well as some manipulation of information structure. The game was in a similar style to party games such as *Taboo* or *Pictionary*.

A pair of subjects were presented a scenario prompt through an incomplete stimulus sentence in written Chinese on the laptop screen. As demonstrated in Figure 5.1, the blank representing the missing keyword serves as an implicit in situ

Figure 5.1: An example prompt used in the word guessing game. The scenario is "*the (blank to be filled) they built is very stable*". The three keywords are *snowman*, *castle*, and *stairs*.

Speaker ID	Role	Transcript
214	Clue-giver	法国的国王都会住在一个地方。 'The kings of France all lived in a place.'
213	Clue-receiver	他们**堆的**宫殿很稳当。(1a) 'The palace they built is very stable.'
214	Clue-giver	差不多。 'Getting close.'
213	Clue-receiver	他们**堆的**皇宫很稳当。 'The royal palace they built is very stable.'
213	Clue-receiver	他们**堆的**城堡很稳当。(1b) 'The castle they built is very stable.'
214	Clue-giver	对。 'Yes.'

Table 5.1: An excerpt of the transcript of the conversation between speakers 213 and 214 during the word-guessing game. The Clue-receiver (213) had three attempts, highlighted in light grey, with proposed keyword noun phrase underlined. The actual targeted phrases *duī de* in this scenario are in bold.

WH-question. One subject, the Clue-giver, was given three correct keywords in each scenario and tried to have their partner guess all the given keywords through sketches or verbal hints without using the morphemes that comprise these keywords. The other subject, the Clue-receiver, attempted to reconstruct the scenarios by filling in potential keywords to complete the sentence, having received clues from their partner. A transcript excerpt of a conversation between two participants is displayed in Table 5.1. While the missing keywords attracted their attentions, the actual speech sequences in which the study was interested in were the target phrases such as *duī de* 堆的 marked in bold in this excerpt. As the guessing attempt increased, the targeted speech sequence was repeated. In this excerpt, the targeted phrase *duī de* was repeated three times. Having exhausted possible keywords or correctly guessed all the keywords, the Clue-giver would confirm the correct answers by reading out the complete sentences with the correct keywords before moving on to the next scenario. The Clue-receivers had maximum 5 attempts in each scenario, considering the time limit of a session. The pair spoke in an uninhibited fashion during the game. There are 16 scenario prompts and $16 \times 3 = 48$ keywords in total.

Whether or not the Clue-receiver successfully guessed all the keywords does not matter in this study and their performance was not assessed, although a score was given to each pair of subjects based on the correct number of keywords guessed and the number of attempts when I conducted the fieldwork visits in the school. An intended interpretation of the score was an indicator of the pair's teamwork or connection. The inclusion of the gamified scoring system was to motivate and stimulate the pair to cooperate and to boost their engagement and participation. The pair with the highest score was awarded with a box of chocolate at the end of my visit.

One key advantage of this game is that the subjects generally find it entertaining and the relaxed and friendly atmosphere sparks naturalistic conversation. It was held that excited and emotionally charged speech might override the constraints of a recording setting and the observer effect (Labov, 1964). Such a technique on the one hand allowed us to control the tokens of requisite phonetic features that yielded meaningful comparison, on the other hand directed subjects' attention away from their own speech and maximally kept them engaged. One slight drawback is that participants can be over excited about the game and occasionally burst into laughter.

5.2.1.2 Manipulating the linguistic contexts

Tonal coarticulatory effects have been examined by manipulating the neighbouring tones of neutral tones, such as Y. Chen and Xu (2006), Q. Li and Chen (2019), and Sun and Shih (2021). In the previous research, f_0 realisation of a neutral tone is heavily affected by the preceding tone, whereas some anticipatory effect from the following tone has also been reported such as the assimilatory effect when the following tone is a high falling Tone 4 and dissimilatory effect when the following tone is a low Tone 3 in Standard Mandarin.

This chapter investigates the acoustics of neutral tone in two Mandarin varieties by focusing on three types of neutral tone constructions:

A) Disyllabic phrases with the neutral tone syllable *de*, notated as [X-*de*];
B) Trisyllabic phrases with disyllabic reduplicatives followed by the neutral tone syllable *men*, notated as [X_1X_2-*men*];
C) Quadrisyllabic phrases with the aforementioned trisyllabic phrases followed by the neutral tone syllable *de*, notated as [X_1X_2-*men de*].

In all above phrases, the preceding tone of a neutral tone X or X1, was manipulated to be one of the four lexical tones in each Mandarin variety ($X \in [T1, T2, T3, T4]$), whereas the following tone after *de* in sets A) and C) varied randomly. Efforts were made to have the same segmental composition for the X syllable and at least the voicing part of the X syllable was controlled. This allows for an examination of the influence of the preceding tone category on the phonetic realisation of a neutral tone syllable and to compare the findings with the corpus analysis in Chapter 4. In set B), all trisyllabic phrases [X_1X_2-*men*] preceded a Tone 4 syllable *zhèng* with a high starting pitch in both Mandarin varieties. B) and C) can thus be grouped as [X_1X_2-*men*X_4] phrases, in which X_4 is either *zheng* or *de*. This enables us to examine the impact of the following syllable on the f_0 contour of a neutral tone syllable. The inclusion of consecutive neutral tone sequence is conducive to assessing the magnitude of the tonal coarticulatory effects.

These constructions also provide insights on another two aspects of neutral tone production: On the one hand, the two sets of [X-*de*] phrases with different voicing parts explore the potential effects of voicing segments on f_0 realisation of neutral tone. On the other hand, the [X-*de*] phrases can be compared to [X_1X_2] phrases to see whether 'toneless' *de* is different from neutral tone syllable X_2 in reduplication.

In addition, focus may affect f_0 contours such as the post-focus compression effect in Mandarin (Y. Xu, 1999). In this study, most targeted phrases were positioned before the expected narrow focus, since pre-focus f_0 is known to be similar to the neutral focus f_0 deprived of compression effect (e.g. Y. Chen and Gussenhoven, 2008). This is demonstrated in examples of targeted sentences (11) in their speech, taken from the conversation excerpt displayed in Table 5.1, where Example (11a) is the first attempt from speaker 213, and (11b) is the third attempt and the correct answer. The constituents of the proposed answers (underlined) are usually the new information given the prompt and naturally in completive focus position (Drubig, 2003), as in (11a). If a wrong word was introduced and rejected, the updated word might also be in contrastive focus, as in (11b).

(11)	a.	*tāmen*	*duī*	*de*	*gōngdiàn*	*hěn*	*wěndang.*
		3PL	built	DE	palace	very	stable

'The palace they built is very stable.' *Completive focus*

	b.	*tāmen*	*duī*	*de*	*chéngbǎo*	*hěn*	*wěndang.*
		3PL	built	DE	castle	very	stable

'The castle they built is very stable.' *Contrastive focus*

The specific targeted phrases are further introduced and displayed in §5.2.2 Materials. A full list of the 48 sentences with keywords is presented in Appendix B.1.

5.2.2 Materials

The target sentences in the word guessing game were designed to be rich in neutral tones. In the following examples in (12), the neutral tone syllables in Standard Mandarin are marked in bold.

(12) a. *tā**men** duī **de** chéngbǎo hěn wěn**dang**.*
3PL built DE castle very stable
'The castle they built is very stable.'

b. *mèi**mei-men** zhèng tán-**zhe** jítā.*
sister-PL PROG play-DUR guitar
'Sisters are playing guitars.'

c. *yé**ye-men** **de** tù**zi** táo-**le.***
grandpa-PL GEN rabbit-NOM escape-PERF
'Grandpas' rabbits escaped.'

In Example 12a, there are three disyllabic sequences that contain a neutral tone syllable, while in 12c, five out of eight syllables are neutral tone syllables and the initial phrase *yéye men de* with the clitic *de* includes three consecutive neutral tone syllables.

5.2.2.1 X-*de* set

One of the neutral tone syllables this study focuses on is the toneless clitic particle *de* 的. It is the most frequent word in Mandarin Chinese, ranked top one with regard to both its frequency and dispersion in different registers in the 50-million-word corpus of R. Xiao (2009, p. 20). The particle *de* marks modification in pre-nominal strings in various contexts and has several roles including marking possession, adjectives, prepositions, and nominalisation (C. N. Li & Thompson, 1981). Given the diverse properties and distributions, the categorical status of *de* has been inconsistent in the literature and it is glossed in various forms including DE, POSS/GEN, and so on. The particle *de* is sometimes analysed as a determiner (Simpson, 2001), a head-initial complementiser (Y. A. Lin, 2010), or a subordinator (Paul, 2015) amongst many others. The *de*-marked modifiers as shown in our data can be uniformly analysed as full-fledged relative clauses (Sproat and Shih, 1987; Y. A. Lin, 2010).

Two sets of phonetically controlled disyllabic sequences in the clitic *de* construction in the form of [X-de], shown in Table 5.2, were embedded in the target

	T1	T2	T3	T4	Total
/ueɪ/	[tueɪ tə] *duī-de* 堆的 built	[xueɪ tə] *huí-de* 回的 returned	[tʰueɪ tə] *tuǐ-de* 腿的 leg's	[tueɪ tə] *duì-de* 兑的 redeemed	
Plastic Mandarin	87	73	60	89	309
Standard Mandarin	56	49	40	64	209
/a/	[ta tə] *dā-de* 搭的 built	[ta tə] *dá-de* 答的 answered	[ta tə] *dǎ-de* 打的 made	[ta tə] *dà-de* 大的 large	
Plastic Mandarin	68	78	69	82	297
Standard Mandarin	54	50	46	56	206
Total (PM)	155	151	129	171	606
Total (SM)	110	99	86	120	415

Table 5.2: Disyllabic speech materials containing the neutral tone syllable *de*. The number of these disyllabic tokens in Plastic Mandarin and Standard Mandarin in the data is shown.

sentences (e.g. 12a). Each set features a consistent vowel in the X syllables[1]. Efforts were made to find X syllables with the same onset consonants and within the range of high frequency words. High frequency words account for a significantly large portion of coverage in both written and spoken texts (Tříšková, 2017), and it is easier to create speaker-friendly stimulus sentences with them. Given the tonotactic gaps, the onset consonants of X syllables in the /ueɪ/ set differ slightly. All X syllables, except for *duì* 兑 'redeemed', are within top 1400 most commonly used characters (R. Xiao, 2009). All [X-*de*] sequences but *tuǐ* 腿的 'leg's' appeared right before the keywords and hence in the pre-focus position [2], as in Example (11).

[1]The IPA transcription of Mandarin Chinese syllables varies between studies, depending on the phonological analyses. The triphthong in the upper set in Table 5.2 can be transcribed as a glide followed by a diphthong. For example, [tweɪ tə] for *duī de*. According to Duanmu (2007), it can also be [tʷeɪ tə]. The triphthong is conveniently adopted here because I will consider the vocalic parts [ueɪ] and [a] as instances of the categorical factor *Vowel* in the statistical analysis in this chapter. Without going into a fuller discussion of phonological structure, which is not strongly relevant to the analyses carried out here, [ueɪ] and [a] are just labels for two distinct types of voicing portion with f_0 after the alveolar stop consonant.

[2]It was difficult to create multiple suitable scenarios with *tuǐde* in a similar pre-focus position, so *tuǐde* was placed after a pause following a focused topic phrase at the beginning of a sentence. See Appendix B.1 for full materials.

Disyllabic phrases containing the neutral tone syllable *de* in the target sentences were mined from the recordings, using a python script and the time-aligned transcripts[3]. There are 606 Plastic Mandarin tokens and 415 Standard Mandarin tokens. Table 5.2 presents a breakdown of the number of speech tokens by categories of lexical tones and syllables in the datasets of Plastic Mandarin and Standard Mandarin respectively.

5.2.2.2 X_1X_2-*men*X_4 set

Table 5.3 presents polysyllabic phrases with consecutive neutral tone syllables in the target sentences of the word guessing game (e.g. 12b and 12c). The upper set features kinship terms with two consecutive neutral tone syllables. Many of the colloquial kinship terms X_1X_2 in Mandarin Chinese are reduplicative words in the diminutive form. The second syllable X_2 is considered as a duplicate of the first syllable X_1 and it carries neutral tone. A kinship term can take on another neutral tone syllable, *men*, the plural suffix, which makes it a common phrase with consecutive neutral tones. These kinship terms were chosen on the basis of the following criteria: first, they are high frequency words; second, they are fully voiced with all sonorant segments; third, a kinship term was selected under each tonal context. All the trisyllabic plural kinship terms are followed by the Tone 4 particle *zhèng* in the sentences (see 12b), indicating progressive aspect. Some analysed *zhèng* as an adverb 'just' (Paul, 2015). The particle *zhèng* does not usually cliticise to the left. The bottom set increases the number of neutral tone syllables from two to three by adding the particle *de* 的 to the right of the same trisyllabic phrases, denoting possessive constructions. All these four-syllable sequences appear at the beginning of an eight-syllable sentence, in pre-focus position.

These polysyllabic phrases were mined from the recordings. There are 547 Plastic Mandarin tokens and 372 Standard Mandarin tokens. Table 5.3 presents a breakdown of the number of speech tokens by categories of lexical tones and syllables in the datasets of Plastic Mandarin and Standard Mandarin respectively.

[3]This procedure was documented and openly-available in my online tutorial https://chenzixu.rbind.io/resources/2speechcorpus/

	T1-N+	T2-N+	T3-N+	T4-N+	Total
	[mama mən ʈʂəŋ] *māma-men zhèng* 妈妈们正 mums PROG	[jɛjɛ mən ʈʂəŋ] *yéye-men zhèng* 爷爷们正 grandpas PROG	[naɪnaɪ mən ʈʂəŋ] *nǎinai-men zhèng* 奶奶们正 grannies PROG	[meɪmeɪ mən ʈʂəŋ] *mèimei-men zhèng* 妹妹们正 sisters PROG	
PM	64	71	70	60	265
SM	49	49	44	47	189
	[mama mən tə] *māma-men de* 妈妈们的 mums'	[jɛjɛ mən tə] *yéye-men de* 爷爷们的 grandpas'	[naɪnaɪ mən tə] *nǎinai-men de* 奶奶们的 grannies'	[meɪmeɪ mən tə] *mèimei-men de* 妹妹们的 sisters'	
PM	67	63	77	75	282
SM	43	47	42	51	183
PM	131	134 (133)	147 (143)	135 (134)	547 (541)
SM	92 (91)	96	86 (85)	98	372 (370)

Table 5.3: Polysyllabic speech materials containing consecutive neutral tone syllables. The number of these polysyllabic tokens in Plastic Mandarin and Standard Mandarin in the data is shown. Pinyin transcriptions are based on Standard Mandarin. The numbers in brackets indicate the numbers of tokens used for pitch analysis.

A few tokens without reliable f_0 measurement due to non-modal phonation were removed for pitch analysis.

5.2.3 Acoustic processing

Figure 5.2 illustrates the process of working with speech data collected from the word guessing game in four stages, from the input of long stereo recordings to the output of acoustic measurements of the target speech tokens.

5.2.3.1 Transcription and annotation

All audio recordings were initially two-channel WAV files (44.1 kHz, 16-bit, signed integers) featuring pairs of subjects. A necessary first step towards annotating metadata and extracting tokens for analysis was to extract each channel of an audio file, identify the main speaker in the channels, and save them separately as mono audio files. The speaker IDs introduced in §3.2.4 were used in the annotations to maintain the anonymity of the speakers. As demonstrated in stage (0) in Figure

5.2, the two channels contained speech signals from two microphones, each of which attached to speakers 201 and 202 respectively. Although the same conversation was recorded, the two channels are different. When subject 201 was speaking, the amplitude of her portions of speech in the corresponding (upper) channel is much larger than that in the other (bottom) channel. Speech data of subject 201 were, therefore, based on the upper channel, while speech data of subject 202 on the bottom channel.

The target sentences in each mono audio file were further identified, annotated, and extracted. This is demonstrated in stage ① in Figure 5.2. By listening to the audio and visually examining the acoustic waveform in Praat (Boersma & Weenink, 2017), supplemented by spectrographic analysis where appropriate, the boundaries of the interval of a sentence were manually determined, and placed at a zero-crossing point of the signal. A label (i.e. `b10_3_201a`) that encoded contextual information about the utterance such as the task code (i.e. a/b/c/d), the scenario prompt index (i.e. 1/2/.../16), the set index (i.e. 1/2/3), and the speaker ID (e.g. 201) and role (i.e. q/a), was inserted in the TextGrid immediately after a sentence interval was identified, using a Praat script. Each target sentence interval was extracted and saved as a WAV audio file automatically, with the label being the filename. Such filenames can be easily collapsed into contextual information in a tabular format.

Each audio file such as '*b10_3_201a.wav*' was transcribed in written Chinese, with a space inserted between each two characters. In doing so, the orthographic transcription can be time-aligned with the audio at both phone and syllable level, given that each Chinese character corresponds to a single syllable. Each transcript was saved as a TXT file under the same filename of the corresponding WAV file (e.g. '*b10_3_201a.txt*').

5.2.3.2 Forced alignment

The audio recordings and their transcriptions were then processed by the Penn Phonetics Lab Forced Aligner (P2FA) for Mandarin Chinese (Yuan and Liberman, 2008; 2015) to achieve automatic phonetic segmentation. The audio files were

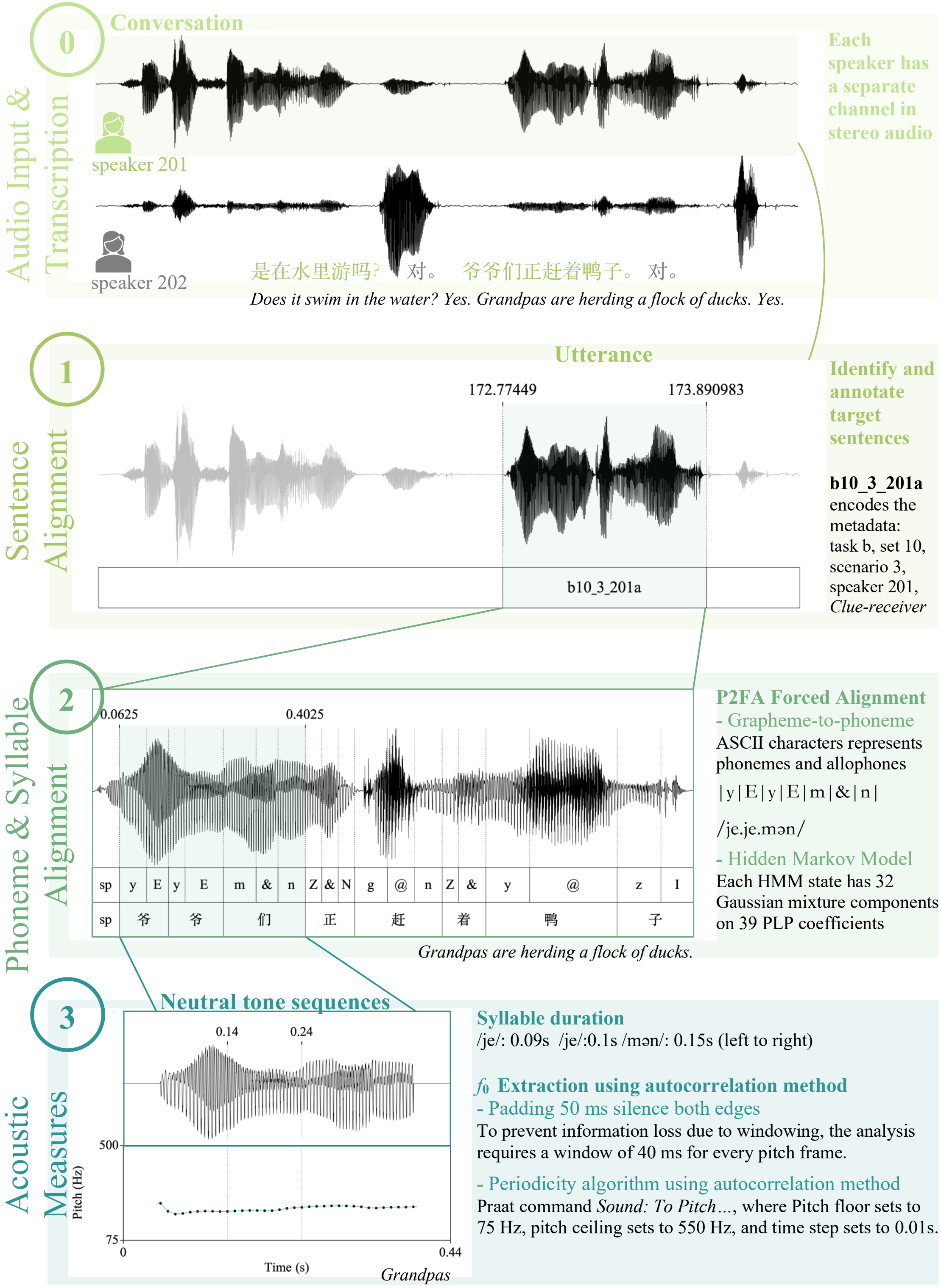


Figure 5.2: An illustration of audio processing steps, from raw stereo audio to a zero-padded speech token *yéyemen* 'grandpas'.

downsampled to 16 kHz to prepare for the use of P2FA. The output of the aligner was a TextGrid file for each audio file, in which the phones and syllables were aligned to the audio, as shown in the first and second tiers of the TextGrid respectively in stage (2) in Figure 5.2. The syllables were represented by Chinese characters. This study is chiefly interested in the syllable tier, from which we can identify where in the audio a syllable begins and ends.

Alignment accuracy is a more complicated issue in connected speech than monosyllables in isolation. Some phoneme transitions are gradual and less consistent. Boundaries between vowels, diphthongs, and approximants can be highly ambiguous (Baghai-Ravary et al., 2009), such as the transition from /aʊ/ to /l/ and from /l/ to /ə/ in *lǎo-le* 'getting old'. Voice qualities and emotional states in spontaneous natural speech may influence the aligner performance. Phonetic reduction, pervasive in conversational speech (Johnson, 2014), has been another persistent challenge in forced alignment (Yuan & Liberman, 2015).

Figures 5.3 and 5.4 demonstrate a major type of misalignment in the sentences in our corpus, which was resulted from phonetic reduction. In the figures, the two highlighted intervals, the neutral tone syllables [tə] (labelled as `d&` in the first tier of the TextGrid) and [ʈʂə] (labelled as `Z&`), correspond to mostly the frication portions in the waveform and spectrogram. By listening to the audio, these syllables were greatly reduced and their onset consonants, unaspirated alveolar plosive [t] and unaspirated retroflex affricate [ʈʂ], were barely audible, suggesting that they were coarticulated[4]. The [t] and [ʈʂ] seemed coalesced with the preceding [n], that is, the nasality extended through the closure period, and the remaining short vowel [ə] attached to the previous syllable, forming [mə̃ň̥ə] and [kʰãňə], and was included in the previous intervals of 们 and 看 respectively in the figures. The highlighted intervals are in fact part of the next stressed and aspirated syllables: it is the burst and partial aspiration of the consonant onset of [tʰu] 兔 in Figure 5.3, and the burst, frication, and aspiration of the consonant onset of [ʈʂʰwɑŋ] 窗 in Figure 5.4. Another type of misalignment included extraneous pauses inserted by P2FA, in

[4] The audio files are available in the online supplementary materials.

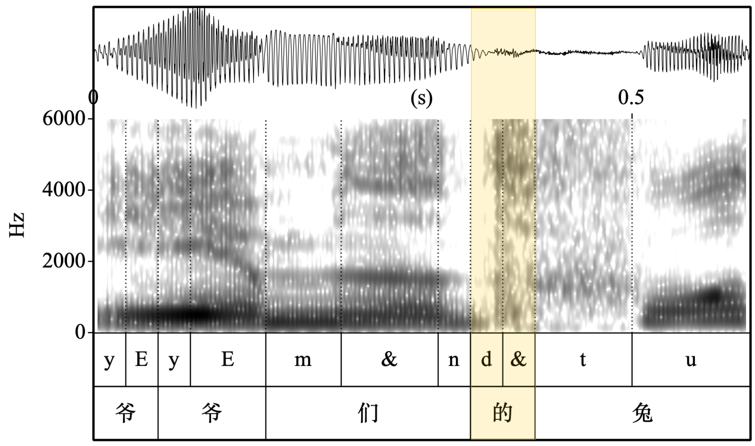


Figure 5.3: Misplaced boundaries of the syllable 的 *de* GEN in ‘grandpas’ rabbit’ by P2FA. The interval is highlighted in yellow background.

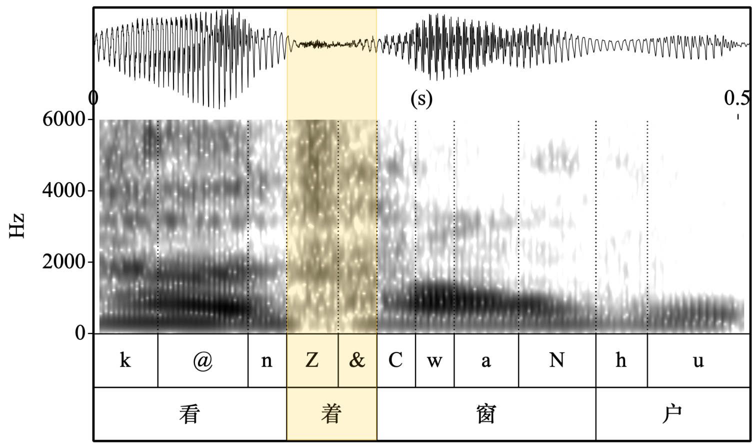


Figure 5.4: Contracted interval of the syllable 着 *zhe* DUR in ‘looking at windows’ by P2FA. The interval is highlighted in yellow background.

places such as the closure period of stops, the frication portion of a velar fricative, or the initiation and termination part of the speech.

Despite these small problems in some TextGrids, P2FA has provided a solid first pass segmentation. It was subsequently verified and manually corrected with a written record, following the practice in many studies (e.g. Zue et al., 1990; Yuan, 2012). Less than 5% of the sentences were noted as ‘bad or wrong alignment'. Utterances with overlapping speech or laughter were not considered in further analysis.

5.2.3.3 Acoustic measurement

The duration of syllables was easily accessible from the forced alignment. Sound intervals can be extracted too given the temporal information. A search script was first created to automatically retrieve the time stamps of phrases containing neutral tones in the audio files from the corresponding TextGrid files. Then a sound trimming script was used to extract the target sound intervals from the recordings, and each output WAV file was assigned a unique filename rich in contextual information. For instance, the trisyllabic token /je je mən/ in stage ③ in Figure 5.2 has the filename '*b10_3_201a爷爷们正s0.2525e0.4025.wav*', where `b10_3_201a` indicates the source file, 爷爷们正 is the transcript of this X_1X_2-*men* phrase plus the following syllable X_4, and `s0.2525` and `e0.4025` specify the starting and ending times of the syllable *men* in the source audio *b10_3_201a.wav*.

As described in §3.3.3.2 and shown in Figure 5.2, 50 ms of silence was further added to both edges of each sound interval. These zero-padded speech tokens constitute the audio data for the subsequent analysis.

f_0 was measured every 10 millisecond where the speech was voiced using Praat's autocorrelation algorithm (Boersma, 1993) through a bespoke Python script. The parameter setting of the pitch tracker were the same as that in §3.3.4. The script also detects f_0 jumps between adjacent voicing periods, flags potential tracking errors, and automatically removes edge f_0 estimates that were unusual, that is, ones that lay much farther away from the token mean f_0. The script calculates the running sample-to-sample difference for each ordered time-series of f_0 measurements and compares each difference to a threshold. It used a maximum f_0 movement threshold of 3 st between adjacent f_0 points to filter out extreme boundary f_0 estimates. This was not a strict threshold according to the findings of previous speech production experiments based on producing oscillating glissando between high and low f_0. Y. Xu and Sun (2002) assessed the maximum speed of pitch change and the fastest excursion speed in their data was 96 st/s (0.96 st/10 ms) in the direction of falling by a female subject. Sundberg (1979) reported the speed of f_0 change in the central 6/8 of an f_0 movement and the average pitch change rate of the fastest group, the female singers, in this study was slightly below 2 st/10 ms. Thus 3 st difference in 10 ms seemed very unlikely and erroneous. Given that the pitch change rate of the fastest individual surpassed 2 st/10 ms, the use of 3 st was adopted. A slightly less strict threshold allowed for flagging and close examination of more tolerated large f_0 changes and less automatic removal of f_0 estimates.

The audio files flagged as containing a potential error were subsequently manually audited, through visual and auditory inspection. They were corrected where applicable, or otherwise removed. A f_0 movement of about 12 st, or an octave, indicated pitch doubling or halving. The method of editing the Pitch object in the Praat PitchEditor introduced in §3.3.4.2 was used to obtain the desired f_0 in such cases. The output of this procedure was an ordered time-series of reliable f_0 measurements for all target speech tokens.

5.2.4 Statistical technique

In previous chapters, Legendre polynomial models effectively summarised the f_0 contour of monosyllables and neutral tone syllables in three interpretable coefficients. Longer phrases, however, have more complex f_0 contours than monosyllables. Polynomial modelling also has its limitations. The greater complexity in a contour may lead to an increase in the minimal number of coefficients required to explain a certain percentage of the variation in the curves. The higher ranking the basis functions, the more intricate and faster varying the patterns, and we might be less intuitive about the interpretation of the high order coefficients. Polynomials tend not to be optimal with long asymptotes and plateaus, nor sensitive to curves that have too many wiggles (Winter & Wieling, 2016). Generalised Additive Mixed Models (GAMMs) under the non-parametric framework were hence motivated and applied to the f_0 contours in this chapter, given its flexibility and adequacy (S. N. Wood, 2006).

In dynamic f_0 time-series f_0 at time t may not be independent from that at time $t - k, k \geq 1$ (Baayen et al., 2017). GAMMs can effectively deal with data with autocorrelation structure (Baayen et al., 2016) and allow simultaneous estimation of temporal non-independence and calibrated inferences in the presence of hierarchical structure. GAMMs represent contours and their differences in both parametric and smooth terms and detect regions where the contours diverge by resorting to the confidence intervals around the estimated difference curve. While parametric terms capture f_0 height-related patterns that hold across the entire contour, smooth terms model patterns related to the shape of the f_0 contour (Sóskuthy, 2021). In our case, the dependent variable is the normalised f_0, modelled as a smooth function over the normalised time, using a sum of weighted special non-linear basis functions across the whole time span. The basis functions used in this study were thin plate regression splines (S. N. Wood, 2003). This process is indeed conceptually similar to polynomial regression in Chapters 3 and 4.

5.2.4.1 Data normalisation

Following the same approach as in Chapter 3, the f_0 measured in Hz of a speaker in a language variety was normalised to semitones above or below the speaker's mean f_0 of that variety. Please see §3.4.2.1 for details. The speaker mean f_0 was consistent with that in Chapter 3 to enable a comparison between tones in connected speech and citation tones.

For time normalisation, a non-linear method was used to allow comparison of individual syllables within a token as well as between tokens: the time measures of syllables in the same position of the same type of phrase was linearly scaled to the same domain. Take disyllabic phrases, for instance—the average duration was computed for the first and second syllables respectively; then the ratio of the average duration between the first and second syllable was maintained and mapped to the normalised time domain of $[-1, 1]$. Durational information was utilised so that the overall shape of a phrasal f_0 track was not greatly distorted. The details about the time normalisation of a specific type of phrase will be presented in §5.4.1.1 following the duration analysis.

5.2.4.2 GAMMs

In LME models, a univariate response or dependent variable is modelled as a sum of a linear predictor $X\beta$, random noise Zb, and a Gaussian error with zero mean ϵ (for review, see §3.4.4). The linear predictor can be a set of independent variables such as $\beta_0 + \beta_1 x_1 + \beta_2 x_2$, and the response variable is often assumed to have a normal distribution. The Generalised Linear Mixed Model (GLMM) extends LME by allowing the response to depend on a smooth monotonic function of the linear predictor and to follow other distributions from the exponential family, such as Poisson, binomial, and gamma, apart from the Gaussian distribution (Baayen et al., 2017).

The Generalised Additive Mixed Model enriches the linear predictor of GLMM so that it is specified as a sum of smooth functions of one or more independent variables such as $\beta_0 + f_1(x_1) + f_2(x_2)$. The smooth function $f(x)$ is a weighted sum of a set of q basis functions B_k defined over the independent variable x (S. N. Wood, 2006):

$$f(x) = \sum_{k=1}^{q} B_k(x)\beta_k \tag{5.1}$$

The relaxation of the linearity assumption in GAMMs enables us to capture the nonlinear patterns that are ubiquitous in phonetic data including f_0 contours with temporal autocorrelational structure (Chuang et al., 2021). The specific GLMM models will be presented in §5.4.1.2 prior to the findings.

5.3 Duration of Neutral Tones in Mandarin Accents

Syllable duration is often associated with prominence, boundary, or language redundancy (Aylett & Turk, 2006). The analysis of timings of different syllables in various linguistic units and their sources of variability is conducive to our understanding of the rhythm of a language. This section answers the research questions *How long is a neutral tone in various contexts in Plastic Mandarin and*

Standard Mandarin? and *How does the duration of neutral tone compare to the preceding lexical tone in the two Mandarin varieties?*

Two types of duration variables were used in the analysis: one was the absolute duration of each syllable measured in seconds, obtained through the forced alignment procedure; the other was the duration ratio of a neutral tone syllable to its preceding lexical tone syllable. For the [X-*de*] set, the duration ratio was the second syllable *de* (neutral tone) to the first syllable (lexical tone) in a disyllabic utterance. For the [X_1X_2-$menX_4$] set, the duration ratios of the reduplicated syllable X_2 and *men* to the initial lexical tone syllable X_1 were included, and notated as $Duration\ Ratio_{X_2/X_1}$ and $Duration\ Ratio_{men/X_1}$.

Based on the duration findings in Chapters 3 and 4, the following hypotheses are proposed about duration:

$H_1$1 Neutral tone syllable *de* are in general shorter than lexical tone syllables.

In §3.6.2.4, *de* in citation was the shortest among the other syllables in both Mandarin varieties, and in the corpus data in §4.3, Standard Mandarin *de* is generally shorter than the preceding lexical tone syllables.

$H_1$2 In stressed positions, lexical tone category has influence in syllable duration.

From the duration data in §3.5, there are intrinsic duration differences among different lexical tones.

$H_1$3 The duplicate syllable X_2 is generally shorter than the preceding syllable X_1.

Following the discussion of the corpus data in §4.3, unstressed neutral tone provides a context for shorter vowels in CV syllables.

When comparing two Mandarin varieties, the null hypothesis ($H_0$4) that there is no difference in the ratio of neutral tones to their preceding lexical tones between the two varieties was adopted.

5.3.1 Statistical method

The general procedure for the duration analysis comprises two steps: first, descriptive statistics and data visualisation constitute exploratory analysis; second, we conduct LME models analysis based on our observations in the first step in order to inspect whether any observed differences are statistically significant. I now describe the LME models in detail.

5.3.1.1 X-de set

Three groups of linear mixed effects (LME) models were initially attempted in the analysis with datasets in various scopes, given the unequal sample sizes in different categories and repeated samples from the same individuals. The inclusion of random-effect factors resolved the non-independencies brought up by repeated tokens of the same phrases from the same speakers.

1) $Duration \sim Tone + (1 + Tone|Speaker) + (1 + Tone|Item)$

The first group of LME models examined the dataset of each Mandarin variety separately and focused on the duration difference between lexical tone syllables and neutral tone syllables. As shown in the model formula, absolute duration was the dependent variable, the fixed-effect factor was *Tone* (2 levels: lexical tone and neutral tone), and two random-effect factors *Speaker* and *Item* (eight different phrases) were included with intercept and slope design. The random slopes were dropped in the Plastic Mandarin dataset due to convergence issues.

2) $Duration\ Ratio \sim Preceding\ Tone * Vowel + (1|Speaker)$

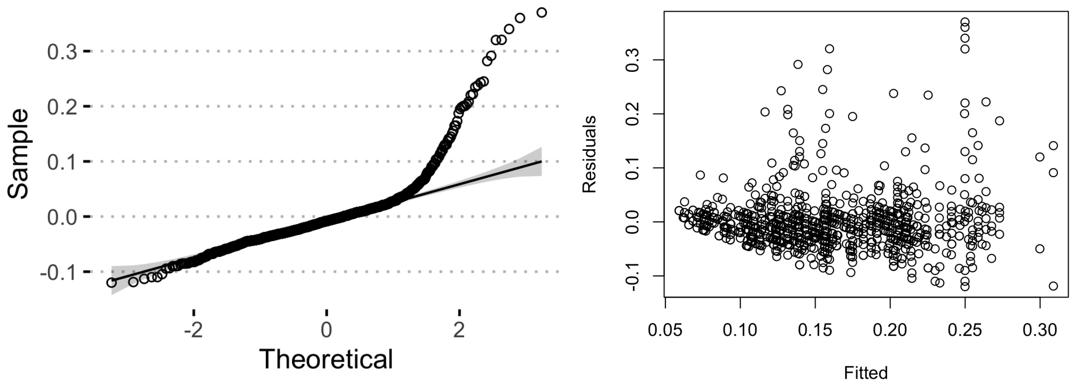
The second group of LME models further explores the effects of preceding tone and vowel in the duration ratio of a [X-*de*] disyllabic utterance in the dataset of each Mandarin variety. Duration ratio was the dependent variable, *Preceding Tone* (T1, T2, T3, and T4), *Vowel* ([ueɪ] and [a]), and their interaction term were included as fixed-effect factors. *Speaker* was the random-effect factor with random intercepts included. Random slopes were not considered given the convergence issues of the models.

3a) $Duration \sim Tone{*}Variety{+}(1{+}Tone|Speaker){+}(1{+}Tone{+}Variety|Item)$

3b) $Duration\ Ratio \sim Variety + (1|Speaker) + (1 + Variety|Item)$

The third group of LME models investigated the difference between the two Mandarin varieties. The datasets were combined and hence a new variable was introduced – language variety, represented as *Variety* in the above formulae 3a) and 3b). Duration and Duration Ratio were used as dependent variables respectively. In 3a), the fixed-effect factor was *Tone* (lexical tone and neutral tone), *Variety* (Standard Mandarin and Plastic Mandarin), and their interactions. By-*Speaker* intercepts and slopes for the effect of *Tone*, as well as by-*Item* intercept and slopes for the effects of *Tone* and *Variety* were included. In 3b), the fixed-effect factor was *Variety*. Random intercepts of *Speaker* and *Item*, and by-*Item* slopes for the effect of *Variety* were included.

For all the LME models, residual plots were checked by visual inspection. The Normal Q-Q plots such as Figure 5.5a reveal a right-skewed sampled distribution, characterised by sample quantiles at the high end being more positive than the expected normal quantiles. In this case the model was also slightly heteroscedastic, with some larger residuals for higher fitted values (see Figure 5.5b). Violations of assumptions are common in real datasets, and model estimates can be robust to violations of assumptions with careful evaluation (Schielzeth et al., 2020). This study tackled the violations in distributional assumptions through robust estimation methods. The models proposed above were re-fit robust linear mixed-effects models using *robustlmm* package (Koller, 2016), and robustness weights were assigned to observations and groups. Random slopes were dropped in the third group of robust models given the convergence issues. The detected potential "outliers" were examined and discussed. More discussion on model selections and robustness weights can be found in the appendix. The full script and results of the statistical analysis are available in the RMarkdown file *Neutral Tone Duration Analysis* in the online supplementary materials.



(a) Normality of residuals **(b)** Homoskedasticity

Figure 5.5: Two residual plots of the LME model 1) for Standard Mandarin data.

5.3.1.2 $[X_1X_2$-***men***$X_4]$ set

Another three groups of robust linear mixed effects (LME) models were employed in the duration analysis of polysyllabic phrases. Datasets of syllable duration were prepared in various scopes including an individual dataset for each Mandarin variety, a combined dataset of both varieties, and subsets of the first three syllables and $[X_1X_2$-*men de*] phrases for each Mandarin variety.

4) $Duration \sim Position * X_1 + (1|Speaker)$

The first group of LME models examined all $[X_1X_2$-*men*] tokens and $[X_1X_2$-*men de*] tokens of each Mandarin variety separately and tested whether the factor of syllable position could be a predictor of syllable duration. As shown in the model formula, absolute duration was the dependent variable, and the main fixed-effect factor was *Position* (trisyllabic phrases: 1, 2, 3; quadrisyllabic phrases: 1, 2, 3, 4) representing different syllables in a phrase. The four types of phrases represented by X_1 were also included in the fixed effects as a control variable, and the interaction term between the two fixed effects were also considered.

5) $Duration \sim X_1 * X_4 + (1|Speaker)$

The second group of LME models focused on the duration of *men* in different contexts: *men* was followed by either the Tone 4 particle *zheng* or another neutral tone syllable *de*, in addition to the four different preceding syllables. The dependent variable was the absolute duration of *men*, and the independent variables were four preceding syllables or disyllabic reduplicatives, represented by X_1 in the formulae (i.e. *māma*, *yéye*, *nǎinai*, and *mèimei*), two following syllables represented by X_4 (i.e. *zhèng* and *de*), and their interactions.

6a) $Duration\ Ratio_{X_2/X_1} \sim Variety * X_1 + (1|Speaker)$

6b) $Duration\ Ratio_{men/X_1} \sim Variety + X_1 * X_4 + (1|Speaker)$

The third group of LME models investigated the difference between the two Mandarin varieties, using the combined dataset of the two Mandarin varieties. Duration Ratios were used as dependent variables. In 6a), the fixed-effect factor was *Variety* (Plastic Mandarin and Standard Mandarin), the syllable type X_1 (i.e. *māma*, *yéye*, *nǎinai*, and *mèimei*), and their interactions. In 6b), the fixed-effect factor was *Variety*, the syllable type X_1, the following syllable X_4 (i.e. *zhèng* and *de*), and the interaction between $X_1 * X_4$.

For all above models, by-*Speaker* random intercepts were included, but not by-*Speaker* random slopes due to convergence issues. The results are presented in the following sections.

5.3.2 The duration of disyllabic phrases with *de* in two Mandarin varieties

5.3.2.1 Central tendencies and distributions

Figure 5.6 presents the distribution of absolute duration of syllables in the [X-*de*] disyllabic utterances in Standard Mandarin, grouped by initial lexical tone categories. Correspondingly, Table 5.4 supplies the mean and standard deviation of both absolute duration and duration ratio of syllables by initial tone categories. On average, the lexical tone syllable X is 195 ms ($\sigma = 58$ ms), while the neutral tone *de* is 128 ms ($\sigma = 84$ ms), about 67% of the duration of the preceding lexical tone syllable. *de* is in general shorter than the preceding lexical tone regardless the tone category in Standard Mandarin.

From Table 5.4, we can also see that the average duration of lexical tone syllables in the first position of [X-*de*] disyllabic utterances by tone categories does not have the duration hierarchy found in the citation tone data in Chapter

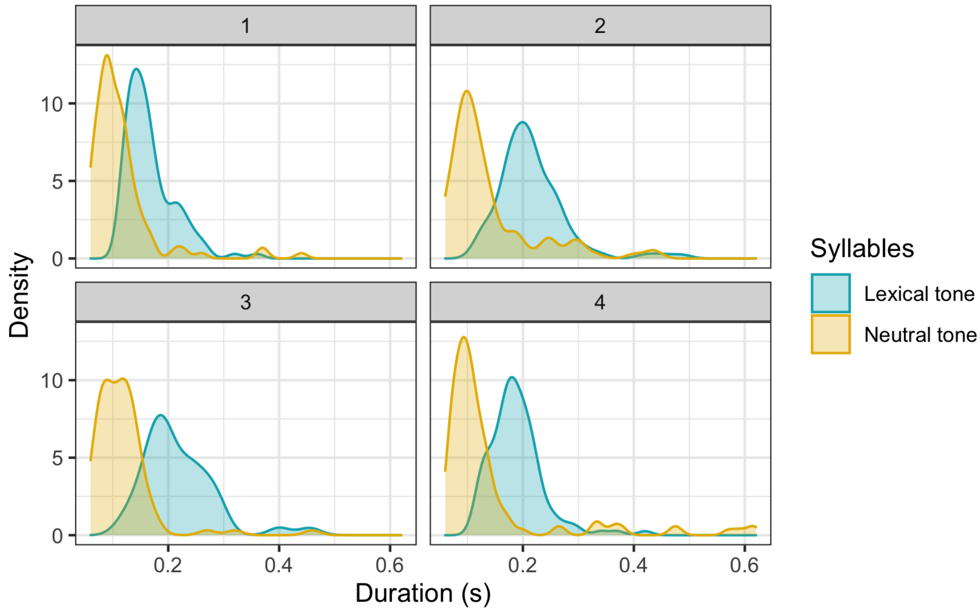


Figure 5.6: Kernel density functions for duration (s) of syllables in [X-*de*] disyllabic utterances in Standard Mandarin by initial tone categories.

Preceding tone X	Lexical tone X $\mu(\sigma)$	Neutral tone *de* $\mu(\sigma)$	Duration ratio $\mu(\sigma)$
Tone 1	0.167 (0.044)	0.114 (0.059)	0.69 (0.32)
Tone 2	0.215 (0.061)	0.138 (0.080)	0.63 (0.27)
Tone 3	0.218 (0.066)	0.117 (0.055)	0.58 (0.32)
Tone 4	0.188 (0.049)	0.142 (0.115)	0.74 (0.51)
All tones	**0.195** (0.058)	**0.128** (0.084)	0.67 (0.38)

Table 5.4: Mean and standard deviation of the absolute duration (s) and duration ratio of the syllables by the preceding tone categories of *de* in the [X-*de*] disyllabic utterances in Standard Mandarin.

3. Here Tone 3 and Tone 2 are similar in duration, about 218 ms (σ = 66 ms) and 215 ms (σ = 61 ms) respectively. The shortest tone among the four tones is Tone 1, about 167 ms (σ = 44 ms).

Figure 5.7 presents the distribution of absolute duration of syllables in the [X-*de*] disyllabic utterances in Plastic Mandarin by initial tone categories. Correspondingly, Table 5.5 supplies the mean and standard deviation of both absolute duration and duration ratio of syllables by initial tone categories. On average, the lexical tone

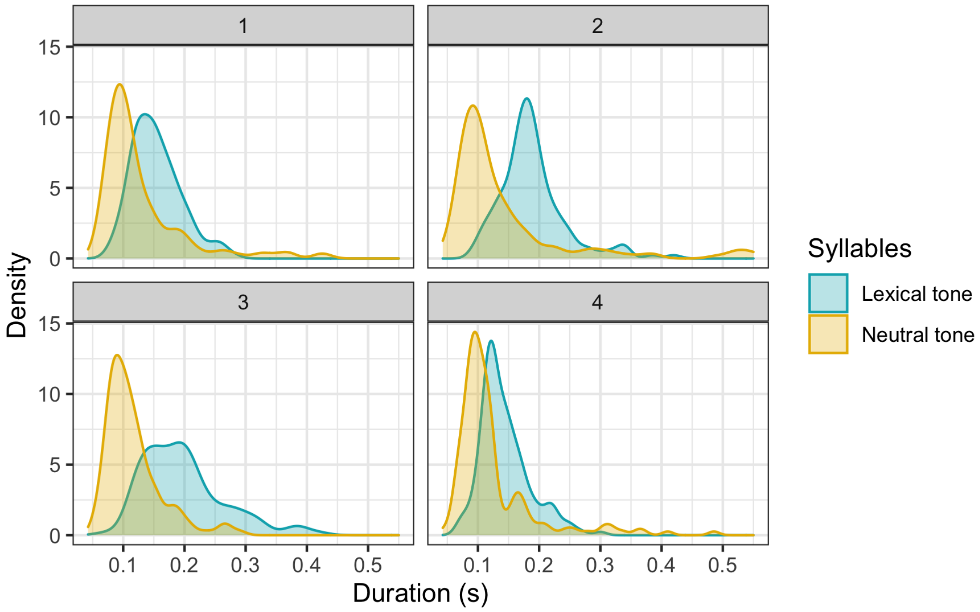


Figure 5.7: Kernel density functions for the duration (s) of syllables in the [X-*de*] disyllabic utterances in Plastic Mandarin by initial tone categories.

syllable X is 169 ms ($\sigma = 56$ ms), while the neutral tone *de* is also 128 ms ($\sigma = 75$ ms), about 77% of the duration of the preceding lexical tone syllable. *de* is in general shorter than the preceding lexical tone regardless the tone category in Plastic Mandarin.

From Table 5.5, we can also see that the average duration of lexical tone syllables in the first position of [X-*de*] disyllabic utterances by tone categories mostly resembles the duration hierarchy found in the citation tone data in Chapter 3, in that the shortest tone is Tone 4, about 144 ms ($\sigma = 41$ ms), followed by Tone 1, about 154 ms ($\sigma = 40$ ms). Similar to Standard Mandarin, Tone 3 and Tone 2 are longer than Tone 1 and Tone 4, about 198 ms ($\sigma = 67$ ms) and 189 ms ($\sigma = 55$ ms) respectively. The duration for citation tones is generally longer than syllables in connected speech in both Standard Mandarin and Plastic Mandarin.

5.3.2.2 Comparisons by tone categories and varieties

The robust linear mixed-effects models suggest that syllable *de* is statistically shorter than a preceding lexical tone syllable ($t(11.16) = -8.02, p < .001$), by about 70.7 ms ±8.8 (standard error) in Standard Mandarin. Similarly, in Plastic

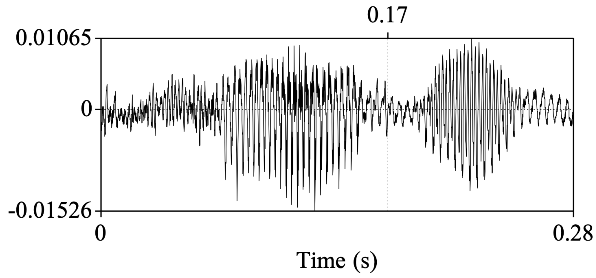
Previous tone	Lexical tone X $\mu(\sigma)$	Neutral tone *de* $\mu(\sigma)$	Duration ratio $\mu(\sigma)$
Tone 1	0.154 (0.040)	0.129 (0.070)	0.83 (0.32)
Tone 2	0.189 (0.055)	0.142 (0.102)	0.74 (0.43)
Tone 3	0.198 (0.067)	0.116 (0.045)	0.64 (0.31)
Tone 4	0.144 (0.041)	0.125 (0.069)	0.86 (0.34)
Overall	**0.169** (0.056)	**0.128** (0.075)	0.77 (0.36)

Table 5.5: Mean and standard deviation of the absolute duration (s) and duration ratio of the syllables by the preceding tone categories of *de* in the [X-*de*] disyllabic utterances in Plastic Mandarin.

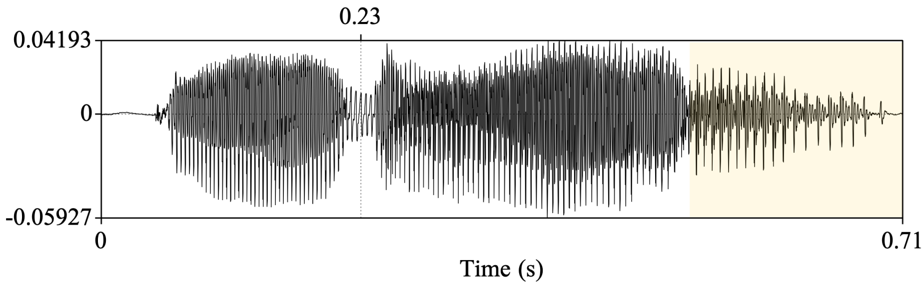
Mandarin syllable *de* is statistically shorter than a preceding lexical tone syllable ($t(6.87) = -22.42, p < .001$), by about 49.5 ms ±2.2 (standard error).

To better understand the concept of robustness weights $w \in [0, 1]$, I scrutinised tokens that had been down-weighted in the models. These tokens indeed have a characteristic departure from the majority. Figure 5.8 illustrates two tokens of the same phrase by the same speaker, which display huge intra-speaker durational differences. The two syllables in token (a) have the maximum robustness weight of 1, which represent the majority, the robust tokens. In (a) the first syllable is 170 ms and the second syllable *de* is 110 ms, 64.7% of the first syllable in duration. The token (b) is rather 'unexpected', especially the second syllable *de* (480 ms), which is much longer than the first syllable (230 ms) as well as other *de* syllables. The *de* syllable in (b) has a low robustness weight of 0.112.

Such a token may be classified as "outliers" by the robust method, it is in fact a token being modelled badly. The large residual may result from variability in the data that was not explained by predictors in the model but might be accounted for by other (unknown) factors. If we put the extracted token (b) back to the sentential context, as shown in Figure 5.9, we may be immediately aware of what factors might have played a role in determining the length of the *de* syllable. The token is followed by a 40 ms pause in the speech, signalled by 'sp' in the TextGrid. The insertion of pause here reinforced the right edge of prosodic boundary of *duì-de*,



(a) $w_{duì} = 1; w_{de} = 1$



(b) $w_{duì} = 0.898; w_{de} = 0.112$

Figure 5.8: Waveform of two tokens of *duì de* [tueɪ tə] 'redeemed' by speaker 202. Creaky portion in (b) is highlighted in light yellow.

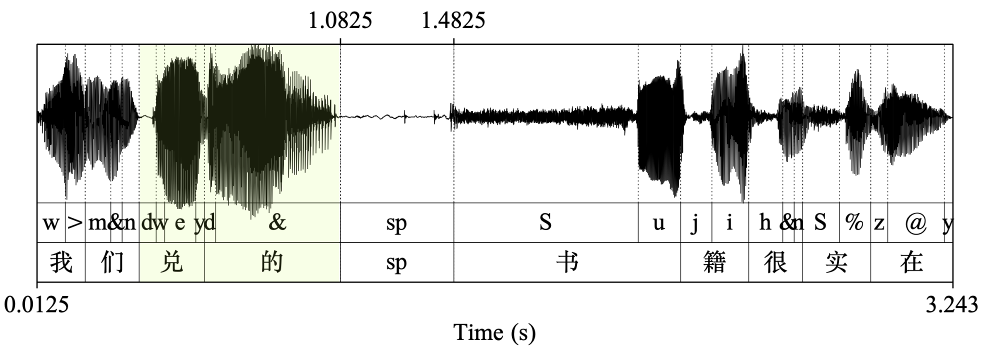
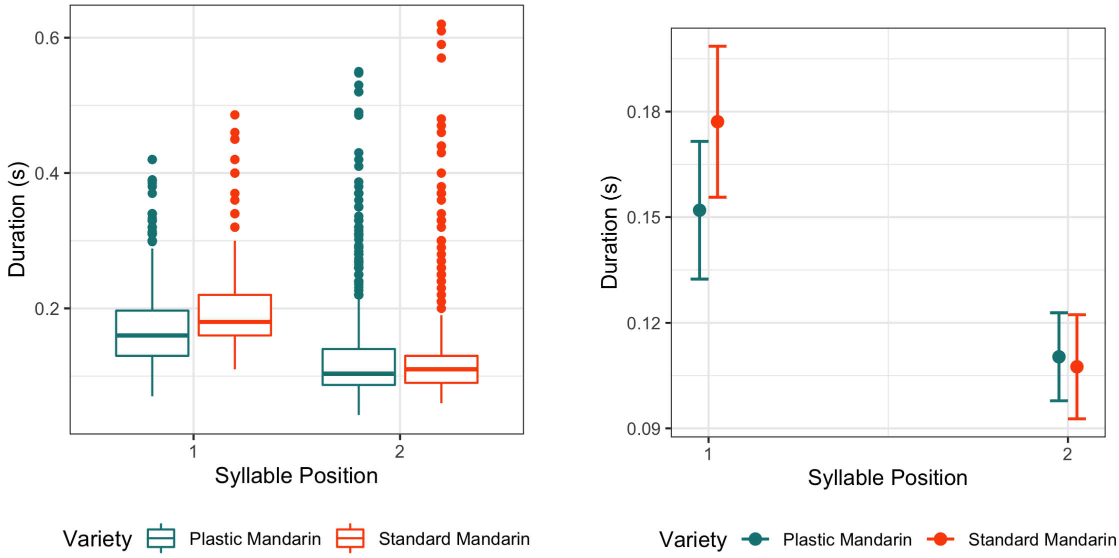


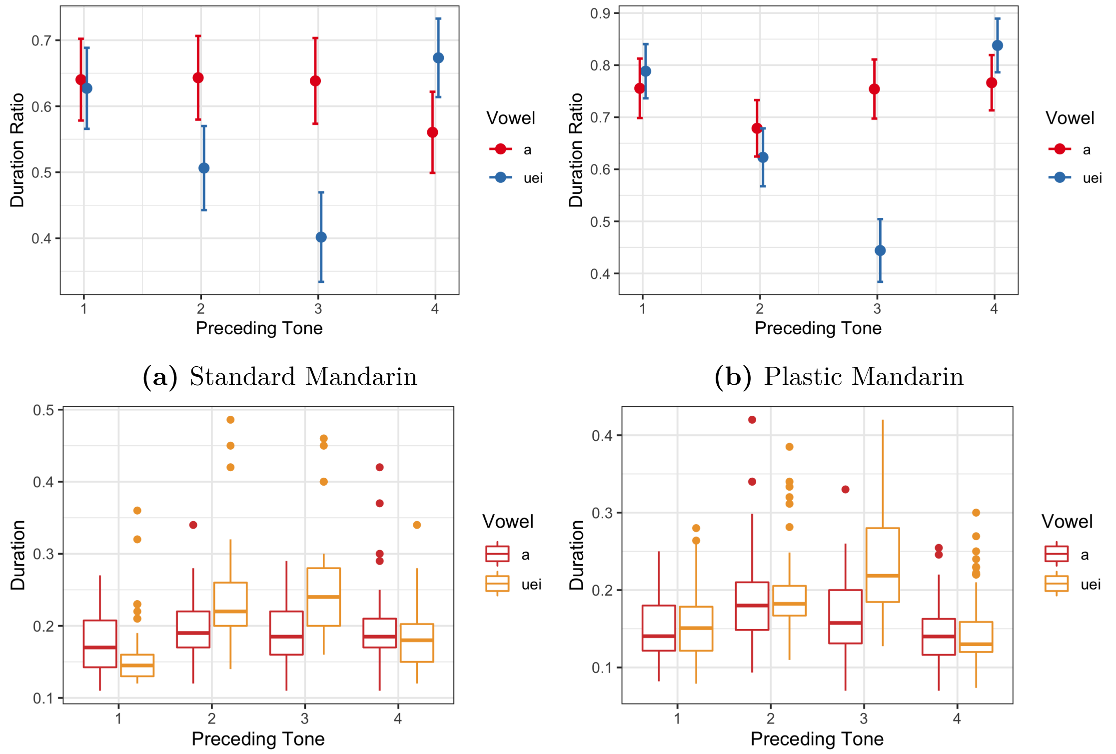
Figure 5.9: Waveform and automatic forced alignment of the sentence containing the disyllabic token (b) *duì de*, highlighted in light green. Time ticks of the pause boundaries are marked. 'The books we redeemed are really good'.



(a) Boxplots of absolute duration **(b)** Marginal effects of $Tone \times Variety$

Figure 5.10: The duration distribution and predicted duration of syllables in [X-*de*] utterances in Standard Mandarin and Plastic Mandarin. Syllable position 1 indicates lexical tone syllables and 2 indicates the neutral tone syllable *de*.

which might be marked by the pre-boundary lengthening of segmental duration (Hayes, 1995). As shown in Figure 5.8b, *de* has a long vowel [ə] (440 ms) and a portion as long as 189 ms at the end is in creaky voice (highlighted). The syllable after the pause *shū* 书 also has unusually long onset [ʃ] (650 ms), annotated with 'S' in the figure. These phenomena altogether suggest that the speech was hesitant, or the speaker was uncertain or still thinking about the key word she was guessing. The lengthened segments around the pause tend to serve to fill the pause partially. The hesitation effect overrides or superimposes on the otherwise shorter duration of the *de* syllables. This example reveals the complexity and vast variation in real-life semi-spontaneous speech data. It also suggests that the robust method is a reasonable and efficient way to find systematic patterns in the majority of productions despite the interference from a small number of undesired productions where other factors may have drastically affected the phonetic form. An alternative way can be manually discarding such data, which involves token-by-token examination and a labelling system with a certain threshold (e.g. the length of pause or silence period after *de*). The latter might be difficult and too laborious for a large dataset.



(a) Standard Mandarin **(b)** Plastic Mandarin

(c) Standard Mandarin **(d)** Plastic Mandarin

Figure 5.11: The predicted duration ratio of syllables (a, b) and duration of the first syllables (c, d) in [X-*de*] utterances by preceding tones and vowels in Standard Mandarin and Plastic Mandarin.

Despite that *de* syllables are generally shorter in both varieties, the third group of LME models suggest that there is a statistically significant difference between the two Mandarin varieties. The duration ratio of *de* to the corresponding first lexical tone syllable in Plastic Mandarin is about 0.118 ± 0.026 (standard error) larger than that in Standard Mandarin ($t(35.4) = -4.46 < .001$). In other words, the duration difference between *de* and the preceding syllable tends to be larger in Standard Mandarin than Plastic Mandarin. Figure 5.10 further presents the comparison of syllable duration in the [X-*de*] utterances in the two Mandarin varieties. While the *de* syllables have similar duration in both Mandarin varieties, the duration of their preceding lexical tone syllables tends to differ.

Figure 5.11 shows the effects of the preceding tone categories and vowel types in the duration ratio of *de* to the preceding lexical tone syllable in the two Mandarin varieties, estimated in the second group of robust LME models. In both Mandarin varieties, the disyllabic phrases starting with the Tone 3 syllable [tʰueɪ] with a triphthong vowel tend to have lower duration ratio than that with the Tone 3 syllable [ta]. The categorisation by vowel was misleading because the consonant of the first syllable in the disyllabic phrases was different in the [T2-de] and [T3-de] phrases due to tonotactic gaps[5]. The difference in duration ratio in question may in fact be largely attributed to the longer syllable duration of Tone 3 syllable [tʰueɪ] with an aspirated plosive compared to Tone 3 syllable [ta], as shown in (c, d) in Figure 5.11. From these findings, we cannot safely conclude whether it is the vowel or the intrinsic duration of different tone of the preceding syllable that influences on the duration ratio of the neutral tone syllable *de*.

5.3.3 The duration of polysyllabic phrases with *men* in two Mandarin varieties

5.3.3.1 Central tendencies and distributions

Figures 5.12 and 5.13 present the distribution of absolute duration of syllables in the [X_1X_2-*men*X_4] quadrisyllabic utterances, grouped by initial and last lexical tone categories, in Standard Mandarin and Plastic Mandarin respectively. Correspondingly, Tables 5.4 and 5.5 supply the mean and standard deviation of absolute duration of syllables by initial and last tone categories. These visualisation and descriptive statistic summaries give us preliminary insights about the duration data.

For Standard Mandarin, we can observe the following:

1) The reduplicated syllable X_2 with an average length of 125 ms ($\sigma = 32$ ms) tends to be shorter than the preceding syllable X_1, with an average length of 154 ms ($\sigma = 50$ ms).
2) While the reduplicated syllable X_2, plural particle *men*, and possessive particle *de* in the X_4 position are all considered as neutral tone syllables, their duration tends to differ from each other. The *de* in the X_4 position tends to be shorter than the reduplicated syllable X_2, as shown in the figure where the blue peak

[5]There is no [tueɪ] in Tone 2, and only one rare character duǐ [tueɪ] in Tone 3.

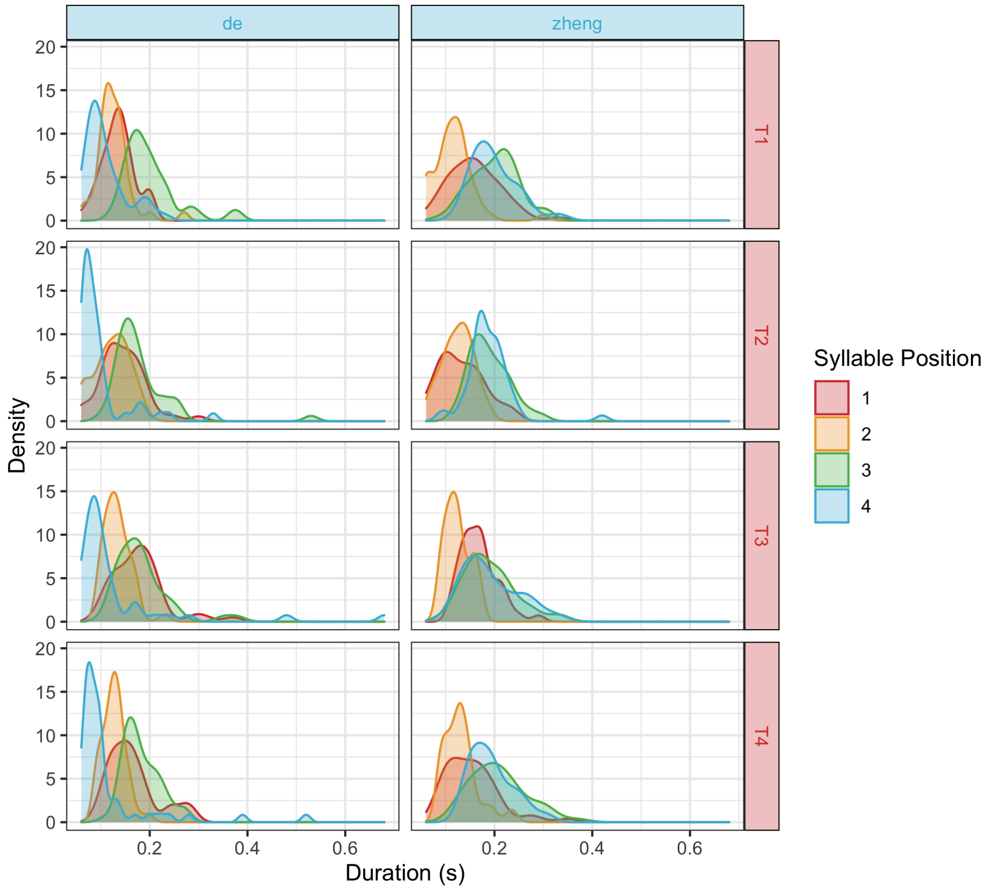


Figure 5.12: Kernel density functions for the duration (s) of syllables in the quadrisyllabic utterances Standard Mandarin by initial (red) and last (blue) tone categories.

in the 'de' column is always to the left of the yellow peak. The neutral tone *men* with a mean of 192 ms ($\sigma = 52$ ms) tends to be longer than the lexical tone syllable X_1, which is consistent with findings in Chapter 4.

3) The neutral tone *men* does not seem to differ from the lexical tone syllable *zheng* in the X_4 position.
4) Tone 3 syllables in the X_1 position tend to be longer than other X_1 syllables.
5) The *men* followed by another neutral tone syllable *de*, with a mean of 186 ms ($\sigma = 54$ ms), tends to be shorter than the *men* followed by *zheng* ($\mu = 197$ ms, $\sigma = 54$ ms).

Likewise, the observations about Plastic Mandarin are as follows:

1) Similar to Standard Mandarin, the reduplicated syllable X_2, with an average length of 122 ms ($\sigma = 35$ ms), tends to be shorter than the preceding syllable

X_1	X_4	Lexical tone X_1 $\mu(\sigma)$	Neutral tone X_2 $\mu(\sigma)$	*men* $\mu(\sigma)$	X_4 $\mu(\sigma)$
T1	T0 *de*	0.136 (0.033)	0.126 (0.034)	0.199 (0.055)	0.109 (0.042)
T1	T4 *zheng*	0.160 (0.054)	0.116 (0.041)	0.201 (0.049)	0.199 (0.047)
T2	T0 *de*	0.146 (0.046)	0.123 (0.036)	0.179 (0.065)	0.098 (0.054)
T2	T4 *zheng*	0.132 (0.047)	0.125 (0.032)	0.189 (0.040)	0.188 (0.047)
T3	T0 *de*	0.177 (0.054)	0.134 (0.027)	0.183 (0.055)	0.129 (0.115)
T3	T4 *zheng*	0.168 (0.038)	0.124 (0.025)	0.190 (0.054)	0.192 (0.057)
T4	T0 *de*	0.165 (0.051)	0.127 (0.024)	0.183 (0.037)	0.114 (0.084)
T4	T4 *zheng*	0.153 (0.054)	0.130 (0.035)	0.208 (0.057)	0.194 (0.044)
	T0 *de*	0.156 (0.049)	0.127 (0.030)	0.186 (0.054)	0.112 (0.079)
	T4 *zheng*	0.153 (0.050)	0.124 (0.034)	0.197 (0.050)	0.193 (0.049)
Overall		**0.154** (0.050)	**0.125** (0.032)	**0.192** (0.052)	

Table 5.6: Mean and standard deviation of the absolute duration (s) of the syllables by the initial and last tone categories in the [X_1X_2-*men*X_4] quadrisyllabic utterances in Standard Mandarin.

X_1, with an average length of 154 ms ($\sigma = 50$ ms).

2) The reduplicated syllable X_2, plural particle *men*, and possessive particle *de* in the X_4 position differ from each other in duration. Despite a larger mean, the *de* in the X_4 position is still likely to be shorter than the reduplicated syllable X_2, given that the blue peak in the 'de' column is to the left of the yellow peak. The large standard deviation of *de* (81 ms) may indicate a biased mean (129 ms) resulting from unusually long tokens. While X_2 and *de* tends to be shorter than the lexical tone syllable X_1, the neutral tone *men* with a mean of 179 ms ($\sigma = 52$ ms) tends to be longer than the lexical tone syllable X_1. This is consistent with the Standard Mandarin findings.

3) The neutral tone *men* does not seem to differ from the lexical tone syllable *zheng* in the X_4 position.

4) Tone 3 syllables in the X_1 position tend to be longer than other X_1 syllables.

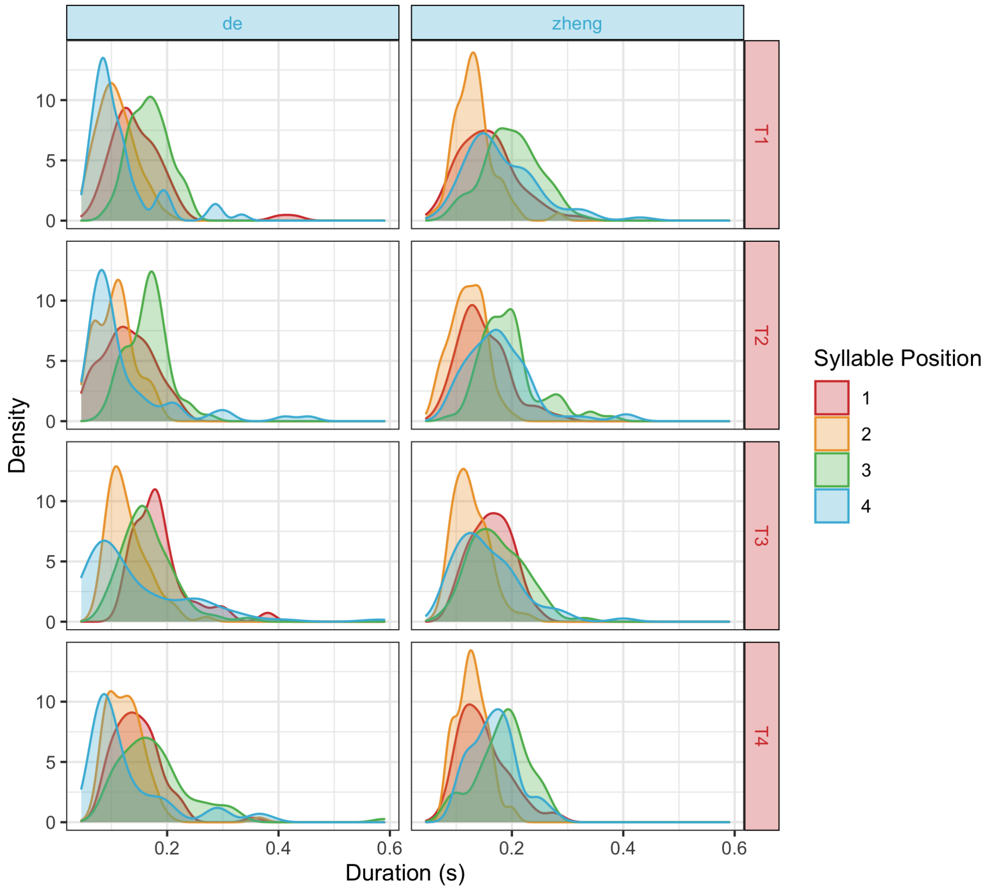


Figure 5.13: Kernel density functions for the duration (s) of syllables in the quadrisyllabic utterances Plastic Mandarin by initial (red) and last (blue) tone categories.

5) The *men* followed by another neutral tone syllable *de*, with a mean of 170 ms ($\sigma = 52$ ms), tends to be shorter than the *men* followed by *zheng* ($\mu = 188$ ms, $\sigma = 50$ ms).

The parallel observations show that the duration patterns of syllables in such quadrisyllabic sequences in the two Mandarin varieties are quite similar. In addition, the distributions in the figures for both Mandarin varieties have exceptionally long right tails, which may signal tokens produced in an undesired fashion. This may also be reflected in the large standard deviations in the X_4 columns in the tables. In the following subsections, LME models are built to explore the sources of variability in the duration data.

X_1	X_4	Lexical tone X_1 $\mu(\sigma)$	Neutral tone X_2 $\mu(\sigma)$	*men* $\mu(\sigma)$	X_4 $\mu(\sigma)$
T1	T0 *de*	0.150 (0.060)	0.108 (0.033)	0.168 (0.035)	0.115 (0.059)
T1	T4 *zheng*	0.157 (0.052)	0.132 (0.036)	0.200 (0.049)	0.182 (0.066)
T2	T0 *de*	0.130 (0.045)	0.107 (0.034)	0.165 (0.037)	0.122 (0.081)
T2	T4 *zheng*	0.148 (0.044)	0.121 (0.030)	0.195 (0.054)	0.177 (0.062)
T3	T0 *de*	0.186 (0.053)	0.129 (0.037)	0.165 (0.046)	0.147 (0.096)
T3	T4 *zheng*	0.163 (0.037)	0.126 (0.031)	0.174 (0.049)	0.157 (0.060)
T4	T0 *de*	0.147 (0.044)	0.126 (0.040)	0.180 (0.075)	0.130 (0.077)
T4	T4 *zheng*	0.148 (0.044)	0.138 (0.027)	0.184 (0.045)	0.169 (0.042)
	T0 *de*	0.154 (0.055)	0.118 (0.038)	0.170 (0.052)	0.129 (0.081)
	T4 *zheng*	0.154 (0.045)	0.127 (0.031)	0.188 (0.050)	0.171 (0.059)
Overall		**0.154** (0.050)	**0.122** (0.035)	**0.179** (0.052)	

Table 5.7: Mean and standard deviation of the absolute duration (s) of the syllables by the initial and last tone categories in the [X_1X_2-*men*X_4] quadrisyllabic utterances in Plastic Mandarin.

5.3.3.2 Comparisons by linguistic contexts and varieties

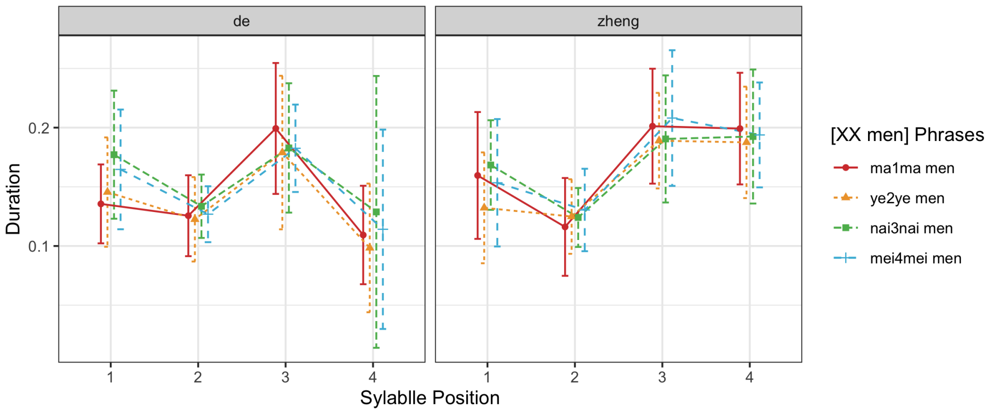
The robust linear mixed-effects models predict that the reduplicated syllable X_2 is on average significantly shorter than the initial syllable X_1 in most cases for Standard Mandarin. More specifically, in the dataset of all X_1X_2 tokens, the second syllable is generally shorter than the initial syllable in the phrases of Tone 1 *māma* by 25.2 ms ±5.2 ms standard errors ($z = 4.85, p < .001$), Tone 3 *nǎinai* by 41.3 ms ±5.4 ($z = 7.68, p < .001$), and Tone 4 *mèimei* by 26.0 ms ±5.0 ms ($z = 5.17, p < .001$). Although the second syllable is on average shorter in the Tone 2 phrase *yéye*, the difference does not reach statistical significance. Yet if we look at the [X_1X_2-*men de*] subset, the second syllable is significantly shorter than the initial syllable in *yéye* ($z = 3.50, p = .039$), but the difference between *māma* does not reach statistical significance. The robust LME models also indicate that on average the final neutral tone syllable *de* in the utterance sequence is statistically significantly shorter than the reduplicated neutral tone syllable X_2 in phrases of *yéye* by 33.3 ms ±6.6 ms standard errors ($z = 5.01, p < .001$), *nǎinai* by 28.9 ms ±7.0 ms($z = 4.12, p = 0.004$), and *mèimei* by 31.8 ms ±6.4 ms ($z = 4.99, p < .001$).

It is a similar case for reduplicatives in Plastic Mandarin. The robust LME models predict that the reduplicated syllable X_2 is on average significantly shorter than the initial syllable X_1 in all phrases in the dataset containing all X_1X_2 tokens: in *māma* shorter by 29.0 ms ±4.6 ms standard errors ($z = 6.29, p < .001$), *yéye* by 23.3 ms ±4.6 ms ($z = 5.09, p < .001$), *nǎinai* by 45.0 ms ±4.4 ms($z = 10.33, p < .001$), *mèimei* by 19.1 ms ±4.6 ms ($z = 4.20, p = .002$). However in the [X_1X_2-*men de*] subset, the duration of the reduplicated neutral tone syllable X_2 does not statistically significantly differ from the duration of the final neutral tone syllable *de* in the utterance sequence in all four types of phrases.

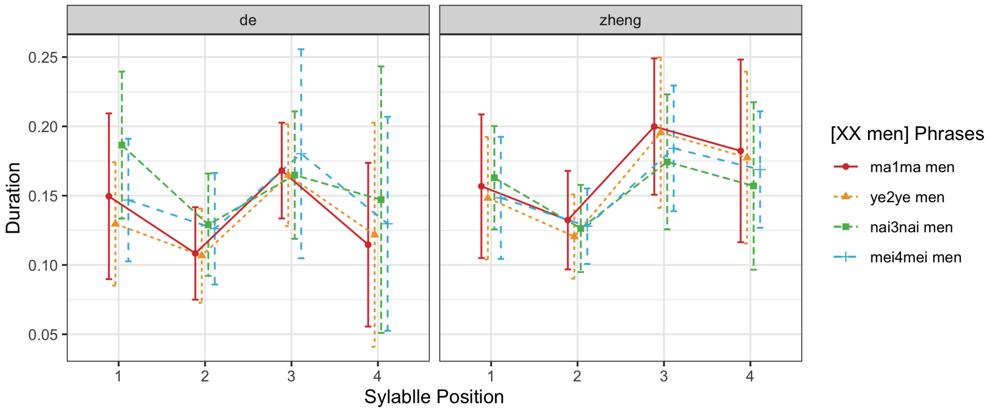
Figure 5.14 shows the absolute duration of the syllables in the two types of quadrisyllabic utterances in both Standard Mandarin and Plastic Mandarin. The mean duration of *men* tend to differ between the two conditions differentiated by the following syllable in both Mandarin varieties. The two types of phrases [X_1X_2-*men de*] and [X_1X_2-*men zhèng*] have different syntactic structures, which may correspond to prosodic domains of distinct levels (Q. Li & Chen, 2019). The longer duration of *men* preceding *zhèng* could potentially be explained by the lengthening effect of prosodic word boundary (marked by #), as demonstrated in a) and b), but in this case it cannot rule out the confound of a different following syllable.

a) *yéye-men# zhèng*
b) *yéye-men-de*

A series of LME models suggest that in Standard Mandarin the differences of predicted average duration between *men* preceding the Tone 4 syllable *zhèng* and *men* preceding neutral tone syllable *de* are not statistically significant except in Tone 2 phrases: the former is longer than the latter by 23.1 ms ±6.9 ms standard errors ($z = 3.36, p = 0.018$). In Plastic Mandarin, *men* preceding *zheng* is on average longer than *men* preceding *de* in only Tone 1 phrases by 30.1 ms ±6.1



(a) Standard Mandarin

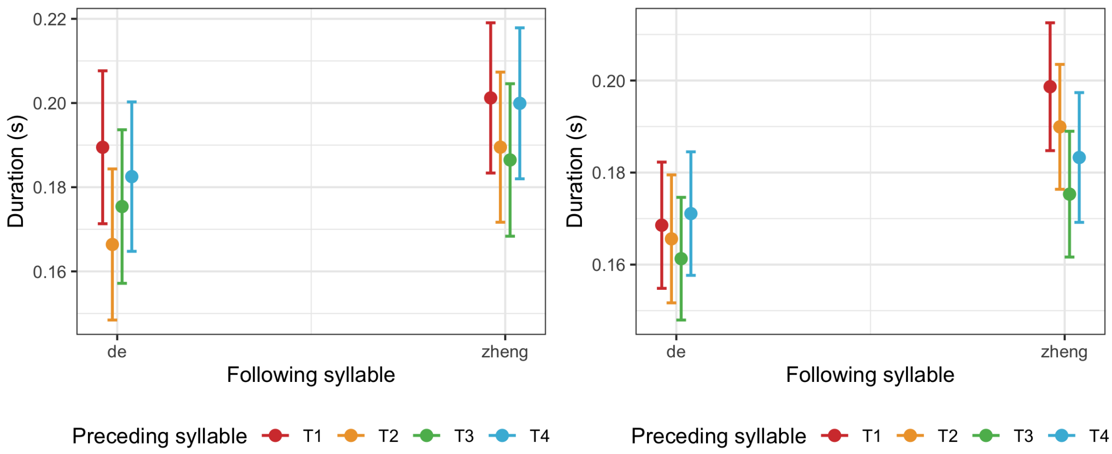


(b) Plastic Mandarin

Figure 5.14: Duration by syllable position in the two types of quadrisyllabic utterances in Standard Mandarin and Plastic Mandarin. The error bars indicate ± one standard deviation.

ms standard errors ($z = 4.48, p < .001$) and Tone 2 phrases by 24.3 ms ±6.7 ms standard errors ($z = 3.66, p = 0.006$). Figure 5.15 shows the predicted duration of *men* by preceding and following syllables. Since such significant difference does not occur consistently in all phrases, we cannot safely conclude the potential effect of prosodic word boundary.

The last group of robust LME models introduced in §5.3.1 concern the duration ratio of the first or second neutral tone syllable to the initial lexical tone syllable. The duration ratio differences between the two Mandarin varieties are not statistically significant.



(a) Standard Mandarin **(b)** Plastic Mandarin

Figure 5.15: The predicted duration (s) of the neutral tone syllable *men* in different linguistic contexts in Standard Mandarin and Plastic Mandarin. The error bars indicate 95% confidence intervals.

5.4 Pitch Contours of Neutral Tones in Mandarin Accents

In Chapter 4, we have observed a converging trend of considerably variable f_0 contours at the end of two neutral tone syllables in a Standard Mandarin corpus. Similar converging effect in Standard Mandarin was previously reported in Z. Li (2003) and Y. Chen and Xu (2006), which suggested a stable tonal target around the mid-low level of the pitch range. This section investigates f_0 realisation of neutral tone in Plastic Mandarin with comparative Standard Mandarin data, and answers the following specific research questions:

1) Does the neighbouring lexical tones have an effect on neutral tone realisation? if so, how?
2) Is there a pitch target for neutral tone in Plastic Mandarin? If so, how does the pitch target compare to that in Standard Mandarin?

This section models f_0 contours of phrases involving neutral tone syllables and their interaction with several predictors, using Generalised Additive Mixed Models (GAMMs). One key factorial predictor is the preceding lexical tone category of a neutral tone, that is, the tone category of X in [X-*de*] phrases and X_1 in [X_1X_2-*men*X_4] phrases. Other factorial factors include the vowel type of the preceding lexical tone syllable, and the following tone category (X_4) of a neutral tone. The full analysis script is available in the online supplementary materials.

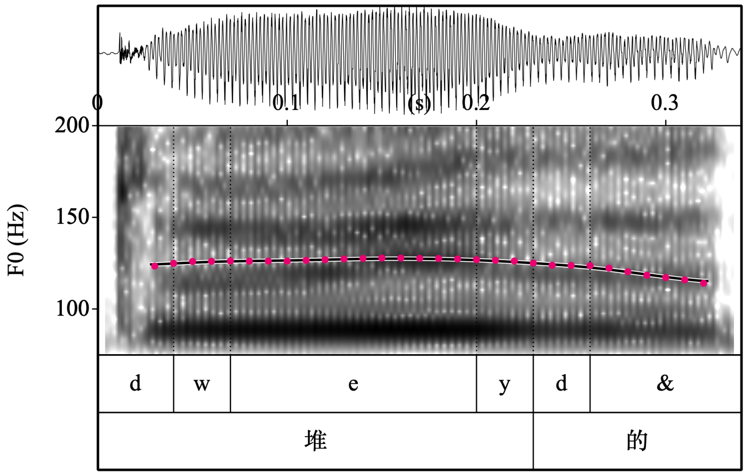
5.4.1 Statistical Method

The f_0 points measured in Hz were converted to semitones using the same method introduced in §3.4.2. The reference f_0 for the conversion is still the speaker mean defined in §3.4.2 based on a balanced inclusion of monosyllables of each tone in a variety. The speaker mean can be conceptualised as the mean of the tonal space of a speaker when speaking the said variety.

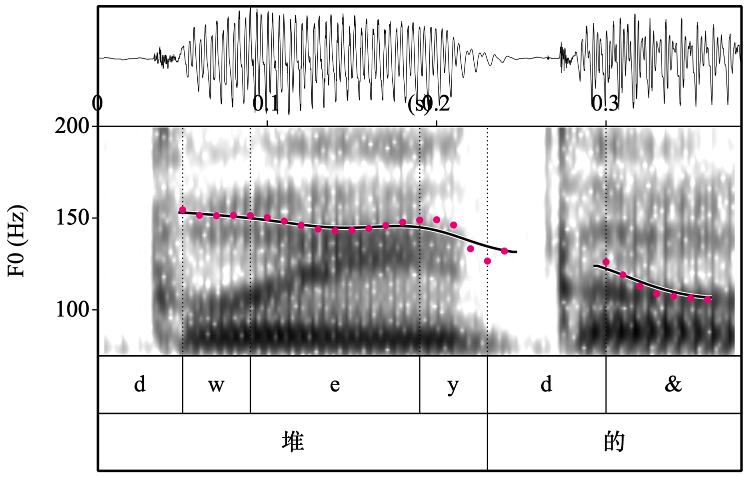
5.4.1.1 Time Normalisation

The acoustic realisation of the disyllabic sequence [X-*de*] varies from token to token. The initial consonant of *de*, a voiceless unaspirated plosive [t], is sometimes realised as completely voiced between the vowels or even deleted. In Figure 5.16 (a), no clear closure period of a stop is shown in the waveform or the spectrogram, while in (b), the closure and release are clearly present, indicated by an almost flat signal followed by irregular noise in the waveform. The segmental realisation has influence over f_0. In (a), the f_0 contour is smooth and continuous while in (b) there is a break in the f_0 contour and a small peak and valley (pink dots) right before the break. Despite the short break, the overall trend of this f_0 contour is smooth, that is, the onset f_0 of *de* tends to be on a track (black line) that is extrapolated from the offset f_0 of the preceding syllable. Such broken curves in this thesis are treated as smooth and continuous curves or functions where the break is interpolated. Therefore, an f_0 contour of an [X-*de*] phrase spans the voicing portion (rhyme) of the first syllable and almost the entire second syllable, except for the creaky pulses at the end of *de* if there are any.

The average duration of the voicing portion of the first syllable is 146 ms (±47.9 ms standard deviation) and the average duration of *de* excluding phrase-final creaky



(a) Speaker 212



(b) Speaker 213

Figure 5.16: The waveform, spectrogram (0-5000Hz), and f_0 measurement of two tokens of *duīde* 'built' by two speakers. TextGrid alignment is generated by P2FA. Black line is the smoothed f_0 curve. Pink dots indicate f_0 measurement.

pulses is 145 ms (±99.6 ms standard deviation) in Standard Mandarin. In Plastic Mandarin, the former is 147 ms (±47.2 ms standard deviation) and the latter is 140 ms (±82.4 ms standard deviation). The ratio between the average duration of these two parts is approximately 1:1 in both Mandarin varieties. The scales of $[-1, 0]$ and $[0, 1]$ were then motivated for the time normalisation of these two parts respectively so that the overall shape would not be greatly distorted.

There is usually no f_0 gaps in the trisyllabic sequence [X_1X_2-*men*] since all the selected kinship terms and *men* are fully voiced. The average duration of the voicing portion where f_0 is measured of the X_1, X_2, and *men* in Standard Mandarin is 148 ms (±56.3 ms standard deviation), 130 ms (±35.3 ms standard deviation), and 198 ms (±56.1 ms standard deviation) respectively. In Plastic Mandarin, the average duration of the voicing portion of the X_1, X_2, and *men* is 152 ms (±58.9 ms standard deviation), 128 ms (±38.7 ms standard deviation), and 186 ms (±58.7 ms standard deviation) respectively. In both Mandarin varieties, the ratio of reduplicative kinship term [X_1X_2] and *men* against the whole trisyllabic phrase is approximately 60% and 40%.

In order to enable visual comparison between [X-*de*] phrases and [X_1X_2] phrases, the scales of $[-1, 0]$ and $[0, 1]$ were then motivated for the time normalisation of [X_1X_2], given only small duration difference between X_1 and X_2. The scale $[1, 2.35]$ was motivated for the time normalisation of *men*, given that *men* is generally longer than X_1 and X_2.

The duration of *de* excluding phrase-final creaky pulses in [X_1X_2-*men de*] phrases is 119 ms (±81.1 ms standard deviation) in Standard Mandarin and 138 ms (±89.2 ms standard deviation) in Plastic Mandarin. As discussed in §5.3.3, *de* in quadrisyllabic phrases is likely to be the shortest syllable in both Mandarin varieties, although the mean duration of *de* may not reflect that, due to a small number of *de* were produced in undesirably long fashion during the game. Therefore, the scale $[2.35, 3.3]$ (a normalised duration of 0.95) was motivated for the time normalisation of *de* in quadrisyllabic phrases in both Mandarin varieties, despite their mean duration difference, to facilitate cross-variety comparison.

Time was normalised in a linear fashion for the relevant part of each syllable in each token. In this way, the overall f_0 contours of the [X-*de*] and [X_1X_2-*men*] phrases can be compared, as well as the f_0 contours of syllables at each position.

5.4.1.2 GAMM construction

Generalised additive mixed models (GAMMs) were fitted to the data to statistically assess differences among f_0 contours of phrases containing neutral tones. In general, the parametric term in our models is related to the overall f_0 heights, and the smooth term is associated with the f_0 shapes. The approach of GAMMs with difference terms (Sóskuthy, 2021) was adopted. In this way, we fit a baseline smooth to f_0 contours of one type of phrase and then other smooths representing their non-linear differences to the baseline smooth. Such models can provide significance testing based on summary statistics of terms extracted from the model. Following Sóskuthy (2017), the variables coding the difference between the sets of contours were converted to ordered factors with contrast coding before the model fitting. In all the models, the f_0 in semitone is the dependent variable. Two broad sets of GAMMs were used: in Set 1 each model was fitted to the dataset of each Mandarin variety separately; in Set 2 each model was fitted to the combined dataset of two Mandarin varieties.

Set 1a: the effects of preceding lexical tone category and vowel

The following terms were included for the analysis of [X-*de*] tokens: the ordered factors of the tone category of X and vowel of X as the parametric terms to capture overall differences between the trajectories, a smooth over the normalised time, a smooth over the overall duration measured in seconds, a tensor product interaction between overall duration and normalised time to account for the influence of duration on the trajectories (see Sóskuthy, 2017), two difference smooth over the normalised time by the factors of X tone category and X vowel respectively, and lastly, random smooth over the normalised time with a special grouping variable that creates separate levels for each combination of speaker, X tone category, and X vowel, since it is possible that the effects of the tone category and vowel of X vary across speakers. Such *Speaker* × *Effects* random smooth (more specifically, *Speaker* × *Tone* × *Vowel*)

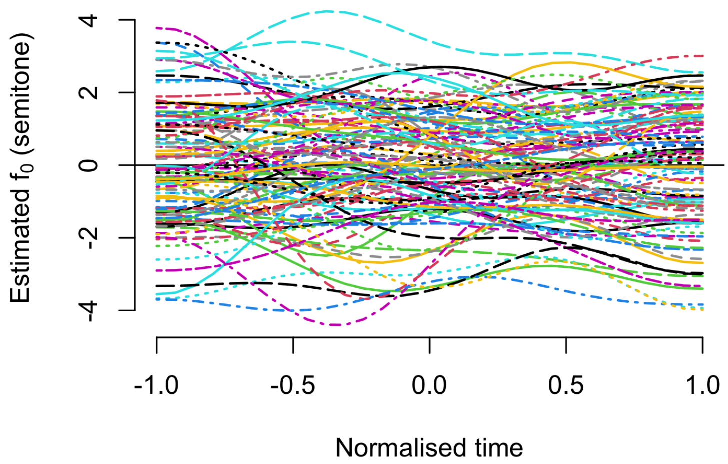


Figure 5.17: Individual by-phrase type (initial tone category and vowel) adjustments to the general f_0 contours in the GAMM for Standard Mandarin disyllabic tokens. The average of these adjustments is approximately 0, i.e. centred.

allows potential speaker-specific trends in the main effects. Figure 5.17 is an example of random factor smooths, which are a nonlinear alternative to random intercepts and slopes in a mixed-effects regression model. They are essential to allowing for the structural variability associated with individual speakers and phrases and thereby prevent anti-conservative (i.e. too low) p-values (Wieling et al., 2016).

Set 1b: the effects of preceding lexical tone category and following syllable

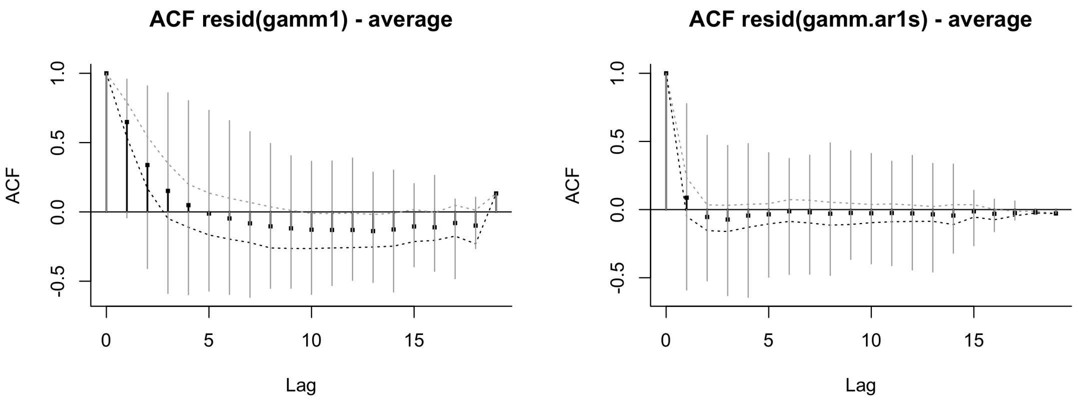
The following terms were included for the analysis of [X_1X_2-*men*] tokens: the ordered factors of the tone category of X_1 and the following syllable of *men* X_4 as the parametric terms to capture overall differences between the trajectories, a smooth over the normalised time, a smooth over the overall duration measured in seconds, a tensor product interaction between overall duration and normalised time, two difference smooths over the normalised time by the factors of X_1 tone category and X_4 syllable respectively, and lastly, random smooths over the normalised time with a special grouping variable that creates separate levels for each combination of speaker, X_1 tone category and X_4 syllable.

Set 1c: the effects of preceding lexical tone category

The following terms were included for the analysis of [X_1X_2-*men de*] tokens: the ordered factors of the tone category of X_1 as the parametric terms, a smooth over the normalised time, a smooth over the overall duration measured in seconds, a tensor product interaction between overall duration and normalised time, the difference smooths over the normalised time by the factors of X_1 tone category, and lastly, random smooths over the normalised time with a grouping variable of *Speaker* to capture nonlinear individual differences.

Set 2: cross-variety effects

Whereas fitting two separate models (one for each variety in Set 1) and visually comparing the patterns might generate insights about cross-variety variations, this approach may not be adequate, since it is not possible to evaluate whether the additional complexity of the language variety factor is warranted. Thus the second set of models is built on the combined dataset of Plastic Mandarin and Standard Mandarin. A new ordered factor variable was created by grouping the X tone category and Mandarin variety, *Tone* × *Variety*, to include variety-specific tone category effect. This combined variable replaced the factor of X tone category in the model of Set 1a and the factor of X_1 tone category in the models of Set 1b and Set 1c, while the rest of specification remains the same, except that the random factor smooth has a two-way grouping variable *Speaker* × *Vowel* in the Set 1a model and *Speaker* × X_4 *syllable* in the Set 1b model. This was because the data structure is different from that in Set 1. In the dataset for each Mandarin variety, each speaker is represented by contours from all tone and vowel groups, in other words, the effect category varies within speakers. However, in the combined dataset, some speakers were in the Standard Mandarin group and others in the Plastic Mandarin group. This set focuses on across-speaker effects, based on the assumption that "there is less variation in contour shapes within individual speakers than there is in the dataset as a whole" (Sóskuthy, 2021, p. 13).



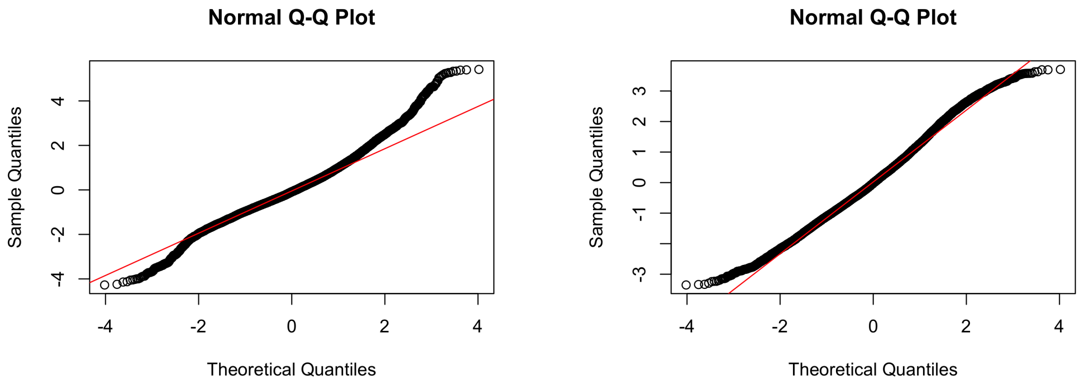
(a) (b)

Figure 5.18: Autocorrelation in the residuals of the models for Standard Mandarin disyllabic tokens. (a) without correction; (b) after correction using the AR1 error model. Lag indicates the number of time steps.

5.4.1.3 Model criticism and robustness

GAMMs were constructed using the *bam* function of the `mgcv` package (S. Wood, 2022). Since autocorrelation in f_0 contours is expected for the time dimension, the autoregressive (AR) error model with an AR1 component was employed to reduce autocorrelation of the model residuals within trajectories. An AR1 model assumes that the errors for immediate neighbouring observations in the contours are correlated (Sóskuthy, 2021). Figure 5.18 illustrates the averaged residuals for all trajectories in the GAMMs with and without the AR1 component for the Standard Mandarin disyllabic dataset. (a) suggests that there is high positive autocorrelation at lag 1. In (b), the autocorrelation at lag 1 has been much reduced and the remaining autocorrelation values are low, between 0 and -0.1. Hence the autoregressive error model enables a more reliable assessment of the model fit and the associated p-values (Wieling et al., 2016). The correlation parameter was set according to the autocorrelation function (ACF) value at lag = 1 (Carignan et al., 2020).

GAMM smooths are penalised regression composed of basis functions, usually splines such as cubic or thin plate regression splines, and the number of basis functions or knots (k parameter) determine the flexibility of a smooth. The default smooth of plate regression splines was employed in the models, considering its



(a) **(b)**

Figure 5.19: The quantile-quantile plot of the residuals of the models for Standard Mandarin trisyllabic tokens. (a) a Gaussian model; (b) a scaled-t model.

fast computation. Since the smallest number of f_0 measurements for a [X-*de*] token was 11 in the Standard Mandarin dataset and 12 in the Plastic Mandarin dataset, the number of knots for these smooths cannot be set higher than 11 or 12 respectively. The chosen k parameters were checked via model diagnosis using the `gam.check()` function.

In addition to the strategy to deal with the inevitable autocorrelation in the residuals, we inspect the normal distribution and homoscedasticity of the residuals, using the `check_resid()` command from the `itsadug` package. Figure 5.19 (a) presents an example quantile-quantile (QQ) plot for the Set 1b model for Standard Mandarin data. The QQ plot suggests that the distribution of the residuals for this model has two heavy tails, which may be indicative of a t distribution. An advantage of GAMMs is that it extends from the generalised linear mixed models which allows different error distribution. We thus remodelled the dataset with a scaled-t distribution, transforming the residuals back to normality (Chuang et al., 2021), and the QQ plot of the updated model is shown in (b). The error distribution is now much closer to a normal distribution in Figure 5.19 (b). The summary statistics in the following sections will be based on such updated models.

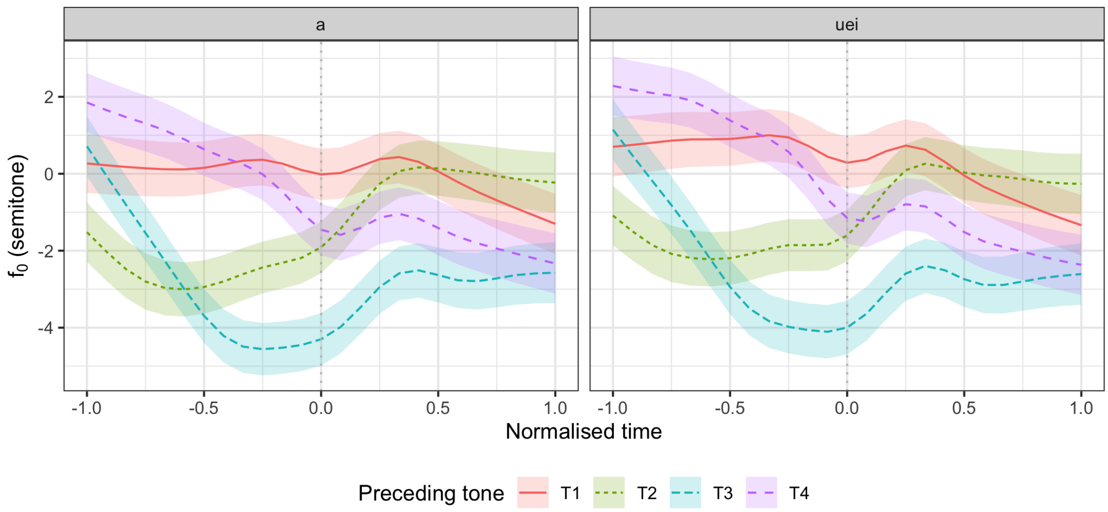
5.4.2 Disyllabic phrases with *de* in two Mandarin varieties

Figure 5.20 presents the fitted normalised f_0 contours for each of the eight phrases in the separate models of Standard Mandarin and Plastic Mandarin. Recall that we defined the first half of each graph, the normalised time domain $[-1, 0]$, for f_0 contours of the lexical tone syllable X and the second half, the normalised time domain $[0, 1]$, for the neutral tone syllable *de*. By comparing the patterns between the [a] set and the [uei] set in each Mandarin variety, the differences are unsurprisingly not substantial. The consistency of tone shapes within a tone category is conducive to differentiating lexical meanings in a tonal language. We now examine the f_0 contours in each Mandarin variety.

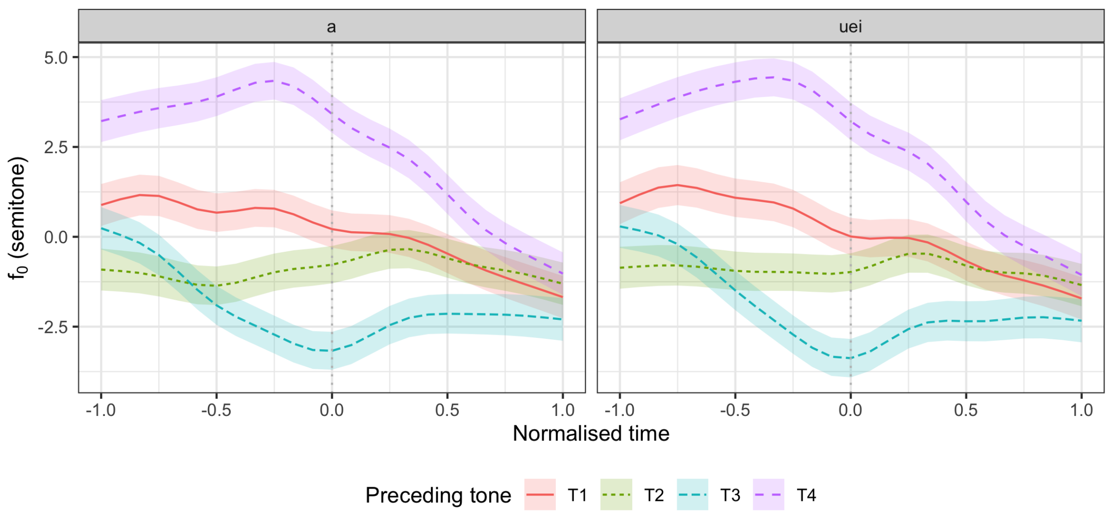
5.4.2.1 Standard Mandarin

The shapes of the first half of these f_0 contours largely resemble the patterns of their corresponding prototypical citation tones, presented in Chapter 3. For Standard Mandarin, Tone 1 here is flat and relatively high compared to Tones 2 and 3, although the estimated normalised f_0 is quite close to 0 (i.e. speaker mean) throughout the syllable. The prototypical citation Tone 1, estimated on the same group of speakers, has an average normalised f_0 of 2 semitones approximately. Tone 2 here generally has a dipping shape where the rise starts at about 40% into the rhyme of a syllable. The degree of rising is much less than that of the citation Tone 2, whose offset is about 2 semitones higher than the speaker mean. Tone 3 here is a large falling tone and the slope of fall decreases at about 70% into the rhyme of a T3 syllable. Tone 4 here is similar to its citation tone with a large fall, although the citation counterpart on average drops to as low as -3 semitones. To sum up, the lexical tone syllable in the stressed position of a disyllabic phrase largely maintains the canonical tone shape, but Tone 2 and Tone 4 tend to have smaller absolute slope and Tone 1 tends to be lower compared to their citation counterparts.

While the falling Tone 3 here may correspond to the low variant of Tone 3 in the literature, the estimated f_0 curve of the [T3-*de*] phrase tends to reveal a carryover effect of Tone 3 on the neutral tone — the canonical dipping shape of



(a) Standard Mandarin /(t)a tə/ and /(C)uei tə/

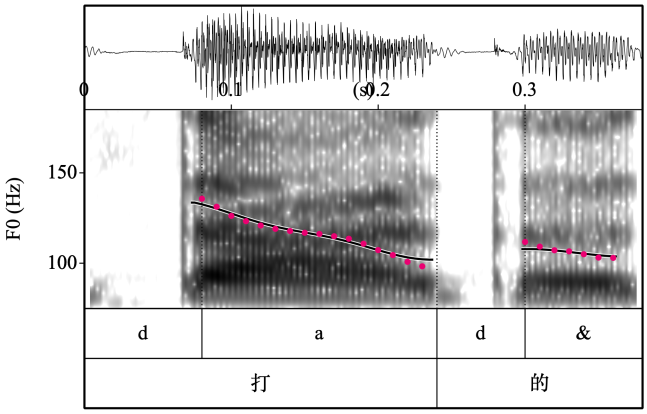


(b) Plastic Mandarin /(t)a tə/ and /(C)uei tə/

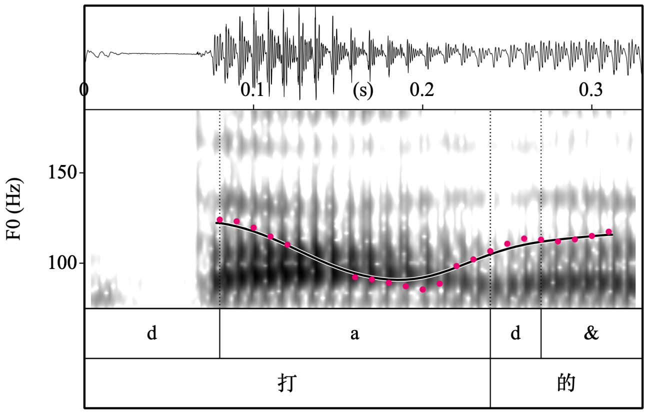
Figure 5.20: Fitted f_0 contours of eight [X-*de*] phrases, grouped by the vowel and tone category of the preceding syllable of *de*, by two different groups of Mandarin speakers. The shades indicate 95% pointwise confidence intervals.

Tone 3 seems to continue into the neutral tone syllable and is completed at about 30% into the second syllable. The end of such effect might be signalled by a small peak (at approximately 0.3 normalised time) in the T3 curves. Indeed, all neutral tone curves have a small peak around 0.3 in Standard Mandarin, which may relate to some f_0 perturbation after a stop consonant.

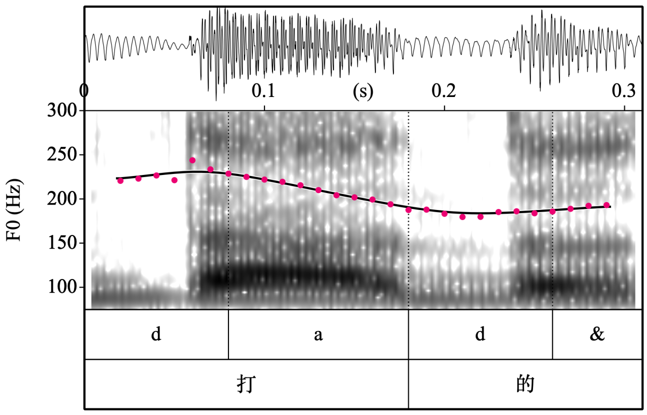
In order to see whether the estimated f_0 curve between $[0, 0.3]$ reflects the actual f_0 or the interpolation to fill the gap of the stop consonant, some random original



(a) Speaker 208



(b) Speaker 207



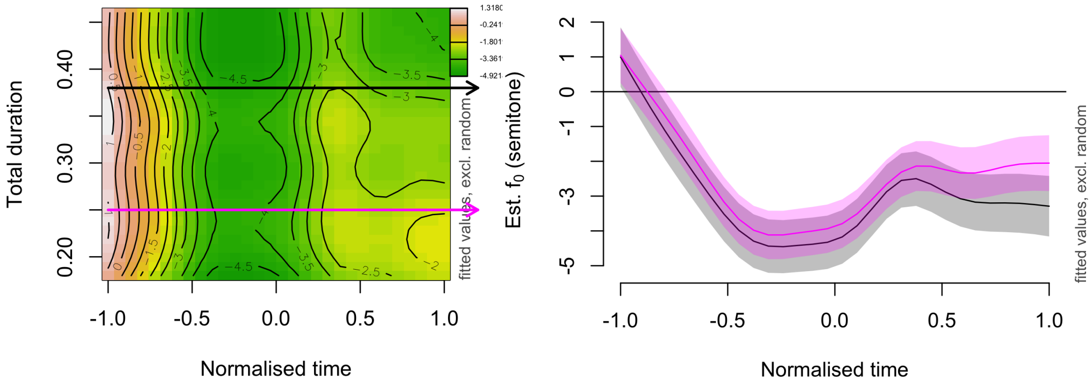
(c) Speaker 205

Figure 5.21: The waveform, spectrogram (0-5000Hz), and f_0 measurement of three tokens of *dǎde* 'made' by three speakers. TextGrid alignment is generated by P2FA. Black line is the smoothed f_0 curve. Pink dots indicate f_0 measurement.

tokens were examined. Figure 5.21 demonstrates three tokens of *dǎde* in different lengths by three speakers. Token (a) has a 60 ms gap (42% of the duration of *de*) without f_0 measurement, while Token (b) and (c) have a continuous f_0 transition from the first syllable to the second syllable. In (a) the initial f_0 point of *de* is about 28 Hz (2.3 semitones) higher than the offset f_0 point of the first syllable. In (b), f_0 starts to rise after 0.2s and *de* also has an onset f_0 that is 21 Hz (3.84 semitones) higher than the f_0 trough of the first syllable. In (c), the starting f_0 of *de* follows the f_0 fall in the first syllable, and it continues to go lower during the majority of the voicing portion of the consonant [t̬]. We also notice that the first syllable *dǎ* in (a) and (b) is longer than that in (c). The three example tokens have demonstrated a diverse realisation in terms of both segments and f_0 contour. They, however, might reveal the general case that the starting f_0 of *de* continues from a dipping shape of the preceding Tone 3 even if there is a gap in voicing resulting from a voiceless consonant. The shorter the first syllable, the more carryover effect is reflected in the f_0 contour of *de*. Tokens (b) and (c) are shorter overall, and their f_0 contour of *de* is different from that of (a), which has a negative slope (f_0 gradually decreases).

The model has indeed included a continuous variable of overall duration. Figure 5.22 (a) illustrates the nonlinear interaction surface of overall duration and normalised time for *dǎde* phrases in Standard Mandarin with a GAMM heatmap, created using the `fvisgam` function of the `itsadug` R package (van Rij et al., 2022). Figure 5.22 (b) is probably more informative by showing the f_0 contours estimated with two different total duration: 0.25s and 0.38s. From Figure 5.22, we can see that in longer tokens, the f_0 of *de* is predicted to decrease, reflected by the darker green colour towards 1 in (a) and falling grey curve in (b). On the contrary, the f_0 of *de* in shorter tokens is predicted to rise.

The fitted f_0 contours of *de* in Figure 5.20 (a) show various patterns. The end of an f_0 contour, though, tend to cluster in the f_0 range of $[-3, 0]$ and there are more overlapping regions in the pointwise confidence bands of different curves, which may indicate that the difference between the curves are not significant.



(a) **(b)**

Figure 5.22: The effect of duration across the normalised time estimated by the GAMM for Standard Mandarin disyllabic tokens of *dǎde*. (a) the interaction surface between overall duration and normalised time; The 0.1 to 0.9 quantile of the whole range of total duration is shown. f_0 in semitone is denoted by colour grade, and regions of equidistant change (here, Δ 0.5 semitones) are denoted by black lines. (b) Two estimated f_0 contours with two different total duration: 0.25 (pink) and 0.38 (black). The corresponding duration values are marked by arrows in (a).

Another visual method of significance testing is to plot the estimated difference between the curves and check whether the confidence interval around it includes zero (Sóskuthy, 2021). Figure 5.23 presents the estimated difference curves between each two groups of the disyllabic Standard Mandarin tokens [ta tə], in which the regions in red indicate where the confidence interval includes zero. In the $[0, 1]$ time domain, the estimated f_0 difference of *de* between between T1-*de* and T2-*de* phrases is mostly not significant. Another similar pair is the T3-*de* and T4-*de* phrases, and their difference curve of *de* gradually rises up to approximate 0. Starting from about 70% into the syllable *de*, their estimated f_0 difference is not significant. The estimated f_0 difference of *de* between between T1-*de* and T3-*de* phrases and T1-*de* and T4-*de* phrases decreases, although the associated confidence interval excludes zero.

The results here are similar to the corpus analysis of disyllabic phrases reported in §4.4.2, in which the f_0 at the end of a neutral tone syllable converges in two groups: in one group the preceding tones of a neutral tone are Tones 1 and 2; in the other group the preceding tones of a neutral tone are Tones 3 and 4. Figure 5.24 shows

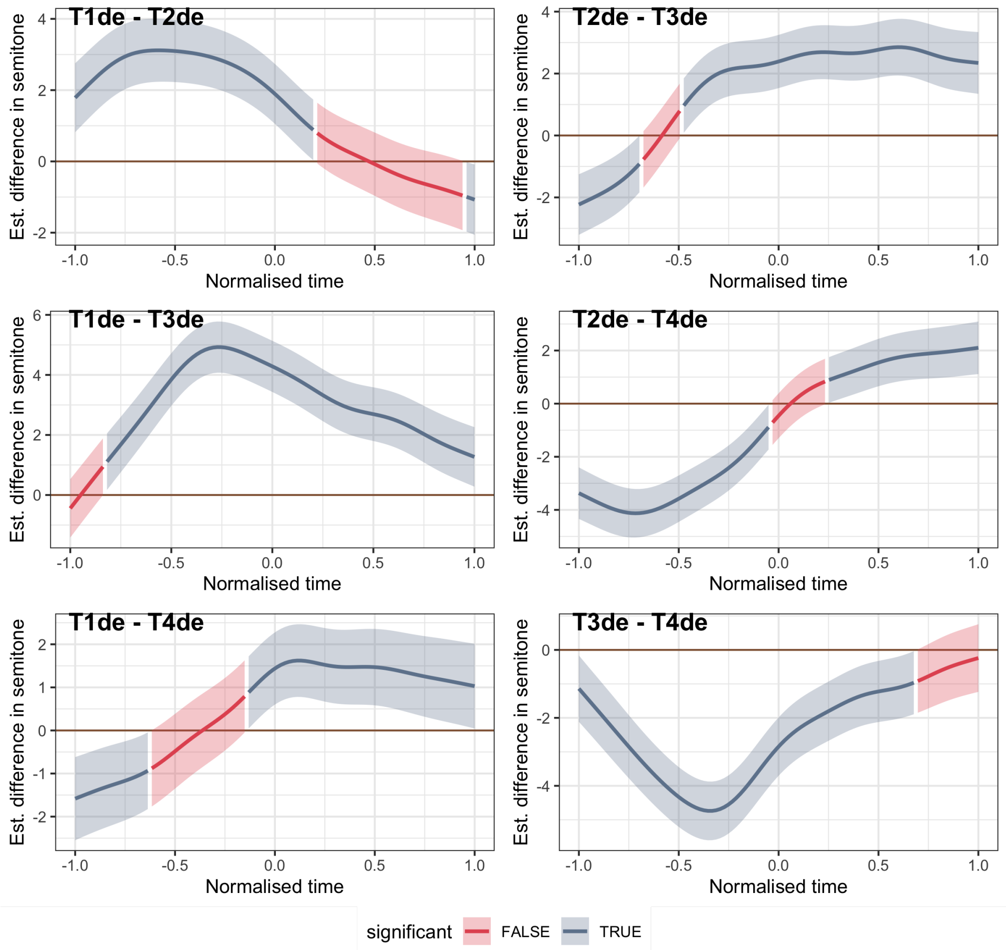


Figure 5.23: The estimated difference f_0 trajectories between each two groups of the disyllabic Standard Mandarin tokens [ta tə] with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red.

the distribution, kernel density estimation of ending f_0 at the end of *de*. Two fitted Gaussian distributions based on such two groups suggest an average ending f_0 of -0.64 semitones ($\sigma = 2.18$ semitones) for T1-*de* and T2-*de* phrases, and an average ending f_0 of -2.15 semitones ($\sigma = 2.27$ semitones) for T3-*de* and T4-*de* phrases.

5.4.2.2 Plastic Mandarin

For Plastic Mandarin, the lexical tones on the left half of the graphs in 5.20 (b) also resemble their citation form very much, especially the high rising Tone 4 almost reaching as high as 5 semitones above the speaker mean. Other tones have slight differences compared to their citation form. Tone 1 here slightly falls, Tone 2 in

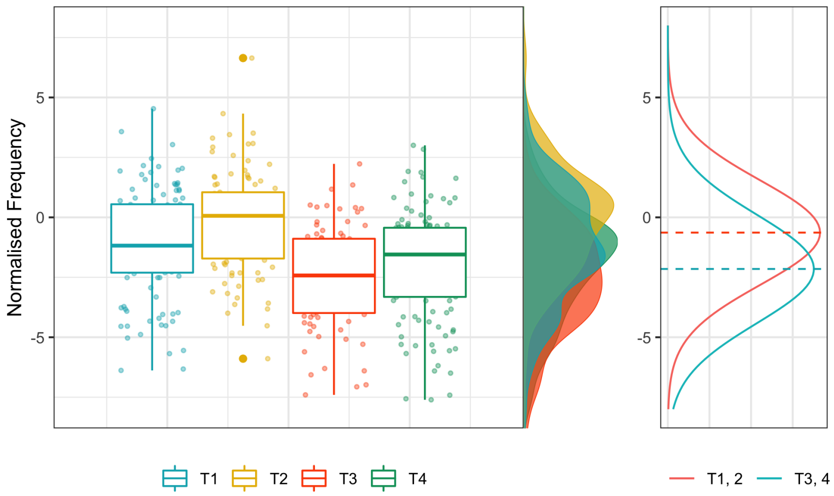


Figure 5.24: The distribution, kernel density estimation of ending f_0 of *de* in disyllabic Standard Mandarin tokens and the fitted Gaussian distributions of two groups. Five potential outliers with an absolute value of f_0 more than 8 were removed in the plot.

huíde here tends to be a level tone. Tone 3 here is a falling tone, but the overall f_0 range is shifted higher than that of its citation form.

The neutral tone *de* half of the graphs show that f_0 contours of different groups converge at the end, approximately in the f_0 range of $[-2.5, 0]$. The converging effect in Plastic Mandarin tends to be visually more obvious than that in Standard Mandarin. Now the difference curves are useful to check whether the f_0 differences at the end of *de* among different groups are statistically significant.

Figure 5.25 presents the estimated difference curves between each two groups of the disyllabic Plastic Mandarin tokens [ta tə]. In the normalised time domain between 0 and 1 ($t^* = [0, 1]$), the absolute value of estimated f_0 difference between all pairs is approaching 0 as the normalised time increases, and between phrases of T1-*de* and T2-*de*, T1-*de* and T3-*de*, T1-*de* and T4-*de*, and T2-*de* and T4-*de*, the associated confidence intervals include zero at the end (marked in red). The region that includes zero, though, is merely 6% of *de* in the latter two comparisons, namely, the f_0 difference between T1-*de* and T3-*de* phrases, and between T1-*de* and T4-*de* phrases. This suggests that the f_0 contours converge at the very end of *de* for these pairs of phrases. The f_0 differences between phrases of T2-*de* and T3-*de*, T3-*de* and T4-*de*, are still statistically significant at the end. Both Tone 3 and Tone 4 syllables

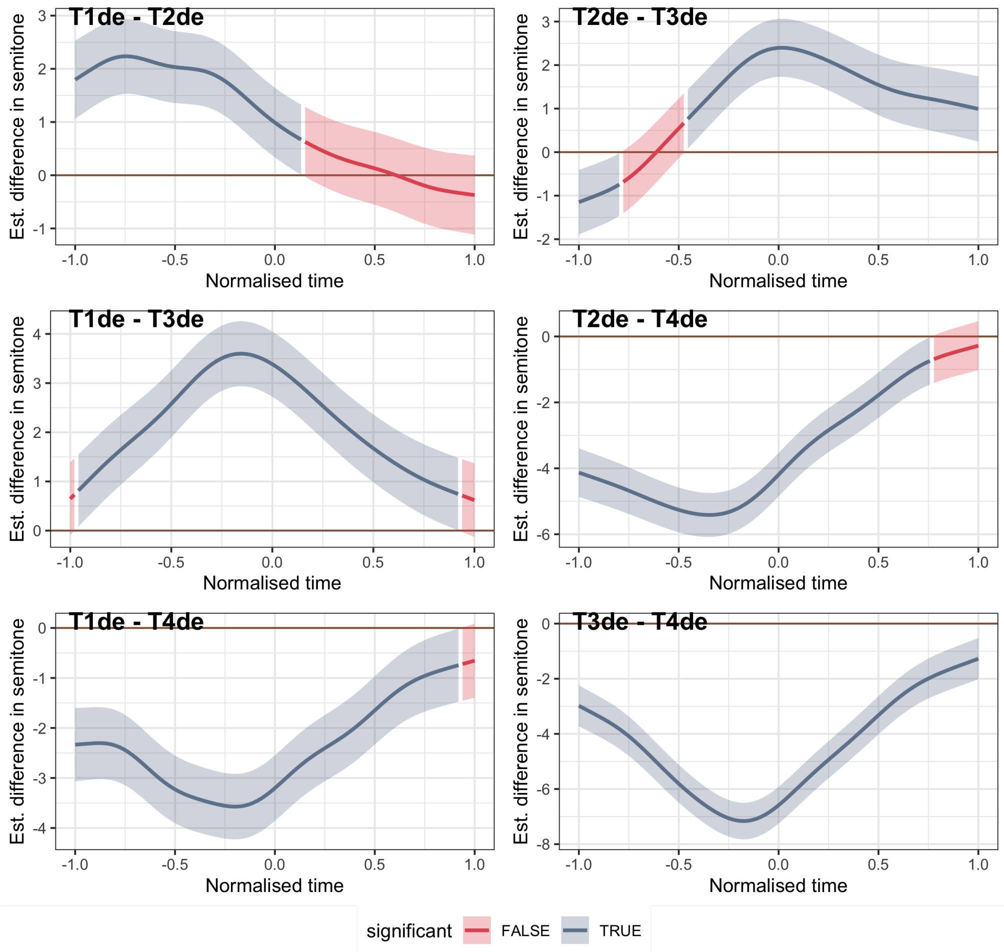
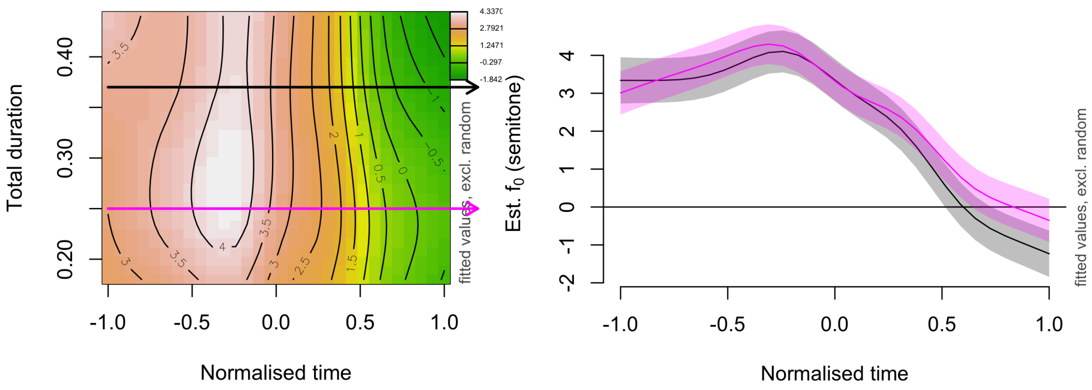


Figure 5.25: The estimated difference f_0 trajectories between each two groups of the disyllabic Plastic Mandarin tokens [ta tə] with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red.

have a large f_0 excursion at the end of the syllable, but in different directions. It may take longer time for these very different f_0 onsets of *de* to converge.

Figure 5.26 illustrates the duration effect for Plastic Mandarin tokens *dàde* 'big'. For the longer tokens, the f_0 drops more in the *de* part, especially in the normalised time range from 0.5 to 1 ($t^* = [0.5, 1]$). This is shown in the heatmap (a) that the black arrow, representing a longer token (370 ms), crosses more black solid lines than the pink arrow that represents a shorter token (250 ms), since black lines mark regions of equidistant f_0 change. The f_0 height at the end of *de* in these two tokens is directly reflected in (b), and the estimated f_0 difference is approximately one semitone.



(a) **(b)**

Figure 5.26: The effect of duration across the normalised time estimated by the GAMM for Plastic Mandarin disyllabic tokens of *dàde*. (a) the interaction surface between overall duration and normalised time; The 0.1 to 0.9 quantile of the whole range of total duration is shown. f_0 in semitone is denoted by colour grade, and regions of equidistant change (here, Δ 0.5 semitones) are denoted by black lines. (b) Two estimated f_0 contours with two different total duration: 0.25 (pink) and 0.37 (black). The corresponding duration values are marked by arrows in (a).

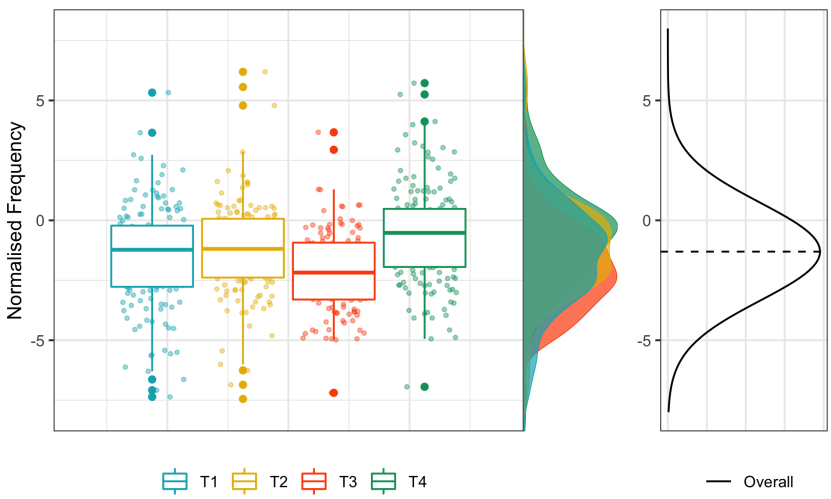


Figure 5.27: The distribution, kernel density estimation of ending f_0 of *de* in disyllabic Plastic Mandarin tokens and the fitted uniform Gaussian distribution. Two potential outliers with an absolute value of f_0 more than 8 were removed in the plot.

In order to estimate a potential converging f_0, Figure 5.27 shows the distribution, kernel density estimation of ending f_0 at the end of *de*. A uniform Gaussian distribution was fitted to the ending f_0 points, and it suggests an average ending f_0 of -1.30 semitones ($\sigma = 2.04$ semitones) across all groups.

5.4.3 Trisyllabic phrases with two consecutive neutral tone syllables

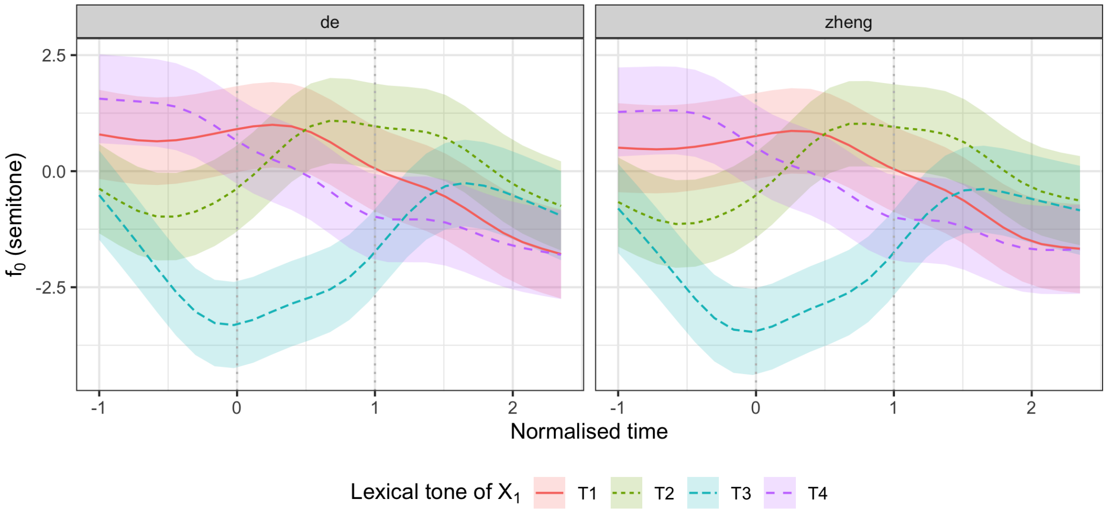
This section reports on the trisyllabic phrases composed of reduplicative kinship term followed by the plural particle *men*. Both the duplicate syllable X_2 and *men* are considered as neutral tone syllables. In this section, we investigate the influence of not only the preceding lexical tone but also the following tone on the f_0 contours of neutral tones.

Figure 5.28 presents the fitted normalised f_0 contours for each of the four phrases in the separate models of Standard Mandarin and Plastic Mandarin. Recall that we have normalised time range $[-1, 0]$ for f_0 contours of the lexical tone syllable X_1, $[0, 1]$ for the neutral tone syllable X_2, and $[1, 2.35]$ for the neutral tone syllable *men*.

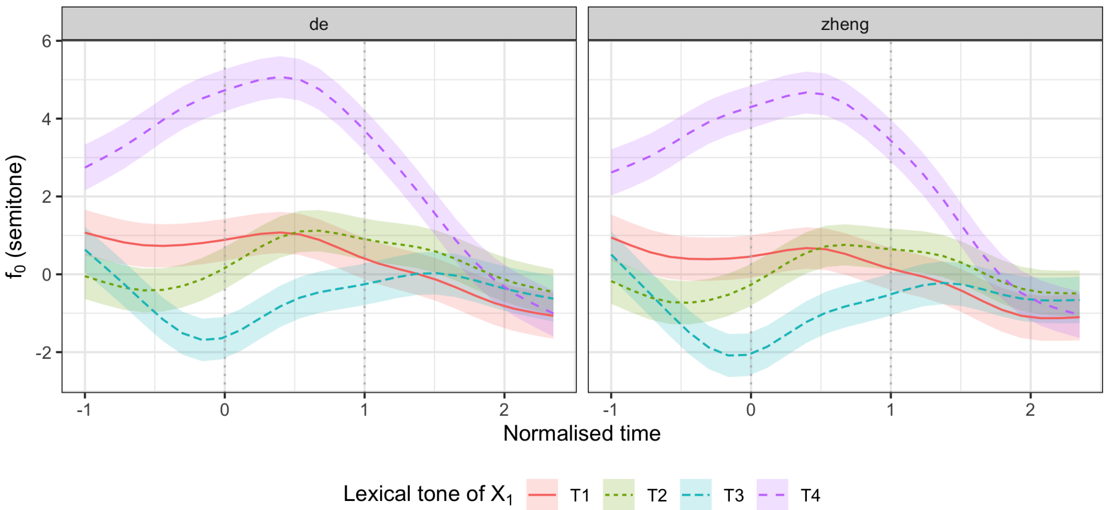
5.4.3.1 Influence of the following tone of *men*

Figure 5.28 separates the estimated f_0 contours according to the following syllable of *men*: neutral tone syllable *de* or Tone 4 syllable *zheng*. A Tone 4 syllable in both Mandarin varieties starts at a relative high f_0. Tone 4 syllable is, therefore, the best choice among other lexical tones to investigate whether the following tone influences the f_0 contour of a neutral tone syllable. In other words, given the its high starting f_0, Tone 4 may elicit evident anticipatory effect, or interpolation if the f_0 contour of a neutral tone syllable is interpolated between the preceding and following lexical tones without a pitch target.

By visually comparing the patterns between the *de* set and the *zheng* set in each Mandarin variety, the differences are not substantial. Three methods were used to test for significant differences between the two sets. First, we set up a nested model which excluded the parametric term and the difference smooth of the following syllable, and compared this to the original model through Chi-square



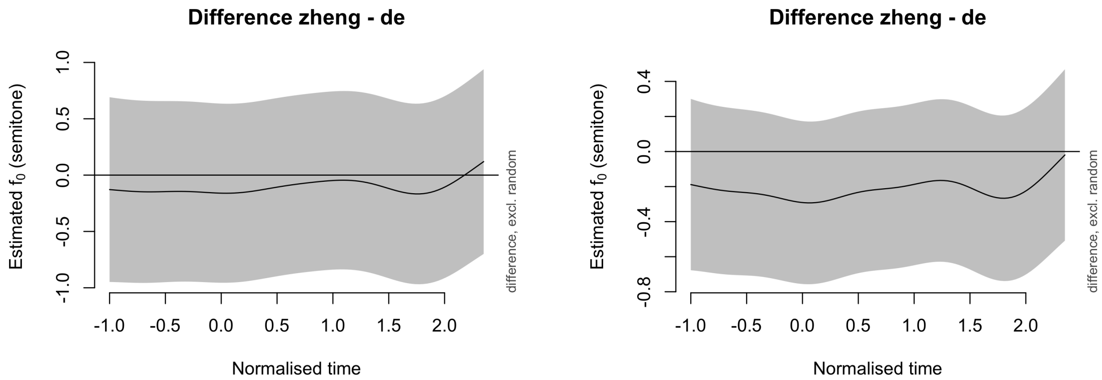
(a) Standard Mandarin



(b) Plastic Mandarin

Figure 5.28: Fitted f_0 contours of four $[X_1X_2\text{-}men]$ phrases, grouped by the following syllable of *men*, by two different groups of Mandarin speakers. The shades indicate 95% pointwise confidence intervals.

test of the Maximum Likelihood scores, using the `compareML()` command from the `itsadug` package. The model comparison in both varieties (Standard Mandarin: $\chi^2(3, 22) = 9.17, p < .001$; Plastic Mandarin: $\chi^2(3, 22) = 20.76, p < .001$) shows that the inclusion of the parametric term and the smooth difference term of the following syllable significantly improves the model fit. This suggests that the following syllable exerts some influence over the f_0 contours. Yet the distribution of the residuals of the models has heavy tails. Having remodelled the data with a



(a) Standard Mandarin **(b)** Plastic Mandarin

Figure 5.29: The estimated f_0 difference trajectory between *meimei-men* in the *de* set and the *zheng* set in each Mandarin variety, with the associated 95% pointwise confidence intervals. The entire area is not significantly different from 0.

scaled-t distribution, the model comparison approach is no longer available[6].

Then, we extracted summary statistics about this parametric and smooth term in the corrected model, using the `summary()` command in `mgcv` package, to check whether they are different from zero. In Standard Mandarin, the parametric term ($t(1) = -0.276, p = 0.78$) is not statistically significant, and the smooth term ($F(1, 5.76) = 2.21, p = 0.03*$) is statistically significant. Similarly in Plastic Mandarin, the parametric term ($t(1) = -0.985, p = 0.32$) is not statistically significant, and the smooth term ($F(1, 6.08) = 2.38, p = 0.02*$) is statistically significant. This suggests that the expected differences are not in f_0 height but in f_0 shape.

Finally, we plot the estimated difference curve between the two sets under each condition of X_1 tone category with the associated 95% pointwise confidence interval. Figure 5.29 demonstrates the difference curve for Tone 4 starting phrases *meimei-men* 'sisters'. We can see that in the normalised time range $[1.5, 2]$ in both varieties, the estimated f_0 difference (concave up shape in the difference curve) is slightly larger than that in $[1, 1.5]$, which reveals that the f_0 may fall slightly lower

[6] Since `bam()` needs the parameter 'discrete=TRUE' for extended distribution families with the AR1 component, which is only available with 'fREML' (fast restricted maximum likelihood estimation) smoothing parameter estimation method. But smoothing parameter 'ML' (maximum likelihood estimation) is needed for model comparison.

then rise at the end in the *zheng* set, so as to gain a positive velocity to reach the upcoming high tone. However, no regions in any of the plots excludes zero.

To conclude, we did not have sufficient statistical evidence supporting the significant differences in f_0 contours in the two sets. This reveals that the following Tone 4 has little impact on the f_0 contour of the trisyllabic phrases in both Mandarin varieties.

5.4.3.2 Influence of the preceding lexical tone

In both Mandarin varieties, the f_0 patterns of the duplicate syllable X_2 in the medial position of the trisyllabic phrases are different from those of *de* in the disyllabic phrases, although both X_2 and *de* are classified as neutral tone syllables. Nevertheless, the overall shape of the contours, grouped by the lexical tone of X_1, bears great resemblance to the shape of [X-*de*] phrases with the corresponding tone.

When we look at $[-1, 1]$ in 5.28(a), the normalised time range for X_1X_2 in Standard Mandarin, we can clearly find four canonical citation tone patterns: a a high falling pattern (Tone 4 SM), a mid-to-high rising pattern (Tone 2 SM), a fall-rise dipping pattern (Tone 3 SM), and a (relatively) high level pattern (Tone 1 SM). The high falling pattern here does not fall as much as the citation Tone 4. Similarly in 5.28(b), the tone patterns in $[-1, 1]$ mostly show the prototypical citation tone patterns of Plastic Mandarin: a high rising pattern (Tone 4 PM), a mid level pattern (Tone 1 PM), and a mid rising pattern (Tone 2 PM). There is also a fall-rise dipping pattern, which is different from the prototypical low-falling Tone 3 in Plastic Mandarin, but similar to Tone 3 in Standard Mandarin. In both Mandarin varieties, f_0 in Tone 3 starting phrases *nainai-men* ‘grannies’ falls in the first syllable, then rises in the second syllable, although it falls to a lower f_0 in Standard Mandarin than in Plastic Mandarin. In fact, except for the Tone 4 starting phrases *meimei-men*, the estimated f_0 contours of the other phrases in Plastic Mandarin resemble those of the corresponding phrases in Standard Mandarin.

The above observations suggest that the reduplication tends to provide increased duration to realise the f_0 contour of the lexical tone of the base syllable X_1.

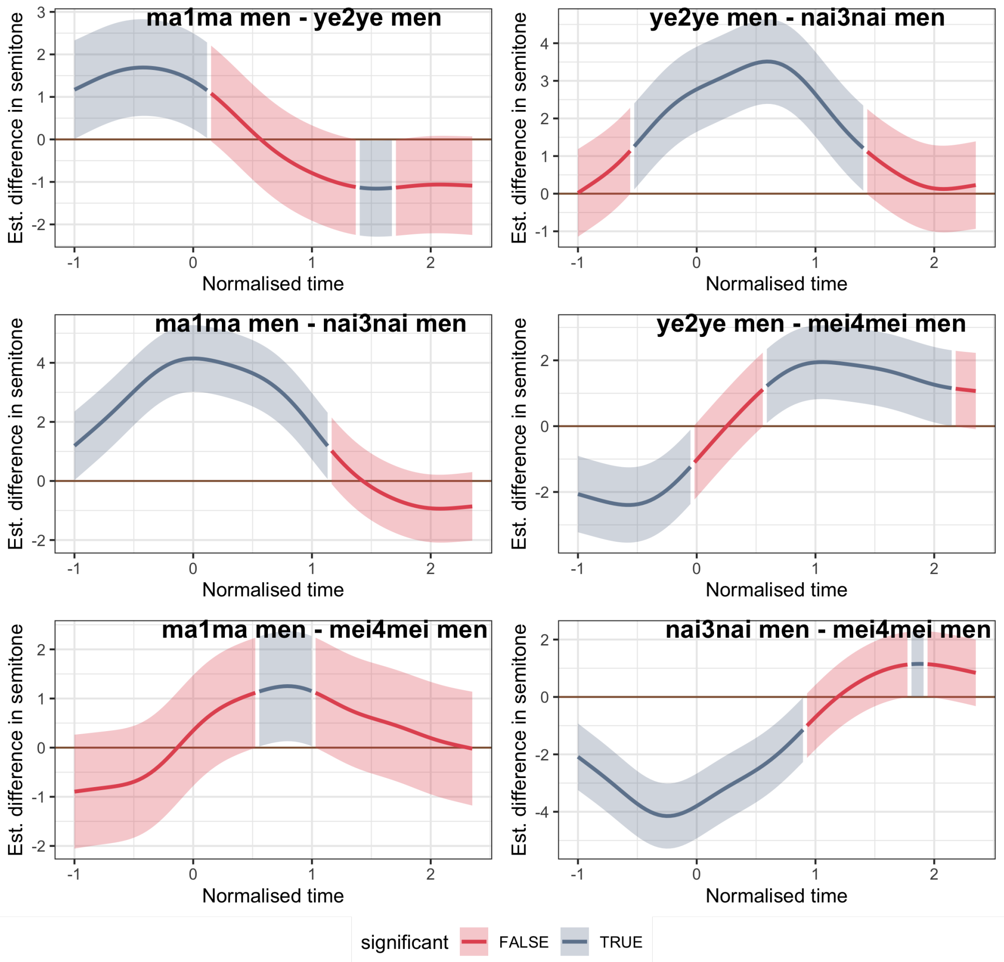


Figure 5.30: The estimated difference f_0 trajectories between each two groups of the trisyllabic Standard Mandarin tokens in the *de* set, with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red. The number in the pinyin transcription indicates X_1 tone category.

In other words, the lexical tone of X_1 has much influence on the f_0 contour of the following neutral tone syllable X_2, and such influence continues into the second neutral tone syllable *men*.

Now we move on to look at the variation in f_0 contours of *men*. Figures 5.30 and 5.31 present the estimated difference curves between each two groups of the trisyllabic tokens in Standard Mandarin and Plastic Mandarin respectively. In both figures, all the difference curves with their associated 95% pointwise confidence intervals include zero towards the end. For Standard Mandarin, in the last 5% of all the difference curves, the normalised time domain $[2.18, 2.35]$, the differences in f_0 between any two groups are statistically not significant. For Plastic Mandarin,

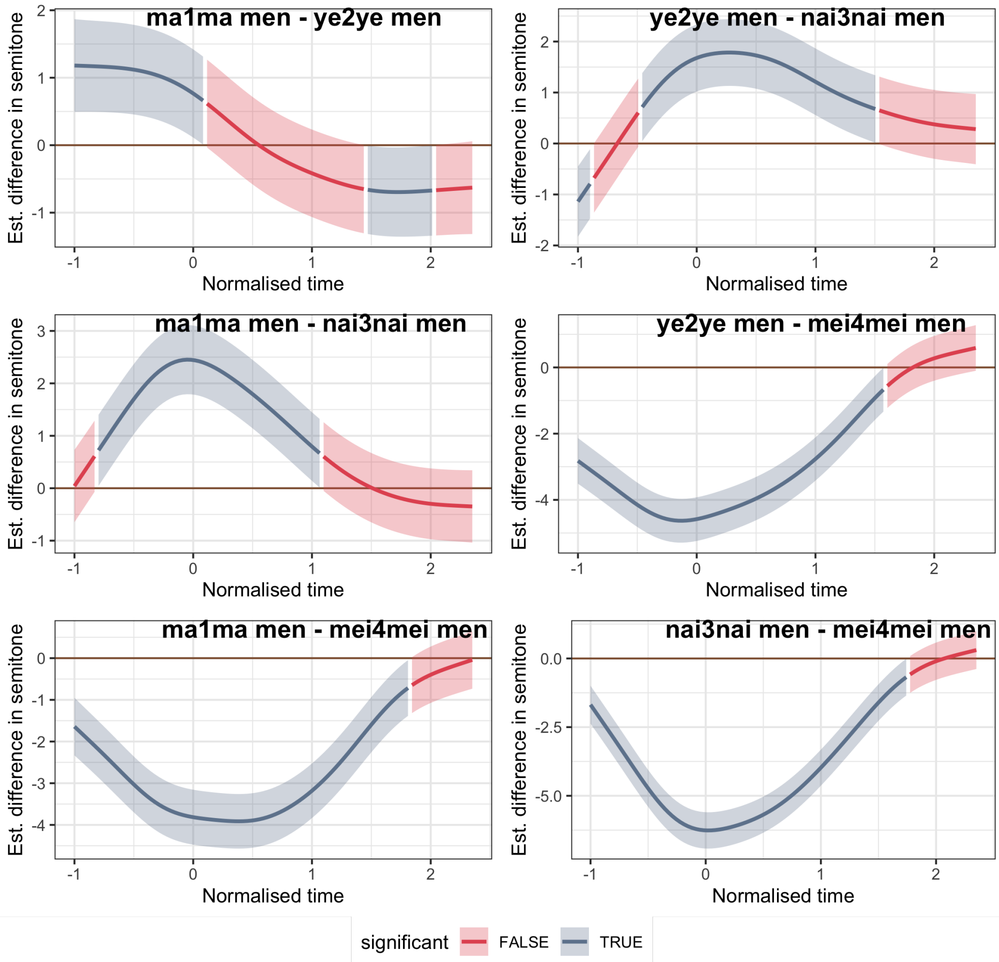


Figure 5.31: The estimated difference f_0 trajectories between each two groups of the trisyllabic Standard Mandarin tokens in the *de* set, with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red. The number in the pinyin transcription indicates X_1 tone category.

in the last approximately 9% of all difference curves, the normalised time domain $[2.05, 2.35]$, the differences in f_0 between any two groups are statistically not significant. This suggests that different f_0 contours converge towards the end of the neutral tone syllable *men* in both Mandarin varieties.

Figures 5.32 and 5.33 show the distribution, kernel density estimation of f_0 at the end of *men* in Standard Mandarin and Plastic Mandarin respectively. A uniform Gaussian distribution was fitted to the ending f_0 points in each dataset. The Gaussian model suggests an average ending f_0 of -1.24 semitones ($\sigma = 2.58$ semitones) across all $[X_1X_2men]$ phrases in Standard Mandarin, and an average ending f_0 of -0.67 semitones ($\sigma = 2.00$ semitones) in Plastic Mandarin.

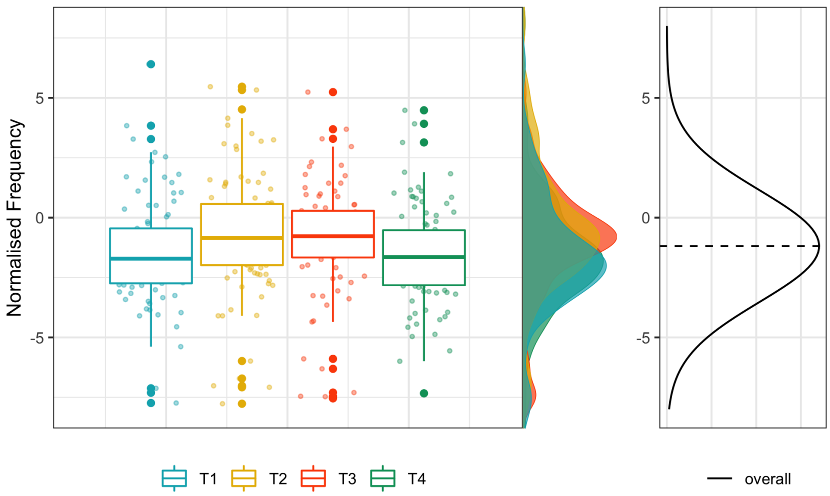


Figure 5.32: The distribution, kernel density estimation of ending f_0 of *men* in trisyllabic Standard Mandarin tokens and the fitted Gaussian distribution. Seven potential outliers with an absolute value of f_0 more than 8 st were removed in the plot.

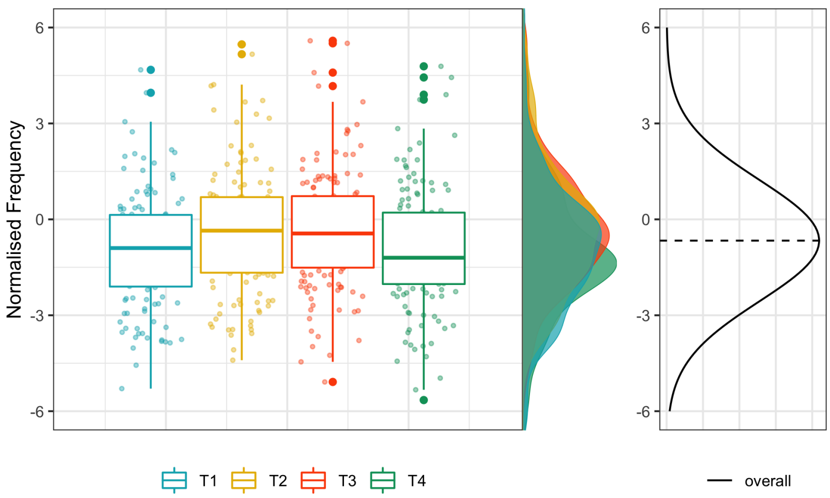


Figure 5.33: The distribution, kernel density estimation of ending f_0 of *men* in trisyllabic Plastic Mandarin tokens and the fitted Gaussian distribution. Five potential outliers with an absolute f_0 value more than 6 st were removed in the plot.

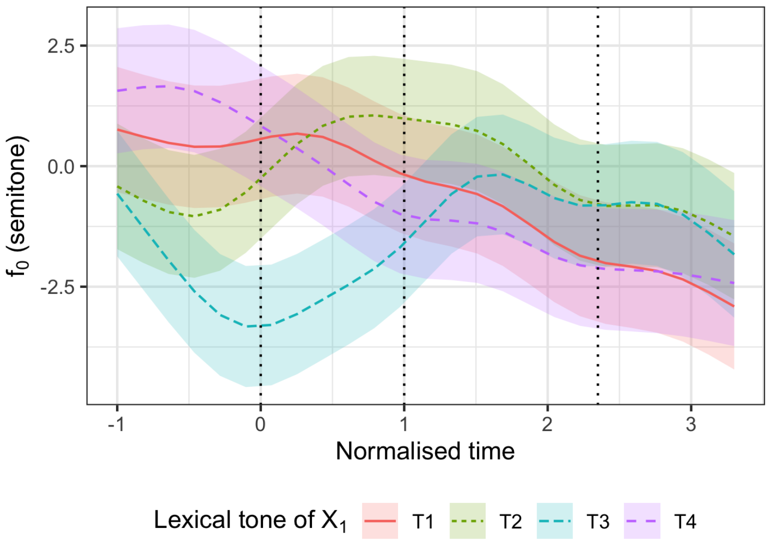
5.4.4 Quadrisyllabic phrases with three consecutive neutral tone syllables

This section continues to examine the subset of $[X_1X_2men]$ phrases, those who are followed by *de*, $[X_1X_2men\ de]$, in which three consecutive neutral tone syllables follow a lexical tone syllable X_1. From the previous section §5.4.3, we have learnt that the four different f_0 contours converge at the end of *men* in both Mandarin variety. This means that the starting f_0 of different contours for the last neutral tone syllable *de* is similar. Will the f_0 contours of *de* be flat if the contours have already converged at a common pitch target?

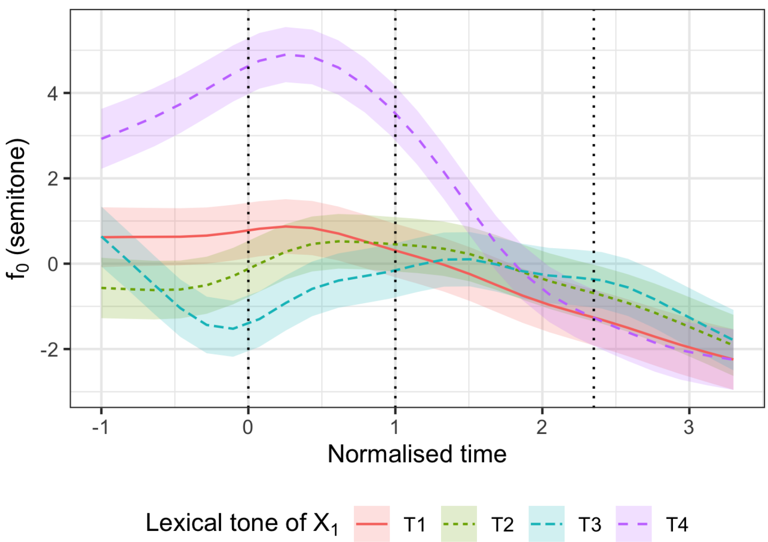
Figure 5.34 shows the model predictions for the four quadrisyllabic phrases in each Mandarin variety. In each sub-figure, the shape of all f_0 contours of *de*, in the normalised time domain from 2.35 (black dotted reference line) to 3.3 ($t^* = [2.35, 3.3]$), is similar – the f_0 contours continue to fall, and their confidence bands overlap with one another. Such a homogeneous f_0 pattern of *de* forms a contrast with the diverse f_0 patterns of the same syllable *de* when it is the second syllable in the disyllabic phrases. This suggests that, in both Mandarin varieties, the lexical tone of X_1 has little impact on the f_0 contour of the third neutral tone syllable *de*, and that the pitch target of *de* is low.

In order to estimate the potential low pitch target, we examine the probabilistic distribution of the final f_0 point of *de*. Figure 5.35 portrays the distribution, kernel density estimation of f_0 at the end of *de* in Standard Mandarin and Plastic Mandarin respectively. A uniform Gaussian distribution was fitted to the ending f_0 points in each dataset. The Gaussian model suggests an average ending f_0 of -2.13 semitones ($\sigma = 2.49$ semitones) across all $[X_1X_2men]$ phrases in Standard Mandarin, and an average ending f_0 of -1.97 semitones ($\sigma = 2.14$ semitones) in Plastic Mandarin.

Recall that in the disyllabic phrases of Plastic Mandarin, the f_0 contours of *de* have already shown a clear converging trend (see Figure 5.20) and the mean of ending f_0 is about -1.30. Here the mean of ending f_0 of *de* is lower, which reveals that the consecutive neutral tones allow ample time to reach a lower pitch target. This matches the duration effect shown in Figure 5.22 and discussed in §5.4.2.

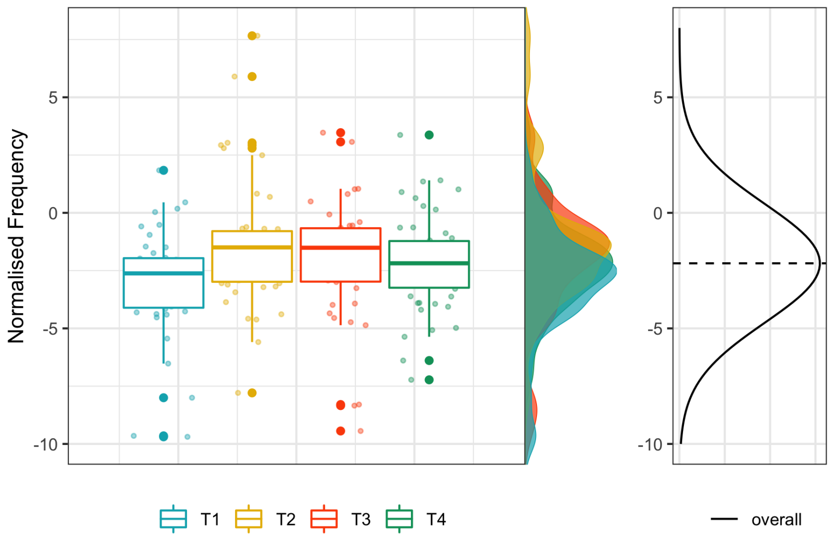


(a) Standard Mandarin

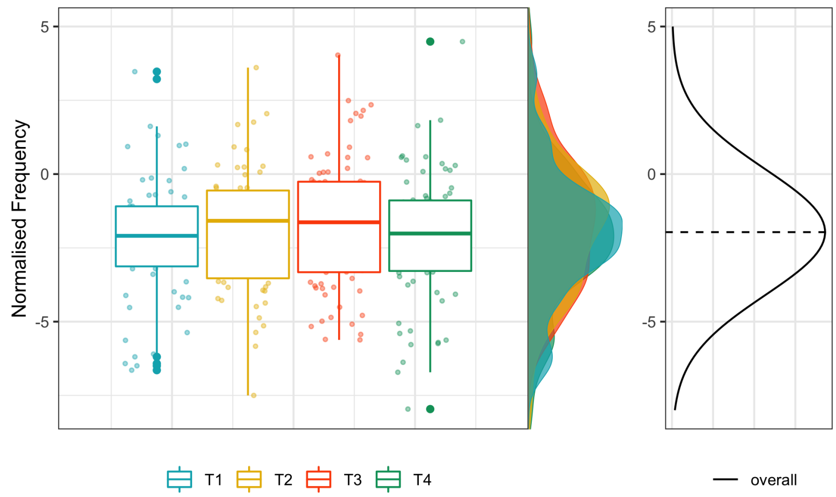


(b) Plastic Mandarin

Figure 5.34: Fitted f_0 contours of four [X_1X_2-*men de*] phrases by two different groups of Mandarin speakers. The shades indicate 95% pointwise confidence intervals. Black dotted vertical reference lines indicate the syllable boundaries.



(a) Standard Mandarin



(b) Plastic Mandarin

Figure 5.35: The distribution, kernel density estimation of ending f_0 of *de* in quadrisyllabic tokens and the fitted Gaussian distribution in the two Mandarin varieties. In (a), one potential outlier with f_0 less than -10 st was removed in the plot.

5.4.5 Comparison of Plastic Mandarin and Standard Mandarin

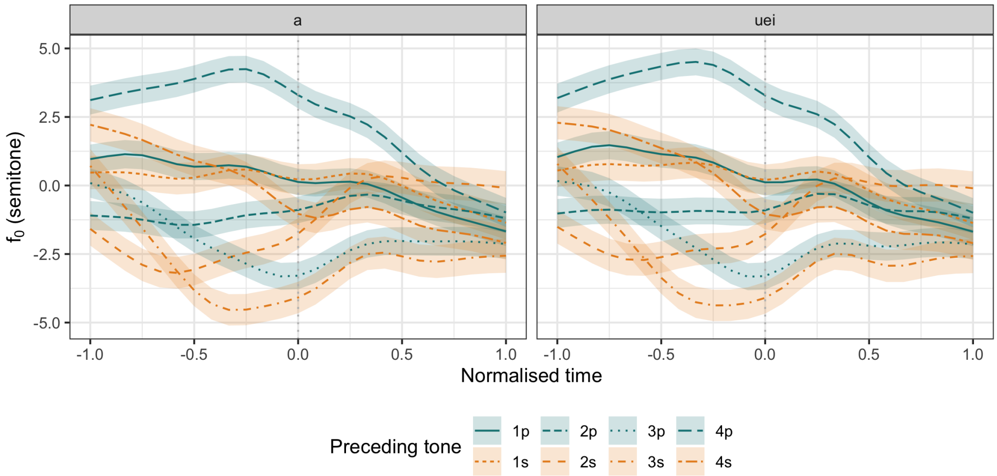
The Set 2 models with the combined data of Standard Mandarin and Plastic Mandarin explore the dialectal variation in the realisation of f_0 contours in the disyllabic, trisyllabic, and quadrisyllabic phrases containing neutral tone syllables. Previous observations have shown that the two Mandarin varieties are similar in the following aspects, despite the different lexical tones:

1) The f_0 contour of a neutral tone syllable is greatly influenced by the preceding tone, but not by the following tone.
2) The different f_0 contours of neutral tone syllables *men* and *de* converge towards a low pitch, and consecutive neutral tones allow ample time to reach a lower target.
3) The f_0 patterns of the duplicate syllable X_2 are different from those of *men* and *de*.

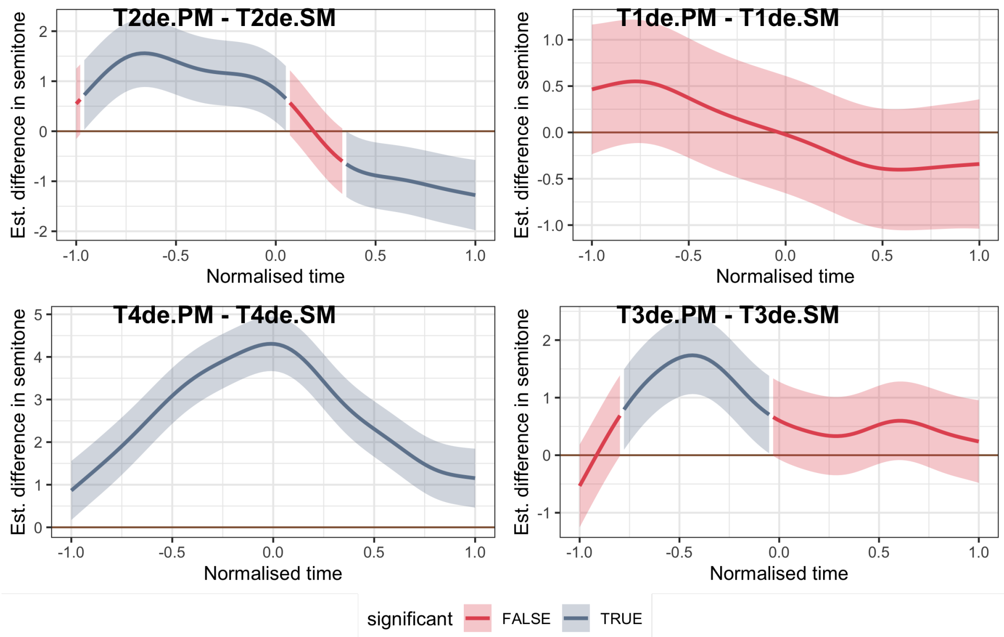
This section mainly investigates whether the differences between Standard and Plastic Mandarin in the observed converging f_0 levels at the end of neutral tone syllables are statistically significant. In other words, is there a similar pitch target for neutral tone syllables in the two Mandarin varieties?

Figures 5.36, 5.37, and 5.38 present the estimated f_0 trajectories of four different groups of disyllabic, trisyllabic, and quadrisyllabic phrases in both Mandarin varieties respectively, and the estimated f_0 difference trajectories between each pair of the same phrase in the two Mandarin varieties. In all three figures, at the end of normalised tokens where f_0 contours converge, Plastic Mandarin contours have much overlap with Standard Mandarin contours, especially in 5.38 (a).

The difference trajectories in Figure 5.36 (b) suggest that the f_0 contours of T1-*de* phrases in the two Mandarin varieties are not statistically significantly different. The f_0 differences between Standard Mandarin and Plastic Mandarin in the *de* part of the f_0 contours of T3-*de* phrases are also not statistically significant, although Tone 3 in Standard Mandarin has a steeper fall and earlier trough. The f_0 differences between the two Mandarin varieties in the Tone 2 and Tone 4 pairs are largely statistically significant. The starting f_0 heights of *de* ($t^* = 0$) in the Tone 1 and Tone 3 pairs are similar (in red), while the starting f_0 of *de* in Plastic Mandarin in the Tone 4 pair is about 4 semitones higher than that in Standard Mandarin. There may not be enough time for a much higher starting f_0 to descend to the same low level, despite a higher decline rate. While the starting f_0 differences of *de* in the Tone 2 pair are small, the f_0 velocity in that moment is different, since Tone 2 in Standard Mandarin has a larger slope of rising than Tone 2 in Plastic Mandarin.

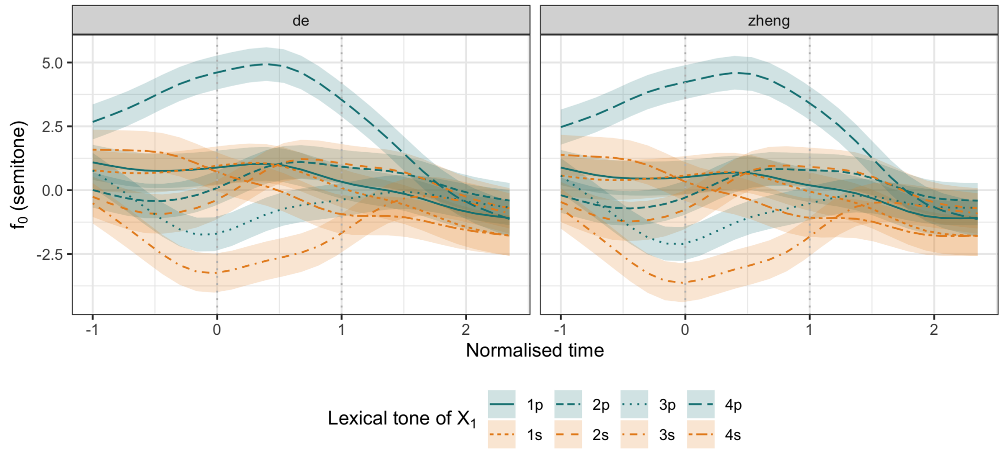


(a)

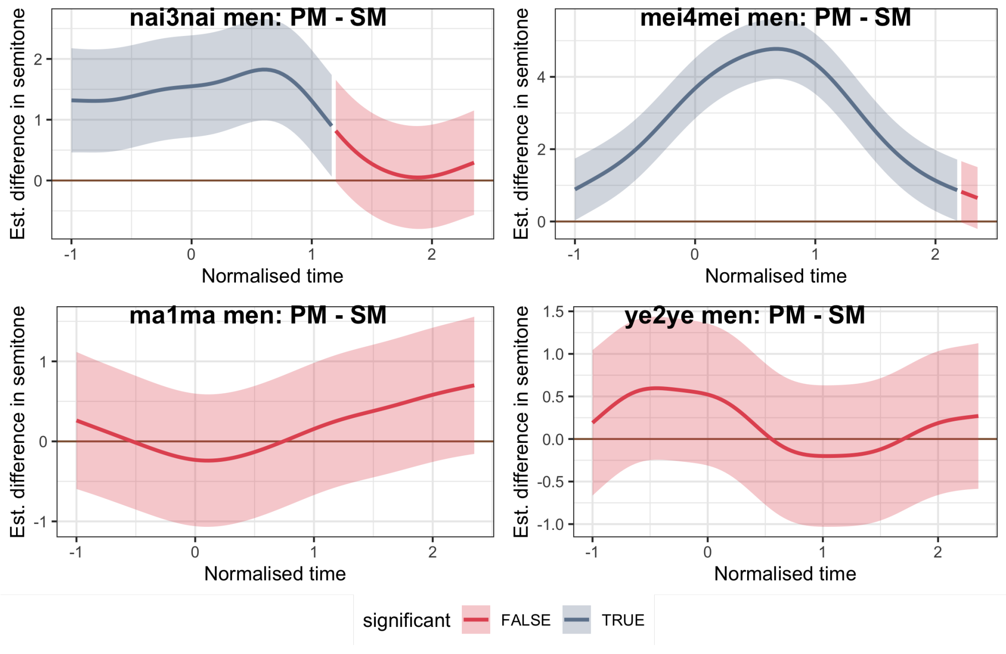


(b)

Figure 5.36: (a) Fitted f_0 trajectories of four different groups of disyllabic phrases in both Mandarin varieties. For the preceding tone of *de*, numbers indicate tone categories, 'p' indicates Plastic Mandarin, and 's' indicates Standard Mandarin. (b) The estimated f_0 difference trajectories between each by-variety pair of *da-de* phrases, with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red. The number in the pinyin transcription indicates the preceding tone of *de*.

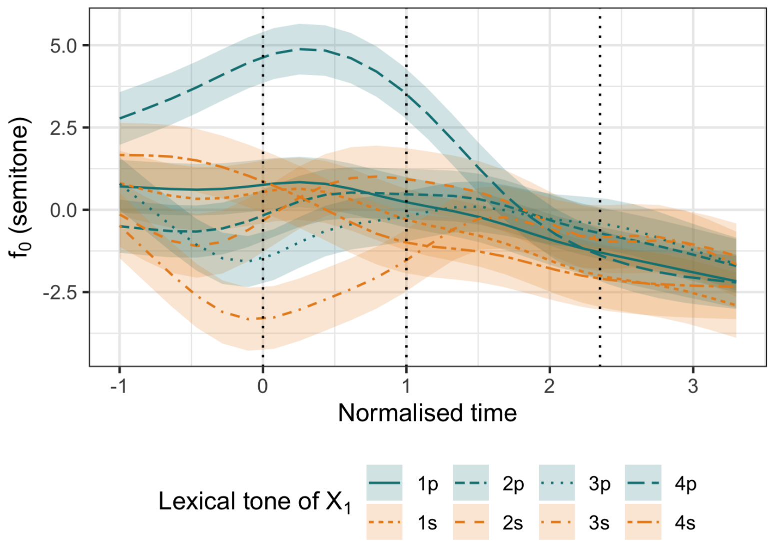


(a)

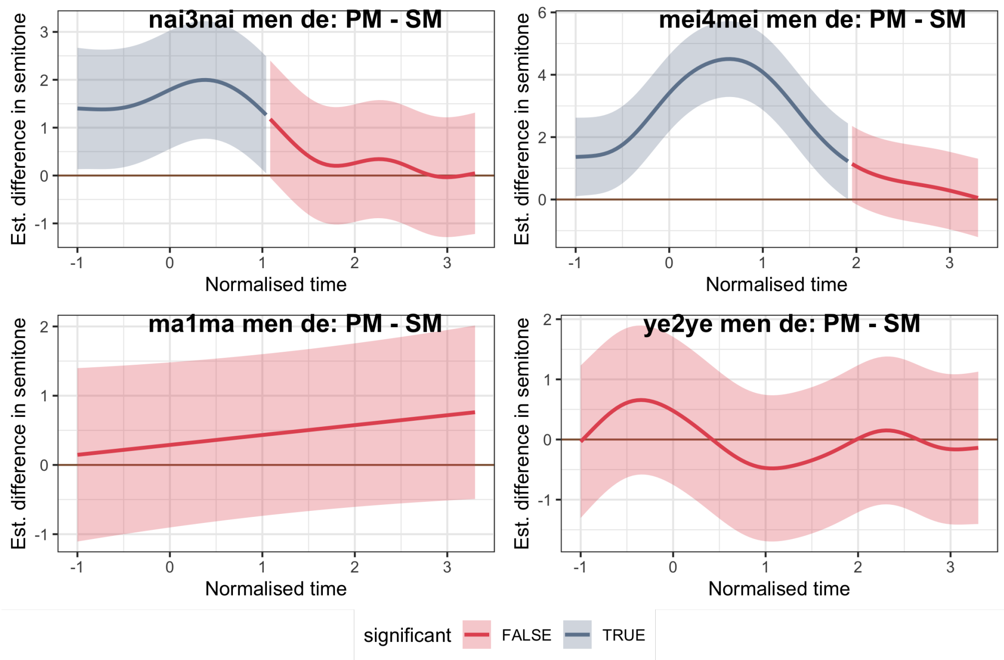


(b)

Figure 5.37: (a) Fitted f_0 trajectories of four different groups of trisyllabic phrases in both Mandarin varieties. For the lexical tone of X_1, numbers indicate tone categories, 'p' indicates Plastic Mandarin, and 's' indicates Standard Mandarin. (b) The estimated f_0 difference trajectories between each by-variety pair of $[X_1X_2 men\ (de)]$ phrases, with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red. The number in the pinyin transcription indicates the tone category of X_1.



(a)



(b)

Figure 5.38: (a) Fitted f_0 trajectories of four different groups of quadrisyllabic phrases in both Mandarin varieties (duration: 540 ms). For the lexical tone of X_1, numbers indicate tone categories, 'p' indicates Plastic Mandarin, and 's' indicates Standard Mandarin. (b) The estimated f_0 difference trajectories between each by-variety pair of $[X_1 X_2 \textit{men de}]$ phrases, with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red. The number in the pinyin transcription indicates the tone category of X_1.

For trisyllabic phrases, Figure 5.37 (b) suggests that the f_0 contours of both *mama-men* and *yeye-men* phrases in Plastic Mandarin are not statistically significantly different from the corresponding phrases in Standard Mandarin. The f_0 differences between Standard Mandarin and Plastic Mandarin in phrases *nainai-men* are not statistically significant in the majority portion of the *men* syllable. Given the drastically different tone patterns of Tone 4, the f_0 differences between Standard Mandarin and Plastic Mandarin in phrases *meimei-men* are mostly statistically significant, and become not significant only at the very end of *men*.

In Figure 5.38 (b) of quadrisyllabic phrases, the percentage of areas marked in red where the f_0 differences between Standard Mandarin and Plastic Mandarin are not statistically significant has increased in phrases *nainai-men de* and *meimei-men de*. The f_0 contours of both *mama-men de* and *yeye-men de* phrases in Plastic Mandarin are, still, not statistically significantly different from the corresponding phrases in Standard Mandarin in their entirety. For all four phrases, the f_0 contours of the *de* part ($t^* = [2.35, 3.3]$) between the two Mandarin varieties do not differ statistically significantly. This suggests that there is a similar low pitch target for neutral tone syllables in Standard Mandarin and Plastic Mandarin.

5.5 Discussion

5.5.1 Approaching a low pitch target

This chapter provides compelling acoustic evidence that the f_0 contour of neutral tone syllables is approaching a low pitch target in both Mandarin varieties. From the model prediction plots of *de* in disyllabic phrases, *men* in trisyllabic phrases, and *men de* in quadrisyllabic phrases in Plastic Mandarin, we can clearly see that various f_0 contours converge at the end of neutral tone sequences towards a pitch level that is lower than the speaker mean, with overlapping confidence intervals. Likewise, the converging effect of f_0 contours of different groups of phrases is strongly present in trisyllabic phrases and quadrisyllabic phrases in Standard Mandarin. It is, though, less obvious in the [X-*de*] disyllabic phrases in Standard Mandarin. In fact, at the

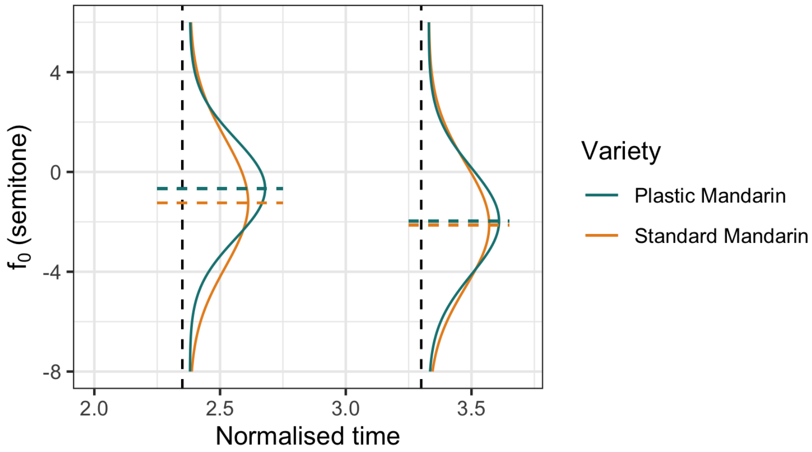


Figure 5.39: The fitted Gaussian distribution of ending f_0 of trisyllabic ($t^* = 2.35$) and quadrisyllabic ($t^* = 3.3$) tokens in the two Mandarin varieties.

end of *de* in the disyllabic phrases, there are some groups of f_0 contours between which the f_0 difference is still statistically significant in both Mandarin varieties.

Duration of a neutral tone sequence tends to play a role in the approximation of a pitch target. The findings suggest that a longer neutral tone sequence allows ample time to approach a lower target. When there is only one neutral tone syllable, the longer token tends to have a lower ending f_0. The short duration of *de* in Standard Mandarin may explain the large variance in the distribution of f_0 at the end of *de* in a disyllabic token. As reported in §5.3.2.1, *de* is on average about 67% of the duration of the preceding lexical tone syllable in Standard Mandarin, and about 77% in Plastic Mandarin.

Figure 5.39 recaps the fitted Gaussian distribution of ending f_0 in the trisyllabic and quadrisyllabic tokens in the two Mandarin varieties. At the end of the quadrisyllabic tokens, the two Gaussian distributions of ending f_0 have a large overlap. The GAMMs analysis reported in §5.4.5 shows that f_0 difference between Standard Mandarin and Plastic Mandarin throughout the entire *de* in the quadrisyllabic tokens is not statistically significant. This suggests that the low pitch target associated with neutral tone sequences in Plastic Mandarin is similar to that in Standard Mandarin.

5.5.2 Interpretations of the low pitch target

While there are systematic variations in the citation form of lexical tones between Standard Mandarin and Plastic Mandarin (see Chapter 3), the pitch target associated with neutral tones, nevertheless, remains stable and constant in this synchronic variation. Where does the target come from?

Our data show that the f_0 contours of neutral tone sequences [X_1X_2-*men*] are greatly influenced by the preceding lexical tone, but hardly influenced by the following lexical tone. The converging effects of neutral tone f_0 contours irrespective of the preceding and following tones rule out the claims that neutral tone is merely an interpolation between its preceding and following tone or a complete spreading from its preceding tone.

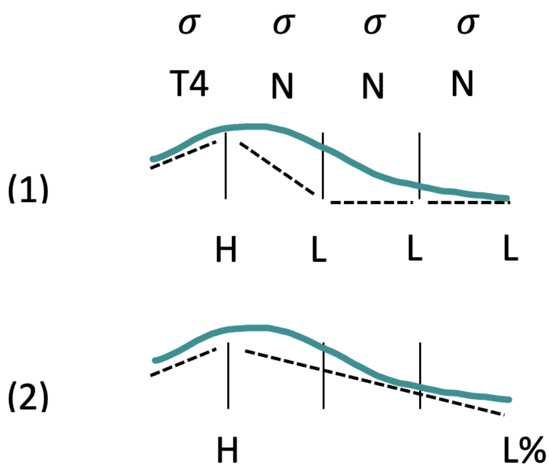


Figure 5.40: The schematic illustration of hypothetical pitch targets for a neutral tone sequence, high rising Tone 4 in Plastic Mandarin followed by three neutral tone syllables. Vertical bars demarcate syllables. Letters H and L indicate the underlying tone target of a syllable. L% represents a low tone target attached to the end of this phrase. The dotted black line is the predicted pitch contour given the underlying tones and the teal solid line is the actual pitch contour.

5.5.2.1 The hypothesis of an underlying low tone

A natural association of a low pitch target is that neutral tone may be underlyingly a low tone, which was proposed by H. Lin (2006). There are two main issues in this account: first, neutral tone does not behave the same way as the other low tone, for example, Tone 3 in Standard Mandarin and Plastic Mandarin. While the low pitch target is usually approached by the end of a Tone 3 syllable, there is a large f_0 variance at the end of a neutral tone. Second, the investigation into sequences of neutral tone suggests that not necessarily every neutral tone is associated with a low pitch target. If all neutral tones were L tones, one would expect the pitch contour of a neutral tone flanked by neutral tones on both sides to exhibit a relatively flat shape with a low pitch, as illustrated schematically in Figure 5.40 (1). Yet our data displayed a falling pattern in all three consecutive neutral tone syllables, despite the variability in the slopes.

5.5.2.2 Attracting the boundary low tone

Another more promising account is that neutral tone, always associated with unstressed syllables, does not have any underlying tone, but attracts a boundary low tone. Sun and Shih (2021) found that anticipatory tonal coarticulation became weak when crossing a strong boundary such as the subject-predicate boundary. Q. Li and Chen (2019) also suggested that the anticipatory effect of the following tone can be blocked by a higher-level boundary. The weak effect of the following lexical tone in our data may indicate a strong boundary effect.

Two examples of sentences containing $[X_1X_2\text{-}menX_4]$ phrases are given in (13), in which the focused words are marked in bold. We can see that in (13a), *mèimei-men* is the noun phrase serving as the subject of the sentence, and precedes the predicate of the sentence. Such subject-predicate boundary was viewed as a high-level or strong boundary in Q. Li and Chen (2019) and Sun and Shih (2021), which may explain the weak effect from the following Tone 4 syllable *zheng*.

(13) a. *meimei-men zheng tán-zhe* ***jítā***.
sister-PL PROG play-DUR guitar
'Sisters are playing guitars.'

b. *meimei-men de* ***zhúzi*** *gāo-le*.
sister-PL GEN bamboo (grow) tall-PERF
'Sisters' bamboo has grown taller.'

Strong boundaries are often defined in prosodic domains, though often synchronised with syntactic structures. In Fitzpatrick-Cole (1996)'s analysis of Bengali, the right edge of focus and the left edge of verbs are both considered as strong phonological phrase boundaries. In intonation phonology, the edges of various prosodic constituents are the docking sites for boundary tones (Pierrehumbert, 1980). For example, Hayes and Lahiri (1991) argue that boundary tones associate to domain boundaries of the Prosodic Hierarchy (Selkirk, 1986; Nespor, 1986), in particular the Phonological Phrase and the Intonational Phrase, in Bengali. Takahashi (2019) argues that a boundary low tone is assigned at the right edge of a phonological word in Shanghai tone sandhi.

A potential prosodic grouping is illustrated in (14) using brackets, and the boundary tone $_L$ is attached to right edge of the outermost prosodic constituent in a recursive structure. The prosodic leanings of neutral tone syllables are leftwards, similar to the encliticising functional words in Germanic languages (Lahiri & Plank, 2022). The location of a neutral tone syllable always coincides with the right edge of a prosodic constituent, and a neutral tone syllable can occur at the right edge of various layers within the prosodic hierarchy. While in (14a) the noun phrase *mèimei-men* precedes a predicate, in (14b) *mèimei-men-de* is the given information preceding a narrow focus (the keywords being guessed). It is likely that both these conditions trigger a strong boundary effect so that the underspecified neutral tone attracts a boundary low tone, as illustrated in Figure 5.40 (2).

(14) a. ((*mèimei*) *men*)$_L$ (*zhèng*)
sister PL PROG

b. (((*mèimei*) *men*) *de*)$_L$
sister PL GEN

One advantage of the boundary tone account is that it differentiates neutral tone from other lexical tones in acquiring a surface pitch target. Boundary tone is a post-lexical process. The distinction between lexical tones and post-lexical processes can conveniently explain their different behaviours in the synchronic variation of Standard Mandarin and Plastic Mandarin—neutral tone does not conform to the systematic variations of tone shape (target) observed in the other lexical tones (recall Chapter 3). If neutral tone is inherently the same as the other lexical tones, how to explain its deviation from the predominant pattern of tonal variation?

Previous literature have highlighted substantial divergences between the neutral tones in Changsha and Standard Mandarin (recall §2.5.1.3). The pitch pattern of neutral tones in Plastic Mandarin in this study did not resemble the reported pitch pattern of neutral tones in Changsha. That the pitch target of neutral tone across the two Mandarin varieties remains invariant may signal that the post-lexical tonal phenomenon associated with neutral tone in Plastic Mandarin is a direct transfer from Standard Mandarin and that the post-lexical processes tend to be more stable than lexical tones in this Mandarin variation.

Another advantage of the boundary account is that it enables neutral tone syllables to undergo different post-lexical processes. Neutral tone syllables such as interrogative particles at the end of a question are most likely to have a high tone H% instead of a low tone in Mandarin varieties.

The boundary account is, nevertheless, not conclusive. Firstly, the phonological account does not explain the varying converging pitch levels at the end of a different number of neutral tone syllables, although all of them are below the speaker mean. The ending f_0 of a single neutral tone in disyllabic phrases was also subject to higher variability than that of a neutral tone sequence. Secondly, despite ample duration in a neutral tone sequence, the average ending pitch of a low boundary tone L% might not be as low as that of a single (underlying) L tone. Thirdly, more data are needed to investigate the effect of different boundary types on neutral tone, controlling for pragmatic contexts such as focus and newness. Fourthly, this study only concerns the potential boundary effects on neutral tone syllables and what happens to full lexical tones at boundaries is yet to be explored in future research.

5.5.2.3 Pitch-related prosodic preplanning

The slow rate of approximating the low pitch target in neutral tones was interpreted as weak articulatory strength (Y. Chen & Xu, 2006). In other words, a low neutral tone tends to be implemented in a 'softer' or 'undershooting' manner compared to a low lexical tone, which may be partially attributed to the short syllable duration of many neutral tone syllables.

The varying converging pitch levels at the end of neutral tone sequences in different lengths may be accounted for in speech production planning, where pitch shifting has been presented as acoustic evidence for pitch-related preplanning in speech production (A. C. H. Chen & Tseng, 2019).

The phonetic form of a word in an utterance is greatly influenced by the prosodic structure and prosodic prominence (Keating & Shattuck-Hufnagel, 2002). Prosodic encodings typically included f_0 heights and duration at edges of prosodic units. Yuan and Liberman (2014) showed strong correlation between the f_0 declination slopes and the length of inter-pause units in English and Mandarin broadcast news speech, that is, longer units tended to have less steep f_0 slopes. No correlation between the length of the inter-pause unit and the final f_0 height was found. Liberman and Pierrehumbert (1984) suggested that utterance-final f_0 height asymptoted to a speaker-based baseline representing characteristics of a speaker's voice. Asu et al. (2016), on contrary, observed that longer intonational phrases in spontaneous Estonian speech exhibited a tendency of lower phrase-final f_0 heights in addition to less steep f_0 slopes, suggesting a larger pitch range may be pre-arranged in longer prosodic units.

In our data, longer neutral tone sequences tended to have a lower final f_0. In phrases with three consecutive neutral tones following a high tone (i.e. T1 in Standard Mandarin and T4 in Plastic Mandarin), the declination rate was faster in the first two neutral tone segment. In the language game setting with semi-spontaneous speech production, the online message planning is likely to be more fragmented compared to read speech and the final f_0 height may be more dependent on the duration of a prosodic unit (A. C. H. Chen & Tseng, 2019). With a longer prosodic unit speakers may be more likely to drop to their pitch low-bound baseline.

The crosslinguistic differences between Mandarin and English in Yuan and Liberman (2014) suggested that prosodic encoding in speech production is linguistically meaningful. This study revealed that pitch-related prosodic encoding including pre-boundary f_0 lowering is similar between the two Mandarin varieties.

5.5.3 Implications for neutral tone typology

The f_0 patterns of the duplicate syllables X_2 tend to differ from the low-pitch target approaching patterns identified for *de* and *men*, reported in §5.4.3. This might be the case that reduplication has a different intonation mechanism and that a copied syllable X_2 has an underlying tone, which is the same as the original lexical tone of X_1. Fitzpatrick-Cole (1996) showed that in Bengali the entire reduplicated phrasal structure including the base phrase and the copy shared one expected contour such as the tune for a focus nucleus (L*H_ϕ), and thereby argued that the entire structure was in a single Phonological Phrase. In a similar vein, the entire reduplicated structure [X_1X_2] here tends to share the lexical tone pattern of X_1 instead of having two copies of a lexical tone. In this case, the so-called neutral tone of X_2 would be some tone sandhi form operated within the domain of reduplication.

Since [X_1X_2] in our data is embedded in trisyllabic phrases [X_1X_2-*men*], the reduplication effect is confounded with the potential influence of the following syllable *men*. It would be ideal to have disyllabic reduplication data [X_1X_2], which form a direct contrast with [X_1*de*] data. This will be a direction for future research. A next step of this research is to examine the pitch contours of neutral tone syllables with underlying tones.

6

Prosodic Variation in the Mandarin Accents

Contents

The phonetic characteristics of Mandarin accents have shifted between the older and younger generations of speakers in Changsha, so have the attitudes towards these non-standard Mandarin accents. The positive associations of Plastic Mandarin may be a concomitant of the gradual status shift of Mandarin from an L2 to an L1 in Changsha. This chapter portrays Plastic Mandarin from the perspective of language variation and change, bringing together the key threads of analysis presented in the Chapters 4 to 6. The main discussion is divided into four parts: §6.1 reviews the analysis of citation tones of Plastic Mandarin, Standard Mandarin, and Changsha, and answers the first set of research questions presented in Chapter

1. §6.2 reviews the analysis of neutral tones of Plastic Mandarin and Standard Mandarin, and answers the second and third set of research questions. Finally, §6.3 and §6.4 discuss the methods of analysing pitch contours and address the importance of open and reproducible practices.

6.1 Cross-varietal Variation in Lexical Tones: Distinct Realisation of Prosodic Prominence

RQ 1: *What are the citation tones of Plastic Mandarin? What are the potential sources of sound change?*

This study characterises the lexical tones of Plastic Mandarin through empirical data. Plastic Mandarin has the same four tonal categories as Standard Mandarin, and the citation tone patterns are summarised in Table 6.1:

Tone	Phonetic Description
1	a mid, level tone [˧]
2	a low to mid rising tone [˨˧]
3	a low, falling tone [˨˩]
4	a high rising tone [˦˥]

Table 6.1: Summary of Plastic Mandarin lexical tones.

The tone patterns of all four tones in Plastic Mandarin differ from their Standard Mandarin counterparts. In connected speech of Standard Mandarin, syllables in prosodically more prominent positions tend to maintain the canonical tone shape more faithfully, while prosodically weak syllables tend to have a reduced or distorted, non-canonical tone shape, especially in fast speech (Shih et al., 2001; Sui, 2016). The difference in the citation tones, therefore, leads to drastically distinct realisation of prosodic prominence.

6.1.1 Phonetic biases in contour changes

One possibility of the formation of Plastic Mandarin tone patterns is that Standard Mandarin tones had some variation by pure chance in Changsha and the randomly mutated tone variants somehow became generalised. Another (not exhaustive) possibility is that Standard Mandarin tones have undergone tonal contour changes in a similar phonetically biased chain shift fashion as reported in Bangkok Thai in the past century (see Pittayaporn, 2018, p. 259). In his analysis of Bangkok Thai tone changes, Pittayaporn (2018) proposed that tone variants arising from the production and perception of lexical tones tended to be crystallised and phonologised.

To explore the synchronic Mandarin variation in the perspective of tone contour change, Table 6.2 summarises the tonal changes by tone categories, by comparing the tone patterns of Plastic Mandarin and Standard Mandarin. When tone patterns are conceptualised as probabilistic approximation of high or low pitch target(s) that anchored to certain time points relative to the vowel or the sonorant part of a syllable, a varied tone pattern can be generated when the pitch target moves in both the temporal domain (leftward or rightward) and/or the pitch domain (upward or downward). Figure 6.1 schematically shows the rightward move of a high pitch target, and the tone pattern of the syllable demarcated by the vertical bars changes from a high falling shape into a high rising shape, when the flanking tones are not high tones. The characteristics of Standard–Plastic tonal variation patterns are also summarised in Table 6.2.

Tone	Standard Mandarin > Plastic Mandarin	Tone Variation
1	high level [˥] > mid level [˧]	downward shifting
2	mid rising [˧˥] > low rising [˨˧]	downward shifting
3	low fall-rise [˨˩˦] > low falling [˨˩]	rightward sliding of valley
4	high falling [˥˩] > high rising [˦˥]	rightward sliding of peak

Table 6.2: Summary of tone changes from Standard Mandarin to Plastic Mandarin. 'Rightward' is referenced to the syllable boundary.

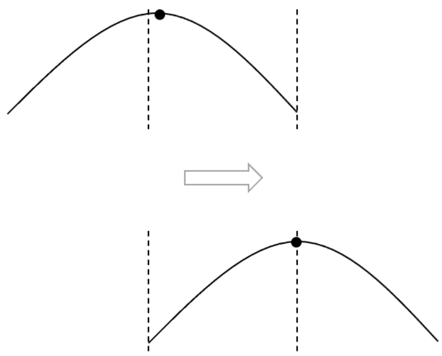


Figure 6.1: Schematic illustration of the right sliding of the peak (an H target) of a pitch contour. The two dotted vertical bars demarcates the prominent syllable with an H target in question. The black contour within the boundaries of the syllable represents its tone pattern. The black dots signal the pitch targets.

These tone variation patterns are in fact some of the most common and recurring patterns of tonal change cross-linguistically, identified in Yang and Xu (2019)'s review of 52 tone change studies in East and Southeast Asia. For example, that the mid-rising tone becomes the low-rising tone (Tone 2) is also observed in Dong'gan, Gaotian Zhuang, Tai Phuan, and Lalo; the contour reduction of the dipping tone, low fall-rise to low falling tone (Tone 3), is also observed in Taiwan Mandarin, Ma'ya, Libo, Taipusi. Many unrelated languages have parallel development of certain types of tone patterns.

Yang and Xu (2019) described the overall crosslinguistic tendencies in tone change as "clockwise[1], levelling, and regress to mid patterns". The variation pattern of Tone 1 in Table 6.2 involves a downward shift to mid level, which matches their description of "regressing to mid". The variation patterns of Tone 3 and Tone 4 in Table 6.2 can be generalised as the rightward sliding of low valley and high peak (or pitch target in both cases), which is a feature of their clockwise pattern. The variation patterns of contour tones also involve contour reduction, that is, flattened contour convexity (Tone 3) and decreased f_0 excursion (Tones 2 and 4). That

[1]This concept is a little opaque. It claims a cyclical tone chain shift. See Yang and Xu (2019) for details and schematic illustration.

the Standard–Plastic tonal variation largely matches the crosslinguistic tendencies strengthens the phonetically based approach to sound change.

Pittayaporn (2018) stated that f_0 peaks tend to be delayed rather than early, introducing a bias in sound change. In Mandarin, for instance, the pitch peak of the rising tone may occur after the syllable boundary in certain environments (Y. Xu & Wang, 2001). Yang and Xu (2019) further attributed such rightward sliding pitch movements to the carryover effects in connected speech and the truncation of tone production. This articulation motivated account may explain the Tone 3 change, since the duration of Plastic Mandarin Tone 3 is much shorter than Standard Mandarin Tone 3 (Recall Figure 3.19 in §3.5.2), and the resulting low falling contour can be regarded as a truncated downward concave shape. However, it does not seem tenable for the case of Tone 4, since there is no evidence of tone truncation.

Despite that the phonetic bias account characterises the general patterns of tone change that are more likely to occur, there are language-specific differences. Furthermore, there is no clear phonetic motivation for the downward shift of Tone 1 and Tone 2.

6.1.2 Identity marking in prosody

The comparison of the tone patterns of Plastic Mandarin and Changsha leads to a potentially more promising account of the mechanism of tone change, contact-induced borrowing, than the phonetic bias account.

Table 6.3 shows the group differences in the aspects of f_0 patterns between Plastic Mandarin and Changsha by tone categories, in which the fewer grey cells in a set of contour attributes, i.e. curvature, slope, and average, the more similar the tone pairs. Together with the individual tone patterns 3.45 displayed in §3.7.4, a striking similarity in f_0 patterns between the Plastic Mandarin Tones 1, 2, and 4 and the corresponding Changsha tones is evident, especially for Tone 4. It suggests that Plastic Mandarin may have borrowed the tone pattern of the corresponding Changsha tone, for Tones 1, 2, and 4. The historical tone categories still play a mediating role in contemporary tonal change. The analysis in the two conditions

Tone	A: $\bar{f}_s$ (PM) $\neq$ $\bar{f}_s$ (CS)			B: $\bar{f}_s$ (PM) = $\bar{f}_s$ (CS)		
1	Curvature	Slope	Average	Curvature	Slope	Average
2	Curvature	Slope	Average	Curvature	Slope	Average
3	Curvature	Slope	Average	Curvature	Slope	Average
4	Curvature	Slope	Average	Curvature	Slope	Average

Table 6.3: Summary of comparison of Plastic Mandarin and Changsha tones produced by the same group of speakers. Two conditions are included: A. speaker mean $\bar{f}_s$ is variety-dependent for the same speaker; B. speaker mean $\bar{f}_s$ is uniform across varieties for the same speaker. Aspects of f_0 contour that are statistically significantly different between Plastic Mandarin and Changsha are highlighted in grey.

of speaker mean $\bar{f}_s$ in Table 6.3 also reveals that the absolute average of Tone 4 differs between Plastic Mandarin and Changsha, but the relative average height of Tone 4 in the tonal system of each variety does not differ significantly. This implies that speakers borrow the tone shape and adjust the height of the 'borrowed' tone in creating a new tonal system, instead of direct copying.

Standard Mandarin syllables with Tone 4, the high falling tone, tend to be judged the most prominent among other tones in both prosodically strong and weak positions (Deng, 2010). Tone 4 is also the most frequent tone in Standard Mandarin (Y. Wu et al., 2020). The change of the Tone 4 tone pattern is, therefore, extremely perceptually salient, and it puts the borrowed Changsha Tone 4 with a distinct high rising pattern in the spotlight. Furthermore, the high rising tone pattern is a shared tone feature among various Xiang varieties in Hunan province. The adaptation of Changsha tone patterns as new canonical tone patterns has allowed Plastic Mandarin to express a shared local and cultural identity and social solidarity. While specialised lexicon associated with local food, festivals, activities, and so on is often the site or marker of cultural identity in a local Mandarin variety, the case of Plastic Mandarin shows that **identity can be marked prosodically through pitch**.

6.1.3 Potential sound change mechanism

The contact-induced borrowing account does not explain why Tone 3 in Plastic Mandarin is very different from that in Changsha. It is likely that other biases or

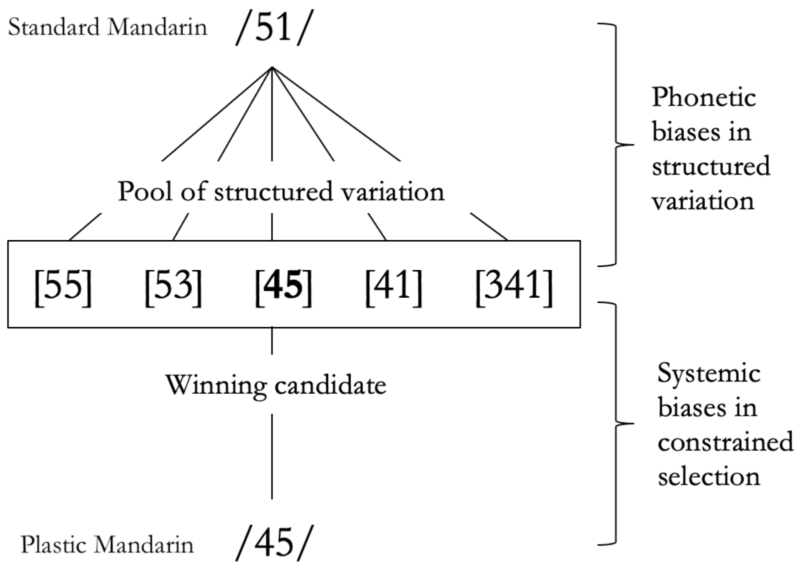


Figure 6.2: Schematic phonologisation of the high rising contour of Tone 4 in Plastic Mandarin.

constraints also exert some influence in the process. The aforementioned accounts may not be mutually exclusive. In addition to the phonetic biases, Pittayaporn (2018) identified three systemic biases in tonal contour changes: contour maximisation, contour accentuation, and avoidance of similar tones, which favoured variants of dynamic tones with a greater and more dramatic f_0 excursion, and tone variants dispersed in phonetic space. Such biases were supported by perception experiments. For example, dynamic pitch excursions improve speech recognition in noise, and speech with dynamic pitch is consistently better perceived than monotone speech (J. Shen & Souza, 2018). The Tone 3 in Changsha [˦˧] lies in the mid-high space of the pitch range, which would not be optimal in maximising perceptual contrast if adopted by Plastic Mandarin, given another two mid-range tones, Tone 1 [˧] and Tone 2 [˧˦], and a high-range tone 4 [˦˥] in Plastic Mandarin. The low falling Tone 3 [˨˩] in Plastic Mandarin, instead, magnify the perceptual distance between Tone 3 and other tones.

The phonetic and systemic biases, as well as the contact-induced borrowing effects, altogether tend to offer a coherent explanation for the synchronic Plastic Mandarin variation. Figure 6.2 schematises the potential phonologisation of the high rising Tone 4 contour, following Pittayaporn (2018). A pool of structured variants motivated by phonetic biases, by borrowing, or by chance are first generated. The variants illustrated in Figure 6.2, which are not exhaustive, are chosen based on common crosslinguistic patterns and phonetic biases such as the effect of syllable onset and peak delay. Among all these variants, [45] [˦˥] was preferred and won out as the primary variant, since it resembles the Changsha Tone 4, the 'signature' Xiang tone. The bottom part of the figure shows the selectional bias, and [45] is selected as the new canonical tone value in Plastic Mandarin.

6.2 Cross-varietal Constancy in Neutral Tone: Stable Transfer of Prosodic Structure

RQ 2: *What are the phonetic patterns of neutral tone in Plastic Mandarin? How do they compare to those of neutral tone in Standard Mandarin?*

The present thesis examines speech sequences with neutral tone syllables in both Mandarin varieties. There is no canonical f_0 contour for a neutral tone syllable in both Mandarin varieties. The shape varies greatly when the syllable is surrounded by different tones, and the preceding lexical tone is a main source of variability. The preceding tone, however, does not fully account for the f_0 pattern of a neutral tone.

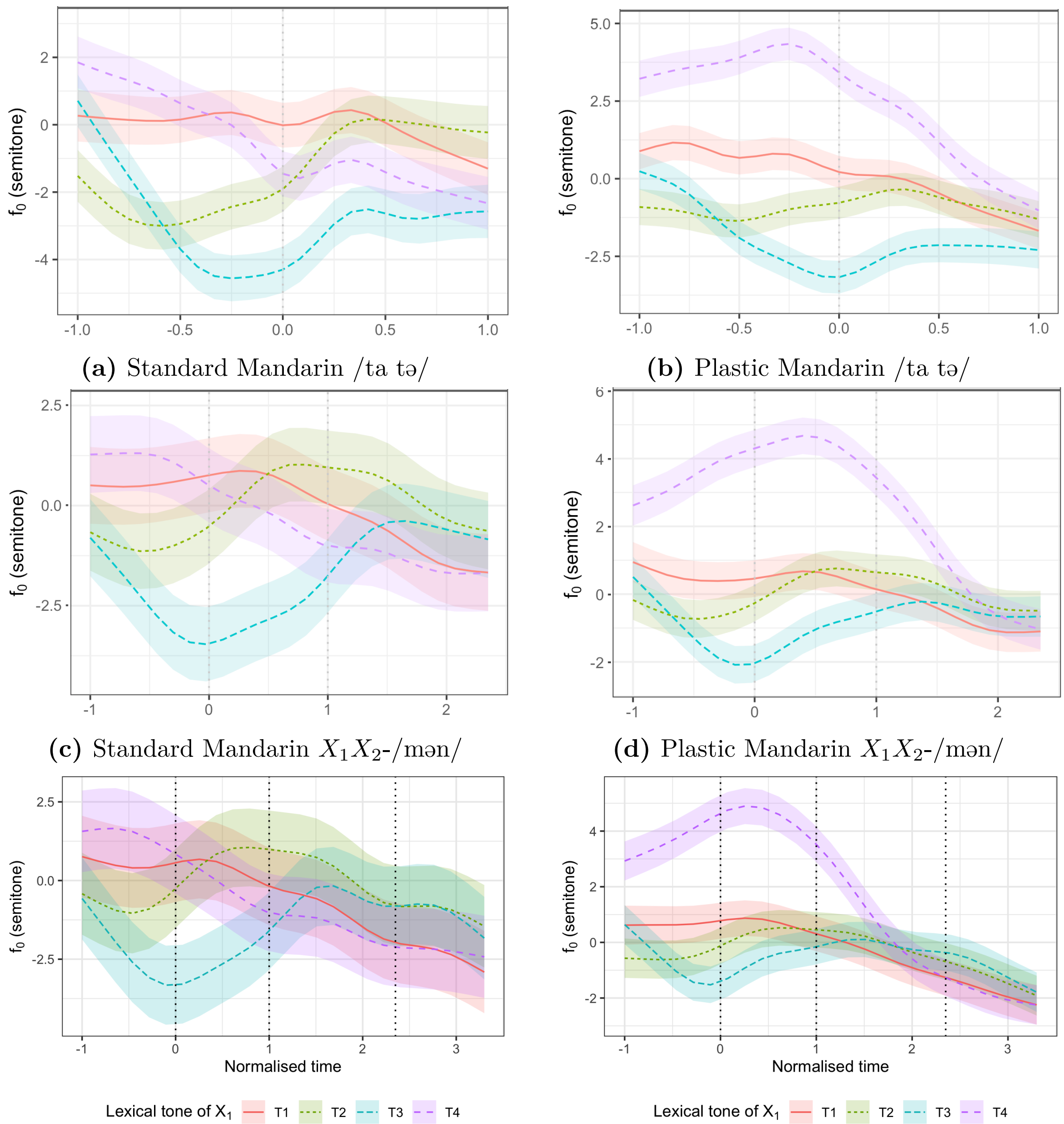
The GAMM analysis reveals that duration is also a significant predictor in the shape of a neutral tone in connected speech. Y. Chen and Xu (2006) predicted that the underlying target for a neutral tone should become more apparent if time becomes more abundant, based on the assumption of target approximation. Our data show that the f_0 trajectory of a longer neutral tone or neutral tone sequence tends to have a falling trend towards the end with a lower final f_0 and the final f_0 is highly likely to be lower than the speaker mean. The consistent falling pattern at the end of f_0 trajectories irrespective of the preceding tone categories indicates that the influence of the preceding lexical tone has gradually tapered off and neutral tone is approaching its own target.

Table 6.4 presents a summary of f_0 contours in the disyllabic, trisyllabic, and quadrisyllabic phrases in both Mandarin varieties, and Figure 6.3 pastes and

	Standard Mandarin		Plastic Mandarin	
	Convergence	Description	Convergence	Description
X-de	×	T1/T2-de contours converge at a higher pitch,T3/T4-de contours converge at a lower pitch.	×	Most contours converged, T3-de contours significantly differ from T2-de and T4-de contours.
X_1X_2-men	✓	All contours converge at a pitch lower than speaker mean, with a mean of −1.24 semitones.	✓	All contours converge at a pitch lower than speaker mean, with a mean of −0.67 semitones.
X_1X_2-men-de	✓	All contours converge at a pitch lower than speaker mean, with a mean of −2.13 semitones.	✓	All contours converge at a pitch lower than speaker mean, with a mean of −1.97 semitones.

Table 6.4: Summary of neutral tone f_0 contours in the disyllabic, trisyllabic, and quadrisyllabic phrases in Standard Mandarin and Plastic Mandarin. When all types of f_0 contours do not statistically significantly differ among each other at the end, it is considered that there is a convergence.

rearranges some of the figures in Chapter 5 here for easy reference. When there are two consecutive neutral tone syllables including the duplicate syllable X_2 and *men*, f_0 contours in various shapes converge at the very end in both Mandarin varieties. When an extra neutral tone syllable *de* is added to such phrases, the converged f_0 contours pattern together and continue falling. It suggests that the carryover effects of the preceding lexical tone syllable X_1 do not reach further than two neutral tones in the data and that neutral tone is inclined to approach a low pitch target in both Mandarin varieties. At the end of three consecutive neutral tones, the average f_0 is approximately 2 semitones lower than the speaker mean in both Mandarin varieties. The average final f_0 of the three-neutral tone sequences, though, is not as low as the lowest f_0 of Tone 3 in Standard Mandarin (see (e) in Figure 6.3). It is approximately 4 semitones lower than speaker mean on average in the citation tone model. In Plastic Mandarin, the average final f_0 of neutral tone sequences can be as low as the average lowest f_0 of Tone 3 in the [X_1X_2-*men de*] phrases, although in the citation tone model the average lowest f_0 of Tone 3 is approaching 5 semitones below speaker mean. This study suggests that **neutral tone is associated a low pitch target in both Mandarin varieties despite its pitch contour is variable**. In connected speech the actual f_0 implementation of a neutral tone often does not reach the bottom of a speaker's pitch range.



(a) Standard Mandarin /ta tə/ **(b)** Plastic Mandarin /ta tə/

(c) Standard Mandarin X_1X_2-/mən/ **(d)** Plastic Mandarin X_1X_2-/mən/

(e) Standard Mandarin X_1X_2-/mən tə/ **(f)** Plastic Mandarin X_1X_2-/mən tə/

Figure 6.3: Recap of f_0 contours in the disyllabic, trisyllabic, and quadrisyllabic phrases in Standard Mandarin and Plastic Mandarin.

Another piece of evidence of a low pitch target comes from the citation form of a neutral tone syllable, when participants were asked to read out the individual neutral tone syllables (recall Chapter 3). Since neutral tone syllables do not have a canonical or prescribed citation tone, speakers may produce them in their most typical or exemplar form resulting from the high frequency of occurrences. In Plastic Mandarin, the typical citation form of syllables [tə], [lə], and [ʈʂə] demonstrates a descending pitch pattern that consistently falls within a lower pitch range compared to the speaker mean, reaching approximately 4 semitones below the speaker mean (see §3.6.2.4). It is a slightly different case in Standard Mandarin, where the pitch patterns of syllables [tə] and [ʈʂə] descend to around 1 semitone below the speaker mean.

RQ 3: *How do full lexical tones and neutral tone interact with the Standard-Plastic language variation?*

Despite the greater surface variability than other lexical tones, neutral tone displays some constant or pertinacious aspects of prosody in this Standard–Plastic variation. Unlike the canonical lexical tones that have systematic contour differences, neutral tone tends to be prosodically conditioned in a similar fashion in the two Mandarin varieties. The strength of the effects of the preceding lexical tones is similar in various phrases between Standard Mandarin and Plastic Mandarin. When the influence of the preceding lexical tones is depleted in a neutral tone sequence, the f_0 patterns of the last neutral tone syllable *de* in both Mandarin varieties are not statistically significantly different and exhibit a falling pattern below the speaker mean (see Figure 5.38 in §5.4.5). The gradual falling pattern suggests that a low pitch target is at the end of the neutral tone sequence. If we assign a low pitch target for every neutral tone syllable in a sequence, it is then very likely to have a low level f_0 pattern for a consecutive neutral tone in a sequence. This thesis, hence, proposes that **there are fewer tonal targets than syllables** in connected speech of Mandarin.

Q. Li and Chen (2019) and K. Huang (2018) also observed a mid-low target for neutral tone sequences in Tianjin Mandarin and Taiwan Mandarin respectively. The convergence of the f_0 patterns of neutral tone in different Mandarin varieties and the emerging low pitch target can be interpreted in two ways: first, in Mandarin variation and change, neutral tone is an invariant low tone and somehow remains unaffected by the systematic tone contour changes observed in other lexical tones; second, neutral tone is underlyingly toneless (underspecified, hence no corresponding tone contour change) but attracts a boundary low tone (Z. Li, 2003). The latter can easily explain why neutral tone tends to have a constant pitch target in different Mandarin varieties when all canonical lexical tones have altered their tone targets or specifications. The perspective of synchronic variation favours the boundary tone interpretation and suggests that neutral tone production involves a mechanism relatively stable in Mandarin variation — the prosodic structure and groupings are similar between Standard Mandarin and Plastic Mandarin. This is not surprising when the morphosyntactic structure is consistent across these two Mandarin varieties. The approximation to a low pitch target in neutral tones may have marked strong prosodic boundaries, and the prosodic grouping in the phrases containing neutral tone syllables are the same between the two Mandarin varieties.

Further empirical research, however, is still needed to confirm and explore boundary effects in a wider variety of neutral tones and other lexical tones. Regardless of the phonological interpretations, the falling pitch pattern of neutral tones is characterised by pitch-related prosodic preplanning and weak articulatory strength.

On the whole, tonal patterns are the main site of phonetic variation along the Standard–Plastic dimension. The adaptation of the Changsha high rising Tone 4 in Plastic Mandarin manifests how local identity is represented and marked prosodically, mainly through pitch. The constancy in the pitch target of neutral tones between the two Mandarin varieties suggests pertinacious prosodic grouping and boundary marking in the process of the language change.

6.3 Methodological Discussion

Two major approaches utilised in this thesis to investigate and predict phonetic variability, especially f_0 contours, in the cross-varietal speech datasets. The conversion of monosyllabic contours into three fully interpretable coefficients can facilitate comparison and classification. For longer contours, Gamm has been a considerably valuable method in analysing the convergence of f_0 contours and determining when the convergence starts. Considering the extremely dynamic nature of speech, mixed models incorporating multiple sources of variability are crucial. In terms of duration analysis, our data often have a right-skewed residual distribution in the process of modelling, which may result from unknown but influential factors. Robust method (Koller, 2016) flags the unusual data predicted by a certain model and enables estimation where the unusual data have only little influence. While assumptions checking is often not required to be reported in the linguistic publications, it is, nevertheless, an indispensable step in statistical model selection and statistical inference interpretation.

The duration analysis points to an interesting phenomenon that the duration hierarchy of citation tones in Standard Mandarin *Tone 3 > Tone 2 >Tone 1 > Tone 4* has been consistent since Dreher and Lee (1968). It shows that duration can be a very stable acoustic attribute over time. For Plastic Mandarin citation tones, Tones 1, 2, and 4 have similar duration to the Standard Mandarin counterparts, but Tone 3 is significantly shorter, with the accompanying contour reduction. In respect of the duration variance of tones, the data show the hierarchy of *Standard Mandarin > Plastic Mandarin > Changsha*, which may suggest that Standard Mandarin is more 'stressed-timed' and less 'syllable-timed' than Changsha. While many previous studies (Cao, 1986, e.g.) claimed that the duration of neutral tone syllables is shorter than the preceding lexical tone syllables, such description does not seem fully accurate. The duration analysis of neutral tone syllables in this thesis indicates that syllable structure plays a role in duration, and that only CV syllables tend to be shorter than the preceding lexical tone syllables when they bear neutral tone. In other words, unstressed syllables tend to have lost or restrained the open-syllable lengthening effect.

Many Plastic Mandarin speakers were identified as "trilinguals" in Standard Mandarin, Plastic Mandarin, and Changsha (Jing & Niu, 2010). An interesting avenue of future exploration would involve examining how these speakers in Changsha produce Standard Mandarin and analysing the potential divergences from the Standard Mandarin spoken by the mainstream northern (mainly Beijing) speakers in this study, so as to further investigate the origin of Plastic Mandarin tones from a language acquisition perspective. Neutral tone in Changsha is another area worth further investigation.

6.4 Open Science and Reproducibility

In the field of quantitative phonetics, there are two major areas of problems: one is statistical power and errors; the other is researcher degrees of freedom.

The consequence of low statistical power can be failure to replicate, which means that the findings of the sample data are not generalisable enough, and the same results cannot be obtained on new data in an independent replication of the study. The number of participants in phonetic studies is generally low given the data collection and annotation is often laborious and time-consuming. Roettger and Gordon (2017) conducted a systematic review of 110 studies of acoustic correlates of word stress on 75 languages published between 1955 and 2017, and found that the majority of studies included 1 to 10 speakers (with the mode being only one speaker), 1 to 40 lexical items, and 1 to 6 repetitions (with the mode being a single repetition). J. Kirby and Sonderegger (2018) illustrated how power, magnitude, and sign errors vary with sample and effect size through a a case study on German incomplete neutralisation, and suggested that the estimation of small effect sizes was particularly subject to statistical power related issues. Winter (2015) further showed that increasing the number of repetitions was not a replacement for including more unique items since a high number of repetitions of few items only led to increasing statistical certainty of the estimates of idiosyncratic differences between the items.

Simmons et al. (2011) constructed the term "researcher degrees of freedom" to refer to methodological and analytical decisions that researchers made in the course of data collection and analysis. For phonetics, the researcher degrees of freedom is high due to the inherent multidimensionality of speech behaviour involving complex interaction between various functional layers (Roettger, 2019). Speech production research, for example, involves decisions of choosing and operationalising phonetic parameters, discarding (correcting) data, selecting independent variables, adjusting model parameters, and so on. The wide variety of choices without enough detailed information on the analysis procedure can result in low reproducibility, that is, the same results cannot even be obtained by re-running the original analysis on the original data. Roettger (2019) warned that exploitation of researcher degrees of freedom introduced strong bias and posed serious challenge to phonetics as an empirical science. In addition, even many published articles failed to outline any procedure for model selection and checks, raising concerns about the robustness and validity of their models.

This research project has been carried out with fundamental concepts and principles of open science in mind. It aims to provide open, honest, and transparent practices in communicating data analytical decisions and findings. The data (subject to ethical restrictions), analysis scripts, and information about the computational environment can be found in the openly accessible research repository hosted on the Open Science Framework (https://osf.io/ufhzt/). More specifically, the workflow consisted of four main parts. First, the data have been packaged and documented under the guidance of Berez-Kroeker, Gawne, et al. (2018) and Berez-Kroeker, Andreassen, et al. (2018). The preservation and sharing of my curated speech data or corpora facilitate the accumulation of knowledge and encourage future work so as to extend their value. Second, the pre-possessing procedure of the speech data was demonstrated step-by-step in an online tutorial (available at https://chenzixu.rbind.io/resources/1forcedalignment/), with modified P2FA forced alignment scripts publicly accessible at a Github repository (https://github.com/chenchenzi/P2FA_Mandarin_py3). Third, the manual corrections of acoustic measurements and some auditory observations of speech tokens were logged in a spreadsheet. Fourth, within the literate programming framework and the aid of R Markdown (Allaire et al., 2022), code snippets and documentation of each research topic were interwoven and integrated in a single source file, where summaries of statistical models, diagnostic tests, and figures were automated and derived. These practices ensure the reproducibility and accountability of my work, and facilitate the validation of data analyses by other researchers. Appendix D.1 provides a comprehensive list of the software and packages used in this thesis, with each of them credited and acknowledged.

With regard to statistical power, I have made the best attempt possible in this project, considering the financial and time limits and the COVID pandemic, to increase the sample size, in terms of the number of speakers and the number of observations from a speaker, as well as the number of different lexical items and the number of repetitions. By combining corpus data and fieldwork elicitation data from a total of 35 speakers, this project has analysed thousands of speech tokens in different conditions.

Appendices

A
Recruitment Documents

Contents

A.1 Recruitment Package Cover

The following pages attached the recruitment poster in Standard Written Chinese (A.1.1) and its English version (A.1.2). The Chinese version was used as the cover of the recruitment package for Plastic Mandarin speakers.

A.1.1 Recruitment Poster (Chinese)

牛津大学语言学文字学及语音学学院
149 Walton St, Oxford, OX1 2HG, UK

徐晨子
chenzi.xu@ling-phil.ox.ac.uk

湖南"塑普"的语言学研究

牛津大学研究操守委员会批准编号: R67136/RE001

湖南"塑普"语言学研究 志愿者招募

每一种语言都有其独特之处。湖南"塑料普通话"（简称"塑普"）作为一种很年轻的地方口音，现在已经越来越普及，尤其是学生群体在学校的交流运用。 此项研究旨在探索湖南"塑普"的一些语言特性。在南雅中学，我们将采集一些地道的"塑普"和长沙话的语音数据。这项研究将有助于我们更好地认识"塑普"，加深我们对语言的接触与变迁的理解。

如果你对"塑普"有兴趣，无听说障碍，达到 16 岁或 16 岁以上，长期生活在长沙，会讲"塑普"和长沙话，就可以参加这项学术研究。 特别欢迎"塑普"和标准普通话都会讲的同学们。若成为志愿者，你将被邀请参加一个约 30 分钟到 45 分钟的语音活动环节，在长沙市南雅中学进行 (具体另行通知)。活动过程中，你需要讲"塑普"和长沙话，并被全程录音。

如果你有兴趣并想了解更多研究咨询，请联系徐晨子（发送电邮至 chenzi.xu@ling-phil.ox.ac.uk），或者通过班主任报名成为志愿者。

活动结束后你将获得一张研究参加证书。非常感谢！

A.1.2 Recruitment Poster (English)

Faculty of Linguistics, Philology and Phonetics
149 Walton St, Oxford, OX1 2HG, UK

Chenzi Xu
chenzi.xu@ling-phil.ox.ac.uk

Characterising Mandarin Accents

Central University Research Ethics Committee (CUREC) Approval Reference: R67136/RE001

VOLUNTEERS NEEDED FOR THE STUDY ON MANDARIN ACCENTS

Each spoken language has its own sound patterns and properties. This project is interested in acoustic properties of different Mandarin varieties. One of them is "Plastic Mandarin" in Changsha, Hunan Province. In Nanya Middle School we will collect some authentic speech data from high school students. This project will help us understand what characterises Plastic Mandarin and how new varieties of speech emerge.

We are looking for healthy volunteers, aged 16 or over to attend a verbal game session with audio recording. You need to be a long-term Changsha resident who speaks "Plastic Mandarin" and Changsha, and preferably Standard Mandarin too. You would be invited to attend one recording session at Nanya Middle School. The session would take about 30-45 minutes of your time. You would be asked to speak in Plastic Mandarin and Changsha during the session and be audio recorded.

If you are interested and would like more information, please contact Chenzi Xu at chenzi.xu@ling-phil.ox.ac.uk or sign up with your head teacher. There is no obligation to take part.

You will be compensated for your time by a certificate of participation. Thank you!

Poster Advert_Version 1.0 December 2019

A.2 Participant Information Sheet

The following pages provide the participant information sheets used in this study. Participant Information Sheet A.2.1 in Standard Written Chinese was prepared for the high-school Plastic Mandarin speakers, and A.2.2 is its English translation. Participant Information Sheet A.2.3 in English was prepared for the Standard Mandarin speakers in Oxford.

A.2.1 Participant Information Sheet (Chinese, PM Group)

牛津大学语言学文字学及语音学学院

PI Name: Professor John Coleman
Primary Researcher: Chenzi Xu [DPhil Student]
Phonetics Laboratory, 41 Wellington Square, Oxford OX1 2JF
+44 (0)1865 270444
chenzi.xu@ling-phil.ox.ac.uk

湖南"塑普"的语言学研究

志愿者须知

牛津大学研究操守委员会批准编号: R67136/RE001

1. 为什么做这项研究?

每一种语言都有其独特之处。湖南"塑料普通话"作为一种很年轻的地方口音，现在已经越来越普及，尤其是学生群体在学校的交流运用。这项研究旨在探索湖南"塑普"的一些语言特性。在南雅中学，我们将采集一些地道的"塑普"语音数据。 这项研究将有助于我们更好地认识"塑普"，加深我们对语言的接触与变迁的理解。

这项研究由牛津大学语言学文字学及语音学学院提供资金。

2. **为什么我被邀请参加这项研究?**

您被邀请因为您达到 16 岁或 16 岁以上，无听说障碍，长期生活在长沙，会讲"塑料普通话"和长沙话（若能讲标准普通话更好）。我们希望在长沙市南雅中学招募约 16 至 20 名年轻志愿者。

3. **我必须参加吗?**

不是。在您决定之前，若有任何疑虑，欢迎向研究员提问。如果您同意参加，您亦有权随时退出，无需提供任何原由，也不会有任何形式的惩罚。退出时，只需提前告知研究员。

4. **如果我参加这项研究，我要做些什么?**

如果您乐意参加这项研究，您将和一位您的同学一起参加一个语音活动环节，大约持续 30 到 45 分钟。活动将会在长沙市南雅中学的一间隔音条件较好的多媒体教室或音乐教室进行。
活动正式开始前，一些关于语言背景和年龄等的个人信息将被采集。当您到达活动地点时， 我会跟您详细介绍活动流程，你将有机会提问。活动中，在您的同意下，我会给您带上便携式麦克风并且全程进行录音。在您的同意下，我可能会拍摄照片。您可以随时暂停或终止语音活动。我们所采集

的语音资料将完全匿名并保持机密，有可能用作之后的学术研究。 如果您仍乐意参与这项研究，请填写《志愿者同意书》。倘若录音效果不佳，可能有回访。

5. **参加这项研究有潜在风险吗?**

参加这项研究不存在潜在风险。您的个人信息将会被拟匿名化处理。

6. **参加这项研究有利益吗?**

参加这项研究没有直接利益。

7. **花销或报酬**

您无需支付任何费用。资料收集完毕后，您将获得一张研究参与证书。

8. **如何处理所收集的我的数据?**

在参与研究过程中你所提供的信息是研究数据。任何可以识别您身份的研究数据包括姓名、年龄、和录音被称作个人数据。

所有研究数据和个人/敏感数据（包括同意书）将会安全转移并机密地储存在牛津大学语音实验室的内部硬盘中。纸质填写的资料将会安全地储存在研究员的私人区域。录音将由研究员转录为文字。研究员及其导师将能获取此研究数据。所有研究数据和记录将在论文发表或研究工作公开发布后至少保留 3 年。牛津大学责任方成员可能获取获取此研究数据，以便监察是项研究。

我希望您能允许我们在研究发表中直接引用您所说的，以匿名的方式或者如果您希望我们标明出处（即您的姓名）。我希望您能同意我们将所收集的语音数据保留并做成语料库， 以便未来的研究。我希望您能同意我们在未来的研究项目中匿名使用这些数据，并且共享给其他研究者（例如以线上数据库的方式）。所有可以识别你的个人信息将会在数据共享给其他研究者或者成果公开后移除或更改。

9. **这项研究会发表么?**

我们希望在学术会议或科学期刊中发表研究成果，这个过程可能是在研究结束后的两到三年内。若您同意参加，您在活动当中的话可能会被直接引用在论文，会议陈述，或学术发表中，用来举例说明某种语言现象。专有名称将会匿名处理，不会出现任何可能暴露参加者的信息。所有研究数据和记录在论文发表或研究工作公开发布后将至少会保留 3 年。

牛津大学致力于为社会传播它的研究，因此建立了一个在线研究材料档案库。这个档案库中包含了成功提交至牛津大学的硕博论文的电子版。这样一个在线档案库给研究者提供了便捷的方式获取论文全文的免费阅读，以便于加强研究的影响和使用。该论文有可能在开放获取的期刊上发表。在发表或存档的文章中，将无法辨认参与者的身份。

是项研究将会写成一篇研究论文，成功提交后将保存在牛津大学印刷版及网上的档案库中，以便于日后使用。若此，这篇论文将能公开获取。

10. **谁审核了这项研究?**

这项研究已经通过牛津大学研究操守委员会的审批， 编号为 R67136/RE001。

11. **如果我有疑虑或者想投诉该联系谁?**

如果您对此研究项目有任何疑问，请联系研究员徐晨子 (chenzi.xu@ling-phil.ox.ac.uk) 或者她的导师 John Coleman 教授(john.coleman@phon.ox.ac.uk) , 我们会尽力回答您的问题。研究员通常会在 10 个工作日内确认您的疑虑并回复您我们将如何处理。如果您对答复不满意或想进行法律起诉，请联系牛津大学社会科学与人文学科研究操守委员会的主席，他将尽快解决事务（**电邮**: ethics@socsci.ox.ac.uk; 地址: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD ）。

12. **联系方式**

若您想和研究员讨论此研究（或者您之后有任何问题），请联系：

徐晨子
牛津大学语言学、文字学和语音学学院
地址： 149 Walton St, Oxford, OX1 2HG, UK
电话: +44 (0)1865 270444
电邮: chenzi.xu@ling-phil.ox.ac.uk

A.2.2 Participant Information Sheet (English, PM Group)

Faculty of Linguistics, Philology and Phonetics

Professor John Coleman
Phonetics Laboratory, 41 Wellington Square, Oxford OX1 2JF
Chenzi Xu [DPhil student]
Oxford phone number: +44 (0)1865 270444
Oxford e-mail: chenzi.xu@ling-phil.ox.ac.uk

Characterising Mandarin Accents

PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee (CUREC) Approval Reference: R67136/RE001

1. ***Why is this research being conducted?***

Each spoken language has its own sound patterns and properties. This project is interested in the acoustic properties of different Mandarin accents. One of them is "Plastic Mandarin", that is, the accent of young people's Mandarin in Changsha, Hunan Province. This project will help us understand what happens when languages are in contact with each other and how new varieties of speech emerge.

This study is funded by Faculty of Linguistics, Philology and Phonetics, University of Oxford.

2. ***Why have I been invited to take part?***

You have been invited because you are a young person aged 16 years old or above, a long-term Changsha resident, able to speak Plastic Mandarin and Changsha, and preferably, Standard Mandarin too, and attending Changsha Nanya Middle School.

3. ***Do I have to take part?***

No. You can ask questions about the study before deciding whether or not to participate. If you do agree to participate, you may withdraw yourself and your data from the study at any time, without giving a reason and without penalty, by advising me of this decision.

4. ***What will happen to me if I take part in the research?***

If you are happy to take part in the study, you will be asked to attend a semi-structured pair recording session together with one of your friends or classmates in a music room in Nanya Middle School. The session consists of a few verbal games. Before the session, some personal information such as age and language background will be collected. When you arrive, I will talk you through the procedure and give you the chance to ask any questions. You will wear a microphone with your speech being recorded during the interview, with your consent. A photo may be taken with your consent.

Participant Information Sheet 1.0 December 2019 Page 1 of 3

The recording session should take approximately 30-45 minutes. You can also ask to pause or stop the interview at any time. Your speech data will be kept confidential, and they may be used in further study on speech. If you are still happy to take part, I will ask you to sign a consent form. There may be a follow-up session if the recordings are not in good conditions.

5. Are there any potential risks in taking part?

There is no anticipated risk associated with participation in this research. To reduce any potential risks, your personal data will be pseudonymised as appropriate.

6. Are there any benefits in taking part?

There will be no direct benefit to you from taking part in this research.

7. Expenses and payments

You will receive a certificate for participation.

8. What happens to the data provided?

The information you provide during the study is the **research data**. Any research data from which you can be identified such as name, age, and audio recordings is known as **personal data**.

The **research data** and **personal / sensitive** data (including consent forms) will be transferred and stored confidentially in the internal drive in the Phonetics Lab, University of Oxford. Hard copies will be stored in the primary researcher's private premises. The recordings will be transcribed by the primary researcher. The researcher and supervisors will have access to research data. All research data and records will be stored for a minimum retention period of 3 years after publication or public release of the work of the research. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

I would like your permission to use direct quotes anonymously or for your name to be attributed to these if you wish in any research outputs. I would like your permission to archive the audio recordings as a speech corpus and retain them for future research. I would like your permission to use anonymised data in future studies, and to share data with other researchers (e.g. in online databases). All personal information that could identify you will be removed or changed before information is shared with other researchers or results are made public.

9. Will the research be published?

The research may be published in conference presentations and scientific journals, but this may be two to three years after the end of the study. If you consent to this, direct quotes from the interviews may be included in the thesis or in presentations and publications, in order to illustrate linguistic points. Proper names will be anonymised and no information will be included that might allow participants to be identified.

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

If you agree to participate in this study, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible.

10. Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number: R67136/RE001).

11. *Who do I contact if I have a concern about the study or I wish to complain?*

If you have a concern about any aspect of this study, please speak to Chenzi Xu (chenzi.xu@ling-phil.ox.ac.uk) or her supervisor Prof. John Coleman (john.coleman@phon.ox.ac.uk), and we will do our best to answer your query. I should acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the relevant chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter in a reasonably expeditious manner:

Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee**; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD; Email: ethics@socsci.ox.ac.uk.

12. Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest.
Further information about your rights with respect to your personal data is available from https://compliance.web.ox.ac.uk/individual-rights.

13. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Chenzi Xu
Faculty of Linguistics, Philology and Phonetics
149 Walton St, Oxford, OX1 2HG, UK
Tel: +44 (0)1865 270444
Email: chenzi.xu@ling-phil.ox.ac.uk

A.2.3 Participant Information Sheet (English, SM Group)

Faculty of Linguistics, Philology and Phonetics

Professor John Coleman
Phonetics Laboratory, 41 Wellington Square, Oxford OX1 2JF
Chenzi Xu [DPhil student]
Oxford phone number: +44 (0)1865 270444
Oxford e-mail: chenzi.xu@ling-phil.ox.ac.uk

Characterising Mandarin Accents

PARTICIPANT INFORMATION SHEET

Central University Research Ethics Committee (CUREC) Approval Reference: R67136/RE001

1. Why is this research being conducted?

Each spoken language has its own sound patterns and properties. This project is interested in the acoustic properties of different Mandarin accents. This project will help us understand what happens when languages are in contact with each other and how new varieties of speech emerge.

This study is funded by Faculty of Linguistics, Philology and Phonetics, University of Oxford.

2. Why have I been invited to take part?

You have been invited because you are aged 16 years old or above, a Mandarin-speaking native who **do not** speak other Chinese varieties (e.g. Cantonese, Xiang Chinese, etc.). Preferably, you were born and bred in Beijing.

3. Do I have to take part?

No. You can ask questions about the study before deciding whether or not to participate. If you do agree to participate, you may withdraw yourself and your data from the study at any time, without giving a reason and without penalty, by advising me of this decision.

4. What will happen to me if I take part in the research?

If you are happy to take part in the study, you will be asked to attend a semi-structured pair recording session together with one of your friends in the Recording Studio in the *Phonetics Laboratory, University of Oxford*. The interview consists of a few verbal games. Before the session, some personal information such as age and language background will be collected. When you arrive, I will talk you through the procedure and give you the chance to ask any questions. You will wear a microphone with your speech being recorded during the interview, with your consent. A photo may be taken with your consent.

The recording session should take approximately 30-45 minutes. You can also ask to pause or stop the interview at any time. Your speech data will be kept confidential, and they may be used in further study on speech. If you are still happy to take part, I will ask you to sign a consent form. There may be a follow-up session if the recordings are not in good conditions.

5. Are there any potential risks in taking part?

There is no anticipated risk associated with participation in this research. To reduce any potential risks, your personal data will be pseudonymised as appropriate.

6. Are there any benefits in taking part?

There will be no direct benefit to you from taking part in this research.

7. Expenses and payments

You will receive £5.

8. What happens to the data provided?

The information you provide during the study is the **research data**. Any research data from which you can be identified such as name, age, and audio recordings is known as **personal data**.

The **research data** and **personal / sensitive** data (including consent forms) will be transferred and stored confidentially in the internal drive in the Phonetics Lab, University of Oxford. Hard copies will be stored in the primary researcher's private premises. The recordings will be transcribed by the primary researcher. The researcher and supervisors will have access to research data. All research data and records will be stored for a minimum retention period of 3 years after publication or public release of the work of the research. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.

I would like your permission to use direct quotes anonymously or for your name to be attributed to these if you wish in any research outputs. I would like your permission to archive the audio recordings as a speech corpus and retain them for future research. I would like your permission to use anonymised data in future studies, and to share data with other researchers (e.g. in online databases). All personal information that could identify you will be removed or changed before information is shared with other researchers or results are made public.

9. Will the research be published?

The research may be published in conference presentations and scientific journals, but this may be two to three years after the end of the study. If you consent to this, direct quotes from the interviews may be included in the thesis or in presentations and publications, in order to illustrate linguistic points. Proper names will be anonymised and no information will be included that might allow participants to be identified.

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

If you agree to participate in this study, the research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible.

10. Who has reviewed this study?

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee (reference number: R67136/RE001).

11. *Who do I contact if I have a concern about the study or I wish to complain?*

If you have a concern about any aspect of this study, please speak to Chenzi Xu (chenzi.xu@ling-phil.ox.ac.uk) or her supervisor Prof. John Coleman (john.coleman@phon.ox.ac.uk), and we will do our best to answer your query. I should acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the relevant chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter in a reasonably expeditious manner:

Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee**; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD; Email: ethics@socsci.ox.ac.uk.

12. Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the study. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest.
Further information about your rights with respect to your personal data is available from
https://compliance.web.ox.ac.uk/individual-rights.

13. Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Chenzi Xu
Faculty of Linguistics, Philology and Phonetics
149 Walton St, Oxford, OX1 2HG, UK
Tel: +44 (0)1865 270444
Email: chenzi.xu@ling-phil.ox.ac.uk

A.3 Information Sheet for Parents / Guardians

The following pages display the participant information sheet for parents or guardians, since most of the Plastic Mandarin participants were below 18 years old. All of the participants were above 16 years old, hence informed consents from parents or guardians were not required. The Chinese version A.3.1 was used in the field trip and the English version A.3.2 is provided here for reference.

A.3.1 Information Sheet for Parents / Guardians (Chinese)

牛津大学语言学文字学及语音学学院

FACULTY OF LINGUISTICS, PHILOLOGY AND PHONETICS
UNIVERSITY OF OXFORD

Professor John Coleman
Phonetics Laboratory, 41 Wellington Square, Oxford OX1 2JF
Chenzi Xu [DPhil student]
+44 (0)1865 270444
chenzi.xu@ling-phil.ox.ac.uk

湖南"塑普"的语言学研究

家长 / 监护人须知

牛津大学研究操守委员会批准编号: R67136/RE001

尊敬的家长:

本人是英国牛津大学语言学博士研究生，正在做一项关于中文语音的学术研究，并将在长沙市南雅中学收集湖南"塑料普通话"和长沙话的语料。我希望能邀请您的孩子参与是项研究，希望您能对此研究给予支持 。在此请您阅读下文，了解研究概况。

研究目标

每一种语言都有其独特之处。湖南"塑料普通话"作为一种很年轻的地方口音，现在已经越来越普及，尤其是学生群体在学校的交流运用。这项研究旨在探索湖南"塑普"的一些语言特性。在南雅中学，我们将采集一些地道的"塑普"语音数据。 这项研究将有助于我们更好地认识"塑普"，加深我们对语言的接触与变迁的理解。

这项研究由牛津大学语言学文字学及语音学学院提供资金。

若想了解更多研究咨询，请发送电邮至 chenzi.xu@ling-phil.ox.ac.uk ，联系主要研究员徐晨子。

为什么我的孩子被邀请参与研究?

只要您的孩子对此研究有兴趣，达到 16 岁或 16 岁以上，长期生活在长沙，能够讲"塑料普通话"和长沙话（同时会讲标准普通话更好），就可以参与研究。我们希望在长沙市南雅中学招募约 16 至 20 名年轻志愿者。

我的孩子必须参加吗?

不是。若您的孩子有任何疑虑，欢迎向研究员提问。如果您的孩子同意参加，他/她亦有权随时退出，无需提供任何原由，也不会有任何形式的惩罚。退出时，只需提前告知研究员。

如果我孩子参加，他/她要做什么?

您的孩子将被邀请和他/她的同学一起参加一个语音活动环节，大约持续 30 到 45 分钟。活动将会在长沙市南雅中学的一间隔音条件较好的音乐教室进行。活动正式开始前，一些关于语言背景和年龄等的个人信息将被采集。活动中，在获得孩子的同意后，他/她将会带上便携式麦克风并且全程进行录音。在您的孩子的同意下，我们可能拍摄照片记录语音数据采集过程。所采集的语音资料完全匿名并保持机密，有可能用作之后的学术研究。

参加的好处和坏处是什么?

参加此项研究不存在潜在的利益或风险。

如何处理研究所收集的数据?

在参与研究过程中您的孩子所提供的信息是研究数据。任何可以识别您孩子身份的研究数据包括姓名、年龄、和录音被称作个人数据。

所有研究数据和个人/敏感数据（包括同意书）将会安全转移并机密地储存在牛津大学语音实验室的内部硬盘中。纸质填写的资料将会安全地储存在研究员的私人区域。录音将由研究员转录为文字。研究员及其导师将能获取此研究数据。所有研究数据和记录将在论文发表或研究工作公开发布后至少保留 3 年。牛津大学责任方成员可能获取获取此研究数据，以便监察是项研究。

我希望您的孩子能允许我们在研究发表中直接引用他/她所说的，以匿名的方式或者如果他/她希望我们标明出处（即您的姓名）。我希望您的孩子能同意我们将所收集的语音数据保留并做成语料库，以便未来的研究。我也希望您的孩子能同意我们在未来的研究项目中匿名使用这些数据，并且共享给其他研究者（例如以线上数据库的方式）。所有可以识别您孩子的个人信息将会在数据共享给其他研究者或者成果公开后移除或更改。

这项研究会发表么?

我们希望在学术会议或科学期刊中发表研究成果，这个过程可能是在研究结束后的两到三年内。若您的孩子同意参加，您的孩子在活动当中的话可能会被直接引用在论文，会议陈述，或学术发表中，

用来举例说明某种语言现象。专有名称将会匿名处理，不会出现任何可能暴露参加者的信息。所有研究数据和记录在论文发表或研究工作公开发布后将至少会保留 3 年。

牛津大学致力于为社会传播它的研究，因此建立了一个在线研究材料档案库。这个档案库中包含了成功提交至牛津大学的硕博论文的电子版。这样一个在线档案库给研究者提供了便捷的方式获取论文全文的免费阅读，以便于加强研究的影响和使用。该论文有可能在开放获取的期刊上发表。在发表或存档的文章中，将无法辨认参与者的身份。

谁组织是项研究?

是项研究主要由牛津大学语言学硕士研究生徐晨子负责，由牛津大学语言学、文字学与语音学学院提供资助。研究已经通过牛津大学研究操守委员会的审批， 编号为 R67136/RE001。

如果有问题怎么办?

如果您对此研究项目有任何疑问，请联系研究员徐晨子 (chenzi.xu@ling-phil.ox.ac.uk) 或者她的导师 John Coleman 教授(john.coleman@phon.ox.ac.uk) , 我们会尽力回答您的问题。研究员通常会在 10 个工作日内确认您的疑虑并回复您我们将如何处理。如果您对答复不满意或想进行法律起诉，请联系牛津大学社会科学与人文学科研究操守委员会的主席，他将尽快解决事务（电邮: ethics@socsci.ox.ac.uk; 地址: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD ）。

下一步我该做什么?

谢谢您对是项研究的支持。请记住您的孩子有权随时退出，无需提供任何原由，退出时请提前告知研究员。若您想了解关于研究的更多资讯（或者之后有任何相关的问题），请联系:

徐晨子
牛津大学语言学、文字学和语音学学院
地址：149 Walton St, Oxford, OX1 2HG, UK
电话: +44 (0)1865 270444
电邮: chenzi.xu@ling-phil.ox.ac.uk

A.3.2 Information Sheet for Parents / Guardians (English)

Faculty of Linguistics, Philology and Phonetics

PI Name: Professor John Coleman
Primary Researcher: Chenzi Xu [DPhil Student]
+44 (0)1865 270444
chenzi.xu@ling-phil.ox.ac.uk

Characterising Mandarin Accents

INFORMATION SHEET FOR PARENTS / GUARDIANS

Central University Research Ethics Committee (CUREC) Approval Reference: R67136/RE001

In partnership with researchers at the University of Oxford, Changsha Nanya Middle School has agreed to take part in a study investigating the characteristics of 'Plastic Mandarin'. We would like to invite your child to be part of this study. We very much hope you would like your child to take part. If your child decides to take part in this study, it is important that you understand why the study is being done and what it will involve.

What are we trying to find out?

Each spoken language has its own sound patterns and properties. This project is interested in the acoustic properties of different Mandarin accents. One of them is "Plastic Mandarin", that is, the accent of young people's Mandarin in Changsha, Hunan Province. In Nanya Middle School we will collect audio recordings of authentic speech data of Plastic Mandarin from high school students. This project will help us understand what happens when languages are in contact with each other and how new varieties of speech emerge.

More information about the project can be obtained by contacting the DPhil student researcher Chenzi Xu at chenzi.xu@ling-phil.ox.ac.uk.

Why has my child been invited to take part?

We are inviting your child to take part because they are a young person aged 16 years old or above, a long-term Changsha resident, able to speak Plastic Mandarin and Changsha, and attending Changsha Nanya Middle School. We are inviting approximately 16-20 young people to take part.

Does my child have to take part?

No. Your child can ask questions about the study before deciding whether or not to participate. Your child may withdraw their data from the study at any time, without giving a reason and without penalty, by advising the researchers of this decision.

What will happen if my child takes part?

Your child will be invited to a semi-structured pair recording session together with a friend or a classmate, which will last for around 30-45 minutes. Before the interview, some personal information such as age and language background will be collected. He or she will wear a microphone with their speech being recorded during the session, which will be held in a music room in the Arts Centre in Nanya Middle School. There may be a follow-up session if the recordings are not in good conditions. Their speech data will be kept confidential, and they may be used in further study on speech.

What are the advantages / disadvantages of taking part?
There is no anticipated benefit or risk associated with participation in this research.

What happens to the data provided?
The information you provide during the study is the **research data**. Any research data from which you can be identified such as name, age, and audio recordings is known as **personal data.**

The **research data** and **personal / sensitive** data (including consent forms) will be transferred and stored confidentially in the internal drive in the Phonetics Lab, University of Oxford. Hard copies will be stored in the primary researcher's private premises. All research data and records will be stored for a minimum retention period of 3 years after publication or public release of the work of the research. Any data transfer will be done securely and with a similar level of data protection as required under UK law.

The researcher and supervisor will have access to the research data. Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the research.
With your child's permission, we may archive the audio recordings as a speech corpus and retain them indefinitely. We may use the audio recordings and transcriptions in future studies, and share data with other researchers (e.g. in online databases) with your child's consent. Given your child's consent, we may use direct quotes anonymously or for their name to be attributed to these if they wish in any research outputs. All personal information that could identify your child will be removed or changed before information is shared with other researchers or results are made public.

Regular summaries of our findings will be given to the school and will be available to interested families.

Will the research be published?
The research may be published in conference presentations and scientific journals, but this may be two to three years after the end of the study.

The University of Oxford is committed to the dissemination of its research for the benefit of society and the economy and, in support of this commitment, has established an online archive of research materials. This archive includes digital copies of student theses successfully submitted as part of a University of Oxford postgraduate degree programme. Holding the archive online gives easy access for researchers to the full text of freely available theses, thereby increasing the likely impact and use of that research.

The research will be written up as a thesis. On successful submission of the thesis, it will be deposited both in print and online in the University archives to facilitate its use in future research. If so, the thesis will be openly accessible.

Who is conducting this research?
The research project is organized by Chenzi Xu of Oxford University, who is a DPhil Student. The research is funded by the Faculty of Linguistics, Philology and Phonetics. This study has been reviewed by and received ethics clearance through the University of Oxford's Central University Research Ethics Committee [R67136/RE001].

What if there is a problem?

If you have a concern about any aspect of this study, please speak to Chenzi Xu (chenzi.xu@ling-phil.ox.ac.uk) or her supervisor Prof. John Coleman (john.coleman@phon.ox.ac.uk), and we will do our best to answer your query. I should acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the relevant chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter in a reasonably expeditious manner:

Chair, **Social Sciences & Humanities Inter-Divisional Research Ethics Committee**; Address: Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD; Email: ethics@socsci.ox.ac.uk

What should I do next?

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Chenzi Xu
Faculty of Linguistics, Philology and Phonetics
149 Walton St, Oxford, OX1 2HG, UK
Tel: +44 (0)1865 270444
Email: chenzi.xu@ling-phil.ox.ac.uk

A.4 Participant Consent Form

The following pages display the participant consent form in Standard Written Chinese (A.4.1) and in English (A.4.2). These were provided to the participants when they arrived at the recording premises.

Faculty of Linguistics, Philology and Phonetics
University of Oxford

A.4.1 Participant Consent Form (Chinese)

PI Name: Professor John Coleman
Primary Researcher: Chenzi Xu [DPhil Student]
+44 (0)1865 270444
chenzi.xu@ling-phil.ox.ac.uk

志愿者同意书

牛津大学研究操守委员会批准编号: R67136/RE001

湖南"塑普"的语言学研究

研究目的: 是项研究旨在探索湖南"塑普"的一些语言特性。

		请在方框内签上您的姓氏
1	我确认我已阅读并理解《志愿者须知》，且有机会向研究员提问并收到满意答复。	
2	我明白参与是自愿属性，我有权随时退出，且无需提供任何原由，也不受任何不良影响。	
3	我了解所采集的研究数据（问卷及录音）可能会被牛津大学研究相关指定人员查看（例如研究员的导师）。我允许这类相关人员获取这些数据。	
4	我了解这个研究项目已经通过了牛津大学研究操守委员会的研究道德批审。	
5	我了解所采集个人信息的使用者、保存方法以及保留时限。	
6	我了解此项研究将被写成论文并且发表。	
7	我知道如何提出问题或投诉。	
8	我同意被录音。	
9	我了解录音数据将会如何用到研究成果中。	

10a	我同意在研究发表中直接引用我的话(标注我的名字)，**或者**	☐
10b	我同意在研究发表中匿名引用我的话。	☐
11	**我愿意成为这项研究的志愿者。**	☐
12	我在此授予研究员我所贡献的数据的所有版权，这些数据可能会在这个研究项目以及未来的项目中使用。	☐
可选择:	如若需要，我同意被拍照。	☐
可选择:	我同意把这项研究所收集的研究数据给研究员做成语料库，可用于其他相关研究。我明白任何资料离开当前研究小组后将会完全匿名化，我将不会被识别。	☐
可选择:	我同意我所提供的个人信息将在一个安全的数据库保存，以便日后相关研究联系。	☐

______________	dd / mm / yyyy	______________
志愿者姓名	日期	签名

Chenzi Xu 徐晨子	dd / mm / yyyy	______________
研究员	日期	签名

A.4.2 Participant Consent Form (English)

PI Name: Professor John Coleman
Primary Researcher: Chenzi Xu [DPhil Student]
+44 (0)1865 270444
chenzi.xu@ling-phil.ox.ac.uk

PARTICIPANT CONSENT FORM

CUREC Approval Reference: R67136/RE001

Characterising Mandarin Accents

Purpose of Study: To characterise linguistic properties of different Mandarin varieties.

		Please initial each box
1	I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
2	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without any adverse consequences or academic penalty.	
3	I understand that research data collected during the study may be looked at by designated individuals from the University of Oxford where it is relevant to my taking part in this study. I give permission for these individuals to access my data.	
4	I understand that this project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.	
5	I understand who will have access to personal data provided, how the data will be stored and what will happen to the data at the end of the project.	
6	I understand how this research will be written up and published.	
7	I understand how to raise a concern or make a complaint.	
8	I consent to being audio recorded	
9	I understand how audio recordings will be used in research outputs	
10	I consent to having my photo taken if needed	
11a	I agree to the use of direct quotes, attributed to my name, in research outputs **OR**	

11b	I agree to the use of anonymised quotes in research outputs	☐
12	I agree to take part in the above study.	☐
13	I hereby assign to the researcher all copyright in my contribution for use in all work stemming from this project and future projects.	☐
Optional:	I agree for research data collected in this study to be formatted as a speech corpus which may be used in other future research studies. I understand that any data that leave the research group will be fully anonymised so that I cannot be identified.	☐
Optional:	I agree for my personal data to be kept in a secure database for the purpose of contacting me about future studies.	☐

______________________	dd / mm / yyyy	______________________
Name of Participant	Date	Signature
______________________	dd / mm / yyyy	______________________
Name of person taking consent	Date	Signature

A.5 Informal Questionnaire

The following pages show the original informal questionnaires in Standard Written Chinese for the Plastic Mandarin speakers and the Standard Mandarin speakers respectively. English translation is provided in grey floating text annotations for reference. The former (A.5.1) was distributed in hard copies and the latter (A.5.2) was distributed through Google Forms.

Faculty of Linguistics, Philology and Phonetics
University of Oxford

A.5.1 Participant Form (PM Group)

PI Name: Professor John Coleman
Primary Researcher: Chenzi Xu [DPhil Student]
+44 (0)1865 270444
chenzi.xu@ling-phil.ox.ac.uk

湖南"塑普"的语言学研究

志愿者信息表

牛津大学研究操守委员会(CUREC)批准编号: R67136/RE001

姓名: Name		年龄: Age	
籍贯: Birth Place		民族: Nation	
在长沙生活了 年 I have lived in Changsha for ____ years.			

请自我评估你的语言水平: Self-evaluation of spoken language proficiency

语言	基本不会 (Not at all)	会一点	一般	良好	优秀 (Fluent)
标准普通话 Standard Mandarin	①	②	③	④	⑤
长沙话 Changsha	①	②	③	④	⑤
其他方言 :Other Chinese varieties	①	②	③	④	⑤

（其他方言没有可不填）

Please indicate the frequency of speaking Standard Mandarin, Plastic Mandarin and other dialects in the following scenarios (please specify the dialect) by filling in the circle.

请在以下的情景中填上使用标准普通话，塑料普通话和方言的频率：

总共分 5 类，由低到高。①代表"从不使用"；⑤代表"一直使用"。

②代表"偶尔使用"；③ 代表"约一半的时间在使用"；④ 代表"经常使用"。

	Never				Always
你和爷爷奶奶见面的时候说什么？（若不适用，可不填） I talk to grandparents in…					
标准普通话 Standard Mandarin	①	②	③	④	⑤
塑料普通话 Plastic Mandarin	①	②	③	④	⑤
方言: Other ________	①	②	③	④	⑤
在家的时候，你和父母说什么？ I talk to parents in…					
标准普通话	①	②	③	④	⑤
塑料普通话	①	②	③	④	⑤
方言: ________	①	②	③	④	⑤

At home, grandparents speak…

爷爷奶奶在家说什么？（若不适用，可不填；若爷爷奶奶使用语言不同，请标注）					
标准普通话 Standard Mandarin	①	②	③	④	⑤
塑料普通话 Plastic Mandarin	①	②	③	④	⑤
方言 1：Other	①	②	③	④	⑤
方言 2：Other (specify if grandma and grandpa speak different varieties)	①	②	③	④	⑤
在家的时候，父母说什么？（若父母使用语言不同，请标注）At home, parents speak…					
标准普通话	①	②	③	④	⑤
塑料普通话	①	②	③	④	⑤
方言 1：______	①	②	③	④	⑤
方言 2：Other (specify if mum and dad speak different varieties)	①	②	③	④	⑤

1 代表"从不使用"；2 代表"偶尔使用"；3 代表"约一半的时间在使用"；4 代表"经常使用"；5 代表"一直使用"。

Never　　　　Always

上课的时候，老师们说什么？ During lectures, teachers speak…					
标准普通话 Standard Mandarin	①	②	③	④	⑤
塑料普通话 Plastic Mandarin	①	②	③	④	⑤
方言：Other	①	②	③	④	⑤
上课回答问题或者作报告，你说什么？ During lectures/presentation, I speak…					
标准普通话	①	②	③	④	⑤
塑料普通话	①	②	③	④	⑤
方言：______	①	②	③	④	⑤
下课的时候，你和老师说什么？ After classes, I talk to teachers in…					
标准普通话	①	②	③	④	⑤
塑料普通话	①	②	③	④	⑤
方言：______	①	②	③	④	⑤
在学校的时候，你和同学说什么？ In school, I talk to classmates and friends in…					
标准普通话	①	②	③	④	⑤
塑料普通话	①	②	③	④	⑤
方言：______	①	②	③	④	⑤
学校外的日常生活中（买东西、打的等），你说什么？After school (e.g. taking taxi or shopping), I speak.					
标准普通话	①	②	③	④	⑤
塑料普通话	①	②	③	④	⑤
方言：______	①	②	③	④	⑤

谢谢参与！

A.5.2 Participant Form (SM Group)

语音实验志愿者招募

牛津大学研究操守委员会(CUREC)批准编号: R67136/RE001

（未分享） 切换帐号

*必填

姓名 * Name

您的回答

邮箱 * E-mail

您的回答

微信 WeChat ID (optional)

您的回答

年龄 * Age

您的回答

民族 * Nation / Ethnic group

您的回答

籍贯 * Birth Place

◯ 北京

◯ 其他:

请问您在北京生活了多少年? How long have you lived in Beijing?

您的回答

请问是否参加普通话水平测试？如有，您的等级是?

Have you ever attended the Putonghua Proficiency Test? If so, what is your grade?

您的回答

How would you evaluate your Putonghua/Standard Mandarin proficiency?

如果自我评估您的标准普通话水平，您会给自己打几分? *

	1	2	3	4	5	
基本不会 I barely speak any Putonghua.	◯	◯	◯	◯	◯	流畅 Fluent

提交

清除表单内容

切勿通过 Google 表单提交密码。

B
Speech Materials

Contents

B.1 A Full List of Speech Materials in the Word-guessing Game (B)

In the word-guessing game, each scenario is represented by an incomplete sentence, with the keyword (in bold here) missing. The Clue-receiver was trying to construct the following sentences, making use of the hints from their friend. The keywords are frequent words and all sentences make sense. Utterances with incorrect keywords also lead to repetitions of the carrier sentences, embedded with speech sequences containing neutral tones.

(1) i. 他们 堆的 雪人 很 稳当
tā-men tuī-de ***xuěrén*** *hen wěndang*
3PL built-DE snowman very stable
'The snowman they built is stable.'

ii. 他们 堆的 城堡 很 稳当
tā-men tuī-de ***chéngbǎo*** *hen wěndang*
3PL built-DE castle very stable
'The castle they built is stable.'

iii. 他们 堆的 台阶 很 稳当
tā-men tuī-de ***táijiē*** *hen wěndang*
3PL built-DE steps very stable
'The steps they built is stable.'

(2) i. 你们 搭的 积木 很 厉害
nǐ-men tā-de ***jīmù*** *hen lìhai*
2PL built-DE toy blocks very awesome
'The toy blocks you built is awesome.'

ii. 你们 搭的 雨棚 很 厉害
nǐ-men tā-de ***yǔpéng*** *hen lìhai*
2PL built-DE shelter very awesome
'The shelter you built is awesome.'

iii. 你们 搭的 舞台 很 厉害
nǐ-men tā-de ***wǔtái*** *hen lìhai*
2PL built-DE stage very awesome
'The stage you built is awesome.'

(3) i. 你 回的 信息 很 马虎
nǐ huí-de ***xìnxī*** *hen mǎhu*
2SG replied-DE message very careless
'The message you replied is careless.'

ii. 你 回的 邮件 很 马虎
nǐ huí-de ***yóujiàn*** *hen mǎhu*
2SG replied-DE email very careless
'The email you replied is careless.'

iii. 你 回的 礼物 很 马虎
nǐ huí-de ***lǐwù*** *hen mǎhu*
2SG returned-DE gift very careless
'The gift you returned is careless.'

(4) i. 我 答的 问卷 很 麻烦
wǒ dá-de ***wènjuàn*** *hen máfan*
1SG answered-DE survey very troublesome
'The survey I answered is a hassle.'

ii. 我 答的 问题 很 麻烦
wǒ dá-de ***wèntí*** *hen máfan*
1SG answered-DE questions very troublesome
'The questions I answered is a hassle.'

iii. 我 答的 试卷 很 麻烦
wǒ dá-de ***shìjuàn*** *hen máfan*
1SG answered-DE exam paper very troublesome
'The exam paper I answered is a hassle.'

(5) i. 这 裤子, 腿的 设计 很 漂亮
zhè ***kùzi****, tuǐ-de shèjì hen piàoliang*
This trousers, leg-GEN design very pretty
'This pair of trousers, the design of the legs is pretty.'

ii. 这 桌子, 腿的 设计 很 漂亮
zhè ***zhuōzi****, tuǐ-de shèjì hen piàoliang*
This table, leg-GEN design very pretty
'This table, the design of the legs is pretty.'

iii. 这 芭比, 腿的 设计 很 漂亮
zhè ***bābǐ****, tuǐ-de shèjì hen piàoliang*
This Barbie, leg-GEN design very pretty
'This Barbie doll, the design of the legs is pretty.'

(6) i. 他 打的 豆浆 很 稀罕
tā dǎ-de ***dòujiāng*** *hen xīhan*
3SG made-DE soy milk very rare
'The soy milk he made is rare.'

ii. 他 打的 领带 很 稀罕
tā dǎ-de ***lǐngdài*** *hen xīhan*
3SG made-DE tie very rare
'The tie he made is rare.'

iii. 他 打的 电话 很 稀罕
tā dǎ-de ***diànhuà*** *hen xīhan*
3SG made-DE call very rare
'The call he made is rare.'

(7) i. 我们 兑的 奖品 很 实在
wōmen duì-de ***jiǎngpǐn*** *hen shízai*
1PL redeemed-DE prize very worthwhile
'The prize we redeemed is worthwhile.'

ii. 我们 兑的 机票 很 实在
wōmen duì-de ***jīpiào*** *hen shízai*
1PL redeemed-DE flight ticket very worthwhile
'The flight ticket we redeemed is worthwhile.'

iii. 我们 兑的 书籍 很 实在
wōmen duì-de ***shūjí*** *hen shízai*
1PL redeemed-DE books very worthwhile
'The books we redeemed are worthwhile.'

(8) i. 大的 屋子 很 舒服
dà-de ***wūzi*** *hen shūfu*
big-DE house very comfy
'Big house is comfy.'

ii. 大的 沙发 很 舒服
dà-de ***shāfā*** *hen shūfu*
big-DE sofa very comfy
'Big sofa is comfy.'

iii. 大的 娃娃 很 舒服
dà-de ***wáwa*** *hen shūfu*
big-DE doll very comfy
'Big doll is comfy.'

(9) i. 奶奶们 正 看着 电视
nǎinai-men zhèng kàn-zhe ***diànshì***
grandma-PL PROG see-DUR television
'Grannies are watching television.'

ii. 奶奶们 正 看着 相片
nǎinai-men zhèng kàn-zhe ***xiàngpiàn***
grandma-PL PROG see-DUR photo
'Grannies are looking at photo(s).'

iii. 奶奶们 正 看着 扇子
nǎinai-men zhèng kàn-zhe ***shànzi***
grandma-PL PROG see-DUR fan
'Grannies are looking at fan(s).'

(10) i. 爷爷们 正 赶着 牛群
yéye-men zhèng gǎn-zhe ***niúqún***
grandpa-PL PROG drive-DUR a herd of cattle
'Grandpas are herding the cattle.'

ii. 爷爷们 正 赶着 苍蝇
yéye-men zhèng gǎn-zhe ***cāngying***
grandpa-PL PROG drive-DUR flies
'Grandpas are whisking the flies off.'

iii. 爷爷们 正 赶着 鸭子
yéye-men zhèng gǎn-zhe ***yāzi***
grandpa-PL PROG drive-DUR ducks
'Grandpas are herding the ducks.'

(11) i. 妹妹们 正 弹着 吉他
mèimei-men zhèng tán-zhe ***jítā***
sister-PL PROG play-DUR guitar
'Sisters are playing guitars.'

ii. 妹妹们 正 弹着 古筝
mèimei-men zhèng tán-zhe ***gǔzhēng***
sister-PL PROG play-DUR guzheng
'Sisters are playing guzheng (a Chinese instrument).'

iii. 妹妹们 正 弹着 钢琴
mèimei-men zhèng tán-zhe ***gāngqín***
sister-PL PROG play-DUR piano
'Sisters are playing the piano.'

(12) i. 妈妈们 正 看着 球门
māma-men zhèng kān-zhe ***qiúmén***
mum-PL PROG guard-DUR goal
'Mums are guarding the goal.'

ii. 妈妈们 正 看着 钱包
māma-men zhèng kān-zhe ***qiánbāo***
mum-PL PROG guard-DUR wallet
'Mums are guarding the wallet.'

iii. 妈妈们 正 看着 单车
māma-men zhèng kān-zhe ***dānchē***
mum-PL PROG guard-DUR bike
'Mums are guarding the bike.'

(13) i. 妹妹们 的 树苗 高了
mèimei-men de ***shùmiáo*** *gāo-le*
sister-PL GEN sapling (grow) tall-PERF
'Sisters' sapling has grown taller.'

ii. 妹妹们 的 风筝 高了
mèimei-men de ***fēngzheng*** *gāo-le*
sister-PL GEN kite (grow) tall-PERF
'Sisters' kite has flown higher.'

iii. 妹妹们 的 竹子 高了
mèimei-men de ***zhúzi*** *gāo-le*
sister-PL GEN bamboo (grow) tall-PERF
'Sisters' bamboo has grown taller.'

(14) i. 爷爷们 的 鹦鹉 逃了
yéye-men de ***yīngwǔ*** *táo-le*
grandpa-PL GEN parrot escape-PERF
'Grandpa' parrot has escaped.'

ii. 爷爷们 的 兔子 逃了
yéye-men de ***tùzi*** *táo-le*
grandpa-PL GEN rabbit escape-PERF
'Grandpa' rabbit has escaped.'

iii. 爷爷们 的 乌龟 逃了
yéye-men de ***wūguī*** *táo-le*
grandpa-PL GEN turtle escape-PERF
'Grandpa' turtle has escaped.'

(15) i. 奶奶们 的 猫咪 老了
nǎinai-men de ***māomī*** *lǎo-le*
grandma-PL GEN cat (grow) old-PERF
'Grannies' cat has grown older.'

ii. 奶奶们 的 驴子 老了
nǎinai-men de ***lǘzi*** *lǎo-le*
grandma-PL GEN donkey (grow) old-PERF
'Grannies' donkey has grown older.'

iii. 奶奶们 的 母鸡 老了
nǎinai-men de ***mǔjī*** *lǎo-le*
grandma-PL GEN hen (grow) old-PERF
'Grannies' hen has grown older.'

(16) i. 妈妈们 的 围巾 到了
māma-men de ***wéijīn*** *dào-le*
mum-PL GEN scarf arrive-PERF
'Mums' scarf has arrived.'

ii. 妈妈们 的 苹果 到了
māma-men de ***píngguǒ*** *dào-le*
mum-PL GEN apple arrive-PERF
'Mums' apple has arrived.'

iii. 妈妈们 的 雨伞 到了
māma-men de ***yǔsǎn*** *dào-le*
mum-PL GEN umbrella arrive-PERF
'Mums' umbrella has arrived.'

B.2 The Path Code Game (C)

B.2.1 Procedure

The picture-based path code game was the third step (Task C) of the recording session. This game was developed considering a wealth of noun phrases containing neutral tone syllables. The pair of subjects were given cards, on each of which there were four items arranged in the four corners, illustrated in Figure B.1. Subjects drew a path that passed every item once independently, then they described their path pattern on each card to one another in a designated manner, followed by an open discussion through which they needed to pick "the odd one out" among the four on each card. There is no correct answer in the "odd one", as long as they justify their answers. Usually the noun phrases were repeated in natural spontaneous utterances in their justification.

They described their path patterns with a specified carrier sentence, demonstrated in Table B.1. The technique ensures two repetitions of each item: once at the beginning of a clause, and once at the end of a clause. Some of these items were mentioned more times in the subsequent spontaneous discussion. Since item names may be varied, each item was introduced once by the interviewer before the game started. In Figure B.1, the four items are *tīzi* [t^hi tsz̩] 'ladder', *lízi* [li tsz̩] 'pear', *lǐzi* [li tsz̩] 'plum', and *lìzi* [li tsz̩] 'chestnut'. The last three words share the same segments, but the first syllable differs in lexical tone. All four words contain the same neutral tone syllable [tsz̩], a noun suffix.

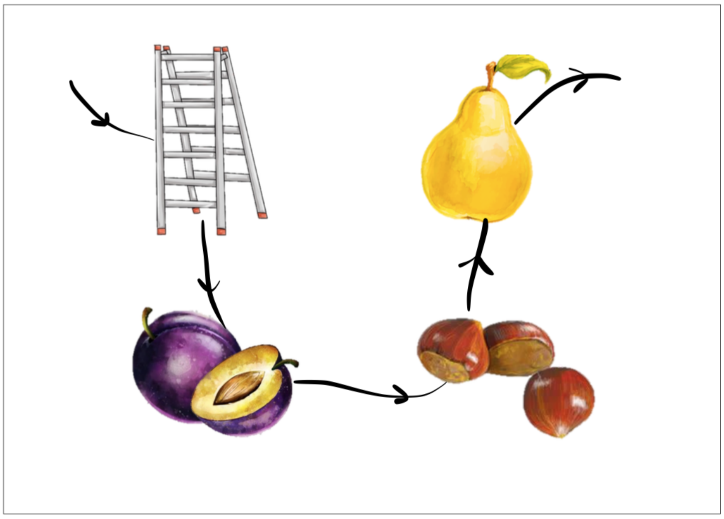


Figure B.1: An example prompt used in the path code game. The four items *tīzi* 'ladder', *lízi* 'pear', *lǐzi* 'plum', and *lìzi* 'chestnut' are arranged in the four corners, and a random path is drawn in black.

Transcript	Translation
起点是梯子。	'The starting point is the ladder.'
梯子的后面是李子。	'After the ladder is the plum.'
李子的后面是栗子。	'After the plum is the chestnut.'
栗子的后面是梨子。	'After the chestnut is the pear.'
梨子结束。	'The pear is the end.'

Table B.1: The carrier sentence of describing the path pattern, demonstrated with the path in Figure B.1. The item fields are underlined.

A scoring system was incorporated as part of the gamification design in this task during the fieldwork visits in the school in order to increase the engagement of high school students. Each pair of subjects had a score, in fact depending principally on luck: the more similar their path codes, the higher their score. The pair with the highest score received a surprise prize (a box of chocolate) in the end of my visit. The similarity of the path code was the number of matching items at given positions. The scoring system was not included in the recording sessions for older Standard Mandarin participants.

Given the uncertainty in the amount of game time and the overall time limit of a recording session, some pairs of subjects skipped the discussion of the most different item.

B.2.2 Disyllabic phrases with neutral tone in the speech materials

Another four sets of phonetically controlled disyllabic noun phrases centred on the nominal suffix *zi*, derived from zǐ 'child' (C. N. Li & Thompson, 1981) and originated as a diminutive suffix (Norman, 1988).

Set	T1-N	T2-N	T3-N	T4-N
/aʊ/	[taʊ tsz̩] *dāozi* 刀子 knife	[tʰaʊ tsz̩] *táozi* 桃子 peach	[tsaʊ tsz̩] *zǎozi* 枣子 date	[maʊ tsz̩] *màozi* 帽子 hat
/i/	[tʰi tsz̩] *tīzi* 梯子 ladder	[li tsz̩] *lízi* 梨子 pear	[li tsz̩] *lǐzi* 李子 plum	[li tsz̩] *lìzi* 栗子 chestnut
/iɛn/	[tɕʰiɛn tsz̩] *qiānzi* 签子 stick	[tɕʰiɛn tsz̩] *qiánzi* 钳子 plier	[tɕiɛn tsz̩] *jiǎnzi* 剪子 scissor	[tɕiɛn tsz̩] *jiànzi* 毽子 shuttlecock
/ɪŋ/	[tɪŋ tsz̩] *dīngzi* 钉子 nail	[tʰɪŋ tsz̩] *tíngzi* 亭子 pavilion	[ɪŋ tsz̩] *yǐngzi* 影子 shadow	[tɕɪŋ tsz̩] *jìngzi* 镜子 mirror

Table B.2: Disyllabic speech materials containing the neutral tone syllable *zi*

zi constitutes the obligatory second syllable of plentiful nouns of any semantic class, but is no longer used productively (C. N. Li & Thompson, 1981). These [X-zi] phrases are presented Table B.2, embedded in the path code game. Each set features a consistent rhyme in the X syllables. The *Modern Chinese Dictionary* (Institute of Linguistics CASS, 2016) was consulted to check that the morpheme *zi* canonically has the neutral tone pronunciation in these words.

Table B.3 contains words with a neutral tone syllable that has a citation tone in isolation.

The *zi* sets were not used in the analysis given that a large number of *zi* tokens are voiceless or with very short voicing portion, which was not ideal for pitch

T1-N	T2-N	T3-N	T4-N
[xi xan] *xīhan* 稀罕 rare	[ma fan] *máfan* 麻烦 complicated	[wən tɑŋ] *wěndang* 稳当 stable	[pʰiaʊ liɑŋ] *piàoliang* 漂亮 pretty
[ʂu fu] *shūfu* 舒服 comfortable	[ʂɻ̩ tsaɪ] *shízai*实在 real	[ma xu] *mǎhu*马虎 careless	[li xaɪ] *lìhai* 厉害 awesome
[fəŋ ʈʂŋ̩] *fēngzheng* 风筝 kite	[lwo po] *luóbo* 萝卜 carrot	[tsweɪ pa] *zuǐba* 嘴巴 mouth	[iaʊ ʂɻ̩] *yàoshi* 钥匙 key
[dəŋ lɔŋ] *dēnglong* 灯笼 lantern	[pʰu tʰaʊ] *pútao* 葡萄 grape	[piɛn tan] *biǎndan* 扁担 carrying pole	[toʊ fu] *dòufu* 豆腐 tofu
[ʈʂʰwɑŋ xu] *chuānghu* 窗户 window	[xu li] *húli* 狐狸 fox	[la pa] *lǎba* 喇叭 speaker	[tsʰz̩ weɪ] *cìwei* 刺猬 hedgehog

Table B.3: Extra disyllabic speech materials containing a neutral tone syllable. The second syllable in these words in isolation has a citation tone.

analysis. These speech materials can be used in future research, when we examine other aspects of neutral tone syllables.

C

Supplementary Statistics

Contents

C.1 Accent Profiling based on Influential Data Analysis

Sometimes influential data can drastically change the interpretation of results from a statistical model. Influential data analysis was performed as part of the model check in the present thesis, as suggested by Winter (2013). Cook's Distance, for example, estimates how much a regression model changes when a data point or a group of data points is excluded ("leave-one-out diagnostics").

Figures C.1 and C.2 present the Cook's Distance estimates of the LME models of Plastic Mandarin tones by speaker and by token respectively. Quite often some data points with a large Cook's Distance value are identified as influential data or "outliers". There are different thresholds or cut-off values for spotting influential data. One rule of thumb is to examine any data that is more than three times

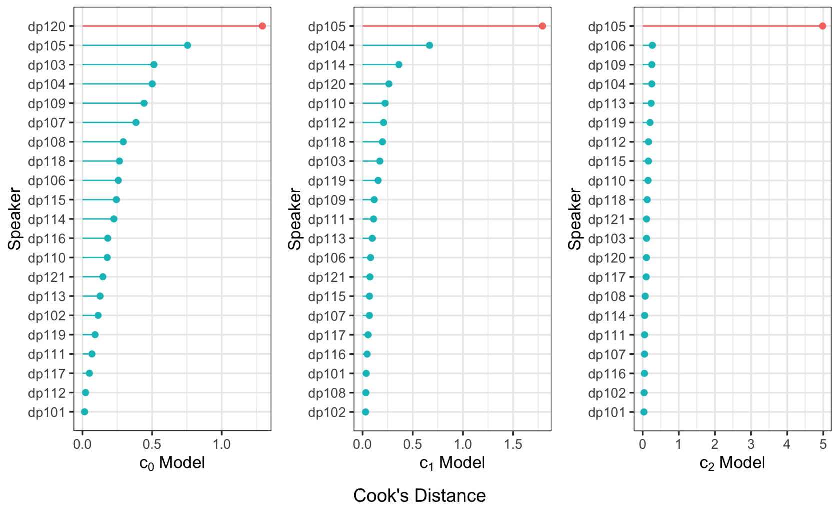


Figure C.1: Cook's Distance by speaker on the LME models of c_i coefficients of Plastic Mandarin tones.

the mean of all the distances. In Figure C.1, the influential speaker is marked in pink. In Figure C.2, the top ten most influential data is labelled.

The influential data analysis was **not** intended to just remove "outliers" and improve the model fit. Instead, it points us to a deep and thorough analysis of these data, to check if they are valid and to understand what might make them different. When there are no technical errors or artifacts in the data, large Cook's Distance values may flag certain individuals' accents or particular utterances. In Figure C.1, for instance, speaker 105 stands out from the group in terms of the slope and curvature of Plastic Mandarin tones, while speaker 120 stands out in terms of the average height of tones. If we look back at Figure 3.46 in §3.7.4, we find speaker 105's pitch patterns slightly different. We can further spot specific tokens that are influential in Figure C.2. Speaker 105's Tone 3 and Tone 4 tokens are influential with regard to the curvature of tones, and speaker 120 had some influential neutral tone tokens regarding the average height. In the analysis, these natural tokens were listened to and preserved, since they reveal some individual variation in speaking Plastic Mandarin.

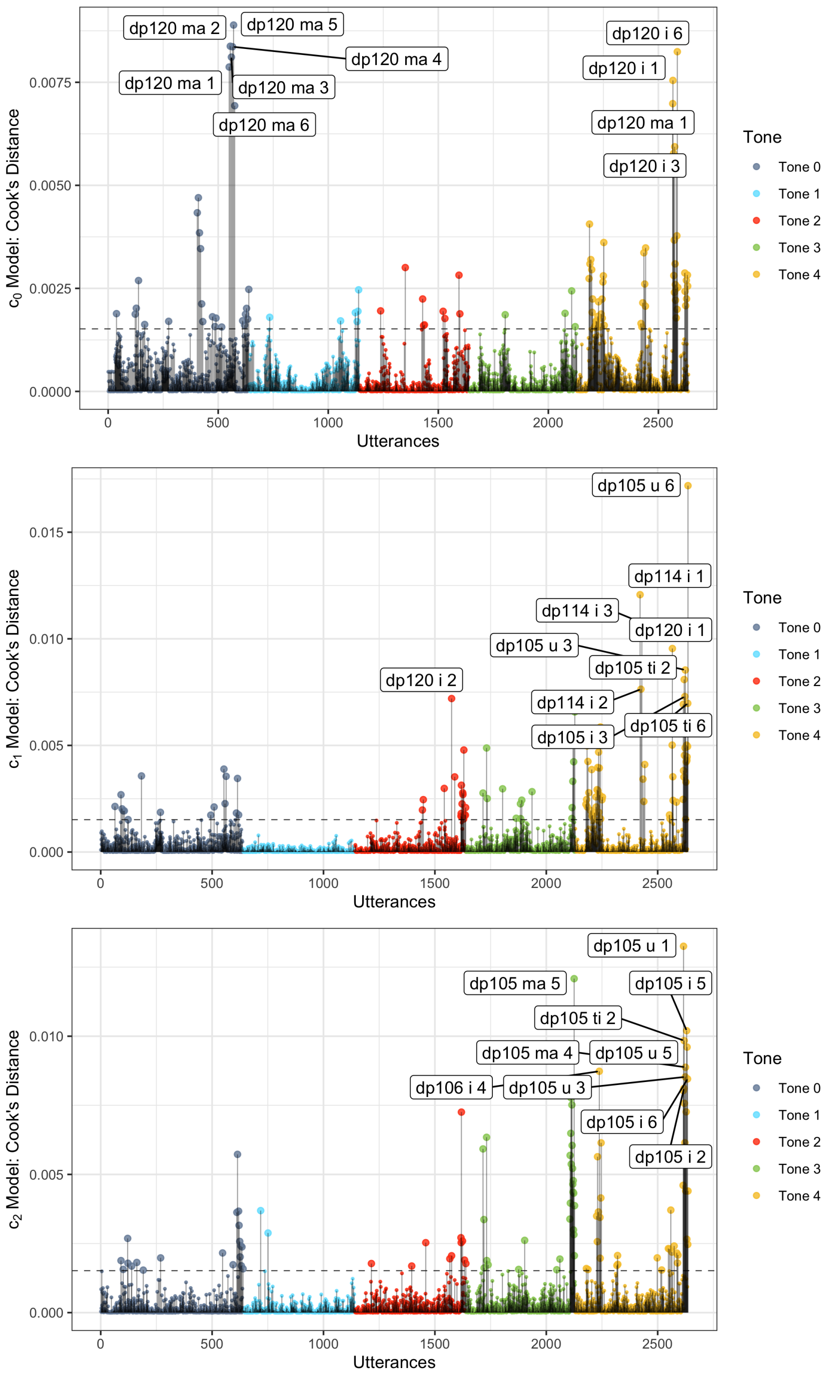


Figure C.2: Cook's Distance by token on the LME models of c_i coefficients of Plastic Mandarin tones. Top ten tokens with highest Cook's Distance were labelled in the format of '(task code) + (speaker ID) + (IPA of the syllable) + (the order in a sequence)'. Dotted line marks a threshold, estimated by $4/thetotalnumberoftokens$.

C.2 Dealing with Violations of Assumptions

Violations of assumptions sometimes occurred during the analysis, and these were detected via Quantile-Quantile plots (QQ-plots) and other residual plots. In the LME analysis of duration data in Chapter 5 for example, the distribution of residuals tend to be right-skewed (illustrated in Figure 5.5), which violated the assumption of the normality of residuals. Figure 5.5 also shows the violation of homoskedasticity.

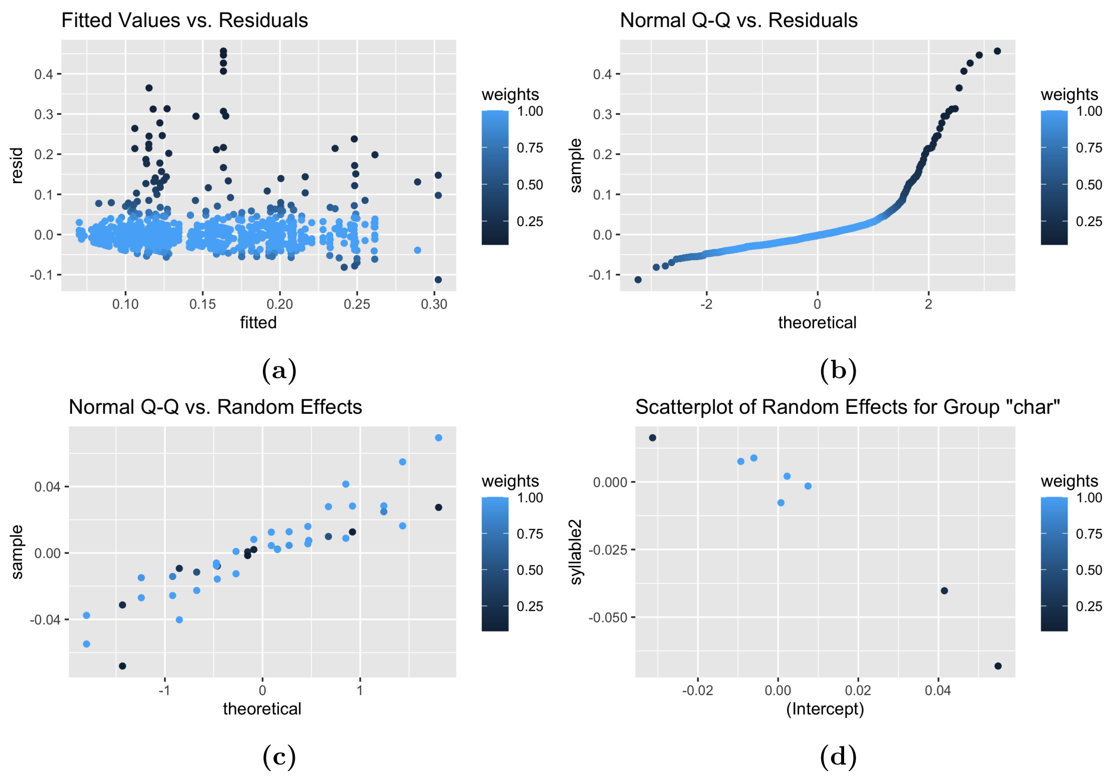


Figure C.3: Residual plots of the robust LME model 1 for Standard Mandarin data. The colouring of blue indicates the robustness weights.

Some of the models were re-fit using generalised mixed-effects models with non-normal continuous probability distributions that are right-skewed such as the lognormal[1] distribution. The lognormal models still had violations of assumptions. Hence, this thesis adopted the robust methods (Koller, 2016). Figure C.3 illustrates the residual plots for robust fit of model 1) for Standard Mandarin data, in which the colouring of blue indicates the robustness weights. The deep blue data points were down-weighted in the robust model. In addition, a study of the robustness

[1]Gamma distribution may be another option.

weights, similar to influential data analysis, gave us insights as to which observations cause the modelling differences.

D

Software and Packages

Contents

D.1 Audio Processing

In this thesis, the following software and packages were used for audio processing: Praat (Jadoul et al., 2018), ESPS (Entropic Research Laboratory, 2006), SoX (Bagwell et al., 2015), HTK (HTK Team, 2006), and Python 3 (van Rossum & Drake, 2009) with standard modules including `pandas`, `numpy`, `os`, `sys`, `glob`, and relevant packages `Parselmouth` (Jadoul et al., 2018), `praat-textgrids` (Nieminen, 2020). I have written a modified Python 3 script of P2FA (Yuan & Liberman, 2008) for Mandarin and multiple Praat scripts to automate some processes.

D.2 Data Wrangling, Visualisation, and Statistical Analysis

Data wrangling, visualisation, and statistical analysis were conducted in R (R Core Team, 2021) and Rstudio (RStudio Team, 2022) with relevant packages: `rmarkdown` (Allaire et al., 2022), `tidyverse` (Wickham & RStudio, 2021), `magrittr` (Bache et al., 2022), `modelr` (Wickham & RStudio, 2020),`ggpubr` (Kassambara, 2020), `ggdist` (Kay & Wiernik, 2022), *ggrepel* (Slowikowski et al., 2021), `ggridges` (Wilke, 2020), `ggsci` (N. Xiao & Li, 2018), `plotly` (Sievert et al., 2021), `gridExtra` (Auguie & Antonov, 2017), `lme4` (Bates, Mächler, et al., 2015), `lmerTest` (Kuznetsova et al., 2017), `pbkrtest` (Halekoh & Højsgaard, 2021), `emmeans` (Lenth, 2021), `broom.mixed` (Bolker et al., 2022), `mgcv` (S. Wood, 2022), `itsadug` (van Rij et al., 2022), `tidymv` (Coretta et al., 2022).

D.3 Thesis Writing

The thesis was written in LaTeX, using the Oxford Thesis Template which was created by Keith Gillow and modified by Sam Evans and John McManigle (available at https://www.oxfordechoes.com/oxford-thesis-template/). The following packages were used: `babel`, `csquotes`, `biblatex`, `booktabs`, `multirow`, `tabularx`, `makecell`, `pinyin`, `CJKutf8`, `tikz`, `caption`, `subcaption`, `cleveref`, `inputenc`, `tipa`, `gb4e`. They are available from the Comprehensive TeX Archive Network (https://www.ctan.org/).

Works Cited

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2022). *Rmarkdown: Dynamic documents for R* (Version 2.14.3). Retrieved July 5, 2022, from https://rmarkdown.rstudio.com

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, *34*(4), 351–366. Retrieved May 9, 2022, from https://www.proquest.com/docview/1299110515/citation/E595D9E63D4D4D97PQ/1

Andruski, J. E., & Costello, J. (2004). Using polynomial equations to model pitch contour shape in lexical tones: An example from Green Mong. *Journal of the International Phonetic Association*, *34*(2), 125–140.

Aston, J. A. D., Chiou, J.-M., & Evans, J. P. (2010). Linguistic pitch analysis using functional principal component mixed effect models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(2), 297–317.

Asu, E. L., Lippus, P., Salveste, N., & Sahkai, H. (2016). F0 declination in spontaneous Estonian: Implications for pitch-related preplanning in speech production. *Speech Prosody 2016*, 1139–1142.

Auer, P. (1998). Dialect levelling and the standard varieties in Europe. *Folia Linguistica*, *32*(1-2), 1–9.

Auer, P. (2011, June 24). Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In *Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations* (pp. 7–42). De Gruyter Mouton.

Auguie, B., & Antonov, A. (2017, September 9). *gridExtra: Miscellaneous Functions for "Grid" Graphics* (Version 2.3). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=gridExtra

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*(5), 3048–3058.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 23.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. N. (2016, January 8). *Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models*. Retrieved February 2, 2022, from http://arxiv.org/abs/1601.02043

Baayen, R. H., Vasishth, S., Kliegl, R., & Bates, D. (2017). The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language*, *94*, 206–234.

Bache, S. M., Wickham, H., Henry, L., & RStudio. (2022, March 30). *Magrittr: A Forward-Pipe Operator for R* (Version 2.0.3). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=magrittr

Baghai-Ravary, L., Kochanski, G., & Coleman, J. (2009). Precision of phoneme boundaries derived using hidden Markov models. *Interspeech 2009*, 2879–2882.

Bagwell, C., Rullgard, M., Robs, & Klauer, U. (2015, February 22). *SoX - Sound eXchange* (Version 14.4.2). Retrieved June 7, 2022, from http://sox.sourceforge.net/Main/HomePage

Bao, H. X., & Chen, H. (2005). The Classification of Xiang group. *Fang yan*, *3*, 261–270.

Bao, Z. (1999). *The structure of tone*. Oxford University Press. Retrieved August 11, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://ebookcentral.proquest.com/lib/oxford/detail.action?docID=430646

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models, 27.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects models using lme4. *Journal of Statistical Software*, *67*(1).

Beijing University. (1995). *Hanyu fangyan cihui* 汉语方言词汇*(Chinese dialects vocabularies)* (2nd ed.). Yuwen chubanshe.

Berez-Kroeker, A. L., Andreassen, H. N., Gawne, L., Holton, G., Kung, S. S., Pulsifer, P., Collister, L. B., the Data Citation and Attribution in Linguistics Group, & the Linguistics Data Interest Group. (2018). The {Austin} Principles of Data Citation in Linguistics. Retrieved July 5, 2022, from https://site.uit.no/linguisticsdatacitation/austinprinciples/

Berez-Kroeker, A. L., Gawne, L., Kung, S. S., Kelly, B. F., Heston, T., Holton, G., Pulsifer, P., Beaver, D. I., Chelliah, S., Dubinsky, S., Meier, R. P., Thieberger, N., Rice, K., & Woodbury, A. C. (2018). Reproducible research in linguistics: A position statement on data citation and attribution in our field. *Linguistics*, *56*(1), 1–18.

Blackburn, C. S., & Young, S. (2000). A self-learning predictive model of articulator movements during speech production. *The Journal of the Acoustical Society of America*, *107*(3), 1659–1670.

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.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound, 14.

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* (Version 6.0.25). Retrieved March 21, 2021, from https://www.fon.hum.uva.nl/praat/

Bolker, B., Robinson, D., Menne, D., Gabry, J., Buerkner, P., Hua, C., Petry, W., Wiley, J., Kennedy, P., Eduard, S., Patil, I., Arel-Bundock, V., Denney, B., & Brunson, C. (2022, April 17). *Broom.mixed: Tidying Methods for Mixed Models* (Version 0.2.9.4). Retrieved July 6, 2022, from https://CRAN.R-project.org/package=broom.mixed

Brunelle, M. (2003). Tonal Coarticulation in Northern Vietnamese. *Proceedings of the 15th International Congress of Phonetic Sciences*, 2673–2676.

Cao, J. (1986). Putonghua qingsheng yinjie texing fenxi 普通话轻声音节特性分析(Acoustic features of neutral tone in Standard Mandarin). *Yingyong sheng xue*, (4), 1–6. Retrieved August 13, 2022, from https://www.cnki.com.cn/Article/CJFD1986-YYSN198604000.htm

Carignan, C., Hoole, P., Kunay, E., Pouplier, M., Joseph, A., Voit, D., Frahm, J., & Harrington, J. (2020). Analyzing speech in both time and space: Generalized additive mixed models can uncover systematic patterns of variation in vocal tract shape in real-time MRI. *Laboratory Phonology*, *11*(1).

Changsha Statistics Office. (2021). *Changsha statistical communiqué for seventh national population census*. Retrieved August 1, 2022, from http://www.yuelu.gov.cn/yl_xxgk/bmxxgkml/qtjj/tjxx/202106/t20210611_10004012.html

Chao, Y. R. (1930). A system of tone letters. *Le Maitre Phonetique*, *30*, 24–27.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press.

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.

Chen, M. Y. (1976). From middle Chinese to modern Peking. *Journal of Chinese Linguistics*, *4*(2/3), 113–277. http://www.jstor.org/stable/23752895

Chen, M. Y. (2000, August 3). *Tone Sandhi: Patterns across Chinese dialects* (1st ed.). Cambridge University Press.

Chen, S. H., & Wang, Y. R. (1990). Vector quantization of pitch information in Mandarin speech. *IEEE Transactions on Communications*, *38*(9), 1317–1320.

Chen, X. (2004). On the necessity and consistency principles of qingsheng words: Metric research on qingsheng words in modern Chinese dictionary 论轻声词界定的必要性、一致性原则. *Yuyan Wenzi Yingyong (Applied Linguistics)*, (1), 112–119. Retrieved June 2, 2022, from http://rdbk1.ynlib.cn:6251/Qw/Paper/263043

Chen, Y. (2010). Post-focus F0 compression—Now you see it, now you don't. *Journal of Phonetics*, *38*(4), 517–525.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, *36*(4), 724–746.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech – Evidence from F0 patterns of neutral tone in Standard Chinese. *Phonetica*, *63*(1), 47–75.

Cheshire, J. (2007). Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics*, *11*(2), 155–193.

Chien, Y.-F., Yan, H., & Sereno, J. A. (2021). Investigating the Lexical Representation of Mandarin Tone 3 Phonological Alternations. *Journal of Psycholinguistic Research*, *50*(4), 777–796.

Chuang, Y.-Y., Fon, J., Papakyritsis, I., & Baayen, H. (2021). Analyzing phonetic data with Generalized Additive Mixed Models. In *Manual of Clinical Phonetics*. Routledge.

Clements, G. N., & Ford, K. C. (1979). Kikuyu Tone Shift and Its Synchronic Consequences. *Linguistic Inquiry*, *10*(2), 179–210. http://www.jstor.org/stable/4178106

Coleman, J., Renwick, M. E. L., & Temple, R. A. M. (2016). Probabilistic underspecification in nasal place assimilation. *Phonology*, *33*(3), 425–458.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America*, *77*(6), 2142–2156.

Coretta, S., van Rij, J., & Wieling, M. (2022, April 18). *Tidymv: Tidy model visualisation for Generalised Additive Models* (Version 3.3.1). Retrieved July 6, 2022, from https://CRAN.R-project.org/package=tidymv

Davidson, L. (2021). Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian. *Journal of Phonetics*, *88*, 101075.

DeFrancis, J. (1986). *The Chinese language: Fact and fantasy*. University of Hawaii Press.

Deng, D. (2010). *Hanyu yunlvci yanjiu* 汉语韵律词研究*(Experimental study of Chinese prosodic word)*. Peking University Press.

Deterding, D. (2006). The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association*, *36*(2), 187–196.

DiCanio, C., Nam, H., Whalen, D. H., Timothy Bunnell, H., Amith, J. D., & García, R. C. (2013). Using automatic alignment to analyze endangered language data: Testing the viability of untrained alignment. *The Journal of the Acoustical Society of America*, *134*(3), 2235–2246.

Dreher, J. J., & Lee, P.-c. (1968). Instrumental Investigation of Single and Paired Mandarin Tonemes. *Monumenta Serica*, *27*, 343–373. Retrieved June 17, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk:2116/stable/40725872

Drubig, H. B. (2003). Toward a typology of focus and focus constructions. *Linguistics*, 1–50.

Duanmu, S. (1994). Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology*, *11*(1), 1–24. Retrieved June 6, 2022, from https://www.jstor.org/stable/4615448

Duanmu, S. (2000a). *The phonology of Standard Chinese*. Oxford University Press.

Duanmu, S. (2000b, January 1). Stress in Chinese. In *Chinese Phonology in Generative Grammar* (pp. 117–138). Brill.

Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). Oxford University Press. Retrieved June 6, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://ebookcentral.proquest.com/lib/oxford/detail.action?docID=415943

Duanmu, S. (2017). Syllable Structure (R. Sybesma, Ed.). *Encyclopedia of Chinese Language and Linguistics*, *4*, 230–236. Retrieved June 2, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk:3150/entries/encyclopedia-of-chinese-language-and-linguistics/syllable-structure-COM_00000404?s.num=0&s.f.s2_parent=s.f.book.encyclopedia-of-chinese-language-and-linguistics&s.q=Syllable+Structure

Eaton, John. W., Bateman, D., Hauberg, S., & Wehbring, R. (2020). *GNU Octave version 6.1.0 manual: A high-level interactive language for numerical computations*. https://www.gnu.org/software/octave/doc/v6.1.0/

Eckert, P. (2017, August 23). Age as a Sociolinguistic Variable. In F. Coulmas (Ed.), *The Handbook of Sociolinguistics* (pp. 151–167). Blackwell Publishing Ltd.

Entropic Research Laboratory. (2006). *Entropic Signal Processing System (ESPS)*. Retrieved June 7, 2022, from http://www.phon.ox.ac.uk/releases

Evans, J., Chu, M.-n., Aston, J. A., & Su, C.-y. (2010). Linguistic and human effects on F0 in a tonal dialect of Qiang. *Phonetica*, *67*(1-2), 82–99.

Firth, J. R., & Rogers, B. B. (1937). The structure of the Chinese monosyllable in a Hunanese dialect (Changsha). *Bulletin of the School of Oriental and African Studies*, *8*(4), 1055–1074.

Fishman, J. A. (1967). Bilingualism with and without diglossia; Diglossia with and without bilingualism. *Journal of Social Issues*, *23*(2), 29–38.

Fitzpatrick-Cole, J. (1996). Reduplication meets the phonological phrase in Bengali. *Linguistic review*, *13*(3-4), 305–356.

Flemming, E. (1995). *Auditory representations in phonology*. ProQuest Dissertations Publishing. Retrieved August 11, 2022, from https://search.proquest.com/docview/304169465?pq-origsite=primo

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, *18*(1), 7–44.

Flemming, E. (2007). The grammar of coarticulation., 23–26. Retrieved August 24, 2022, from http://web.mit.edu/flemming/www/paper/grammar-of-coarticulation.pdf

Frazier, M. (2013). The phonetics of Yucatec Maya and the typology of laryngeal complexity. *STUF - Language Typology and Universals*, *66*(1), 7–21.

Fu, L. (2010). *Fangyan yu Putonghua de jiechu yanjiu* 方言与普通话的接触研究*(A study of the contact between Mandarin and dialects)*. Soochow University. Retrieved April 11, 2023, from http://cnki.cgl.org.cn/kcms/detail/detail.aspx?DbCode=CDFD&filename=2010147048.nh

Fu, Q. J., & Zeng, F. G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing*, *5*(1), 45–57.

Fujisaki, H., Wang, C., Ohno, S., & Gu, W. (2005). Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model. *Speech Communication*, *47*(1-2), 59–70.

Gahl, S. (2008). Time and thyme are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, *84*(3), 474–496.

Gandour, J. (1977). On the Interaction between Tone and Vowel Length: Evidence from Thai Dialects. *Phonetica*, *34*(1), 54–65.

Goldsmith, J. A. (1976). *Autosegmental Phonology*. Massachusetts Institute of Technology.

Grabe, E., Kochanski, G., & Coleman, J. (2003). Quantitative modelling of intonational variation. *Proceedings of Speech Analysis and Recognition in Technology, Linguistics and Medicine*, 000–23.

Grabe, E., Kochanski, G., & Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech*, *50*(3), 281–310.

Graham, C., Buttery, P., & Nolan, F. (2016). Vowel characteristics in the assessment of L2 English pronunciation. *Proceedings of the Annual Conference of the International Speech Communication Association*, 1127–1131 Accepted: 2020-09-09T23:30:10Z.

Gubian, M., Torreira, F., & Boves, L. (2015). Using Functional Data Analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics*, *49*, 16–40.

Hadjipantelis, P. Z., Aston, J. A. D., & Evans, J. P. (2012). Characterizing fundamental frequency in Mandarin: A functional principal component approach utilizing mixed effect models. *The Journal of the Acoustical Society of America*, *131*(6), 4651–4664.

Halekoh, U., & Højsgaard, S. (2014). A Kenward-Roger approximation and parametric bootstrap methods for tests in linear mixed models: The R package pbkrtest. *Journal of Statistical Software*, *59*(1), 1–32.

Halekoh, U., & Højsgaard, S. (2021, March 9). *Pbkrtest: Parametric Bootstrap, Kenward-Roger and Satterthwaite Based Methods for Test in Mixed Models* (Version 0.5.1). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=pbkrtest

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *The Journal of the Acoustical Society of America*, *125*(1), 425–441.

Hartman, L. M. (1944). The Segmental Phonemes of the Peiping Dialect. *Language*, *20*(1), 28.

Haudricourt, A.-G. (1954). Origines asiatiques des langues malayo-polynésiennes. *Journal de la Société des Océanistes*, *10*(10), 180–183.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. London : University of Chicago Press.

Hayes, B., & Lahiri, A. (1991). Bengali intonational phonology, 50.

Hinskens, F. (1998). Dialect levelling: A two-dimensional process. *Folia Linguistica*, *32*(1-2).

Hirst, D., & Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, *15*, 75–85.

Ho, A. T. (1976). The Acoustic Variation of Mandarin Tones. *Phonetica*, *33*(5), 353–367.

Ho, D.-a. (2015). Chinese dialects. In W. S. Y. Wang & C. Sun (Eds.), *The Oxford Handbook of Chinese Linguistics*. Oxford University Press.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, *55*(1), 37–58.

House, A. S., & Fairbanks, G. (1953). The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels. *The Journal of the Acoustical Society of America*, *25*(1), 105–113.

HTK Team. (2006). *HTK Speech Recognition Toolkit* (Version 3.4). Retrieved July 6, 2022, from https://htk.eng.cam.ac.uk/

Hu, Y., & Xu, B. (Eds.). (1995). *Xiandai hanyu (modern Chinese)* (Chong ding ben, di 6 ban). Shang hai jiao yu chu ban she.

Huang, K. (2018). Phonological identity of the neutral-tone syllables in Taiwan Mandarin: An acoustic study. *Acta Linguistica Asiatica*, *8*(2), 9–50.

Huang, S., Liu, J., Wu, X., Wu, L., Yan, Y., & Qin, Z. (1998a). *1997 Mandarin Broadcast News Speech (HUB4-NE)*. Linguistic Data Consortium.

Huang, S., Liu, J., Wu, X., Wu, L., Yan, Y., & Qin, Z. (1998b). *1997 Mandarin Broadcast News Transcripts (HUB4-NE)*. Linguistic Data Consortium.

Humbert, J.-M. (1978, January 1). Consonant Types, Vowel Quality, and Tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 77–111). Academic Press.

Hunan Provincial Bureau of Statistics. (2017). *Changsha population summary and prospect* 长沙人口规模现状和发展趋势. Retrieved August 1, 2022, from http://tjj.hunan.gov.cn/tjfx/sxfx/zss/201706/t20170627_4327150.html

Hunan Provincial Bureau of Statistics. (2021). *Hunan statistical yearbook 2021*. China Statistics Press. Retrieved August 1, 2022, from http://www.enghunan.gov.cn/hneng/AboutHunan/HunanFacts/AdministrativeDivisions/201508/t20150804_1814341.html

Hyman, L. M. (1993). Register tones and tonal geometry. In H. van der Hulst & K. L. Snider (Eds.), *The phonology of tone: The representation of tonal register*. Mouton de Gruyter. Retrieved August 11, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://ebookcentral.proquest.com/lib/oxford/detail.action?docID=3049907

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, *23*(2), 225–257.

Hyman, L. M. (2007). Universals of tone rules: 30 years later. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes. Volume 1, Typological Studies in Word and Sentence Prosody.* (pp. 1–34). De Gruyter Mouton. Retrieved June 2, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://doi.org/10.1515/9783110207569

Hyman, L. M. (2018). Towards a typology of tone system changes. In H. Kubozono & M. Giriko (Eds.), *Tonal change and neutralization* (pp. 203–222). De Gruyter Mouton. Retrieved August 6, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://ebookcentral.proquest.com/lib/oxford/detail.action?docID=5158449

Institute of Linguistics CASS. (2016). *Xiandai hanyu cidian (Modern Chinese dictionary)* (7th ed.). The Commercial Press. Retrieved May 22, 2022, from https://www.cypressbooks.com/proddetail.php?prod=9787100124508

Institute of Linguistics CASS, 中国社会科学院民族学与人类学研究所, & 香港城市大学语言资讯科学研究中心. (2012). *Language Atlas of China* 中国语言地图集汉语方言卷 (2nd ed.).

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge University Press.

Iwata, R. (2018). Chinese tonal neutralization across dialects: From typological, geographical, and diachronic perspectives. In H. Kubozono & M. Giriko (Eds.), *Tonal change and neutralization* (pp. 156–199). De Gruyter Mouton. Retrieved August 14, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://ebookcentral.proquest.com/lib/oxford/detail.action?docID=5158449

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics*, *71*, 1–15.

Jing, S. (2002). *Xiandai hanyu qingsheng dongtai yanjiu* 现代汉语轻声动态究*(Dynamic study of neutral tone in modern Chinese)*. Minzu chubanshe.

Jing, S., & Niu, F. (2010). Changsha difang putonghua guhua yanjiu 长沙地方普通话固化研究:地方普通话固的个案调查(The fossilization of local Putonghua in Changsha: A sociolinguistic survey). *Yuyan wenzi yingyong* 语言文字应用, *4*, 41–49.

Johnson, K. (2014). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Proceedings of the first session of the 10th international symposium on spontaneous speech: Data and analysis* (pp. 29–54). The National International Institute for Japanese Language.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2000). Probabilistic Relations between Words: Evidence from Reduction in Lexical Production. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure.* (p. 26). John Benjamins.

Karlgren, B. (1922). The Reconstruction of Ancient Chinese. *T'oung Pao*, *21*(1), 1–42. Retrieved August 1, 2022, from https://www.jstor.org/stable/4526642

Kassambara, A. (2020, June 27). *Ggpubr: 'ggplot2' based publication ready plots* (Version 0.4.0). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=ggpubr

Kay, M., & Wiernik, B. M. (2022, February 27). *Ggdist: Visualizations of Distributions and Uncertainty* (Version 3.1.1). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=ggdist

Keating, P. (1990, November 30). The window model of coarticulation: Articulatory evidence. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology* (1st ed., pp. 451–470). Cambridge University Press.

Keating, P., & Cohn, A. C. (1988). Cross-language effects of vowels on consonant onsets. *The Journal of the Acoustical Society of America*, *84*(S1), S84–S84.

Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice, 5.

Keating, P., & Kuo, G. (2010). Comparison of speaking fundamental frequency in English and Mandarin. *UCLA Working Papers in Phonetics*, *108*, 164–187. Retrieved March 23, 2021, from https://escholarship.org/content/qt1gh6x943/qt1gh6x943.pdf?t=l921bp

Keating, P., & Shattuck-Hufnagel, S. (2002). A Prosodic View of Word Form Encoding for Speech Production. *UCLA working papers in phonetics*, 112–156.

Kerswill, P. (2020). Contact and new varieties. In *The Handbook of Language Contact* (pp. 241–259). John Wiley & Sons, Ltd.

Kerswill, P., Cheshire, J., Fox, S., & Torgersen, E. (2007). *Linguistic innovators: The English of adolescents in London* (ESRC End of Award Report RES-000-23-0680). Swindon: ESRC.

Kiparsky, P. (1982). Lexical Morphology and Phonology. In I. S. Yang (Ed.), *Linguistics in the morning calm*. Hanshin. Retrieved August 11, 2022, from https://web.stanford.edu/~kiparsky/Papers/Lexical%20Morphology%20and%20Phonology.pdf

Kirby, J., & Sonderegger, M. (2018). Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics*, *70*, 70–85.

Kirby, J. P., & Ladd, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. *The Journal of the Acoustical Society of America*, *140*(4), 2400–2411.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*, *118*(2), 1038–1054.

Kochanski, G., & Shih, C. (2000). Stem-ML: Language-independent prosody description. *ISCA Archive*, 4. http://www.xn--iscaspeech-vh801h.org/archive

Kochanski, G., & Shih, C. (2001). Automated modelling of Chinese intonation in continuous speech. *Proceedings of EUROSPEECH 2001*, 911–914.

Kochanski, G., Shih, C., & Jing, H. (2003). Hierarchical Structure and Word Strength Prediction of Mandarin Prosody. *International Journal of Speech Technology*, *6*(1), 33–43.

Koller, M. (2016). Robustlmm: An R package for robust estimation of Linear Mixed-Effects Models. *Journal of Statistical Software*, *75*(6).

Kong, H., & Wu, S. (2019). Frequency effect and neutralization of tones in Mandarin Chinese. *Journal of Quantitative Linguistics*, *26*(2), 95–115.

Kratochvíl, P. (1968). *The Chinese Language Today: Features of an Emerging Standard*. Hutchinson.

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, *142*(3), 1693–1706.

Kuang, J., & Liberman, M. (2015). Influence of spectral cues on the perception of pitch height. *Proceedings of ICPhS*, 0435.1–0435.5.

Kurpaska, M. (2010, January 18). *Chinese Language(s): A Look through the Prism of The Great Dictionary of Modern Chinese Dialects*. DE GRUYTER MOUTON.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13).

Labov, W. (1964). *The social stratification of English in New York city* (Doctoral dissertation). Columbia University. New York. Retrieved May 17, 2022, from https://www.proquest.com/docview/302149177/citation/DAC3C04D80D94FB9PQ/1

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, *1*(1), 97–120.

Labov, W. (1984). Field methods of the Project on Linguistic Change and Variation. In *Language in use: Readings in sociolinguistics*. Prentice-Hall.

Lahiri, A., & Plank, F. (2022). Phonological phrasing: Approaches to grouping at lower levels of the prosodic hierarchy. In *The Oxford History of Phonology* (pp. 134–162). University Press.

Lahiri, A., & Reetz, H. (2008, August 22). Underspecified recognition. In *Laboratory Phonology 7*. De Gruyter Mouton.

Language atlas of China: Parts I and II. (1987) OCLC: 19755701.

Laniran, Y. O., & Clements, G. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics*, *31*(2), 203–250.

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, *33*(1), 109–112. Retrieved June 23, 2022, from https://www.jstor.org/stable/44526911

Lee, W.-S., & Zee, E. (2008). Prosodic characteristics of the neutral tone in Beijing Mandarin 北京话轻声的韵律特征. *Journal of Chinese Linguistics*, *36*(1), 1–29. Retrieved June 2, 2022, from https://www.jstor.org/stable/23754104

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1).

Lenth, R. V. (2021, March 21). *Emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.5.5-1). Retrieved April 11, 2021, from https://CRAN.R-project.org/package=emmeans

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Li, Q., & Chen, Y. (2019). Prosodically conditioned neutral-tone realization in Tianjin Mandarin. *Journal of East Asian linguistics*, *28*(3), 211–242.

Li, R. (1998). *Chang sha fang yan ci dian. (Changsha dialect dictionary)*. (H. X. Bao, R. Shen, Y. Wu, & z. Cui, Eds.; 2nd ed.). Jiang su jiao yu chu ban she.

Li, W. (1981). Shi lun qingsheng he zhongyin (A preliminary discussion on stressless and stressed syllables) 试论轻声和重音. *Zhongguo Yuwen* 中国语文*(Studies of the Chinese Language)*, (1), 35–40.

Li, X. (2004). *Xinan guanhua yuyin yanjiu* 西南官话语音研究*Phonetic studies of Southwest Mandarin.* Shanghai shifan daxue. Retrieved August 13, 2022, from https://le.cnki.net/kmobile/Master/detail/ZKTM/2004061589.nh
Li, Y., Lee, T., & Qian, Y. (2004). Analysis and modeling of F0 contours for cantonese text-to-speech. *ACM Transactions on Asian Language Information Processing*, *3*(3), 169–180.
Li, Z. (2003). *The phonetics and phonology of tone mapping in a constraint-based approach* (Thesis). Massachusetts Institute of Technology. Retrieved June 2, 2022, from https://dspace.mit.edu/handle/1721.1/17651
Accepted: 2009-01-30T18:38:10Z
Liang, C. (2010). Qingsheng ci de jiazhi (The significance of neutral tone syllables) 论轻声词的价值. *Social Sciences in Guangxi* 广西社会科学, (7), 133–136.
Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language Sound Structure* (pp. 157–233). MIT Press.
Lieberman, P. (1967). *Intonation, perception, and language.* M.I.T. Press.
Lin, H. (2000, January 1). Stress and the Distribution of the Neutral Tone in Mandarin. In *Chinese Phonology in Generative Grammar* (pp. 139–161). Brill.
Lin, H. (2006). Mandarin Neutral Tone as a Phonologically Low Tone. *Journal of Chinese Language and Computing*, *16*(2), 121–134.
Lin, H. (2011). Changsha liandu biandiao zhi mi 沙之(Changsha tone sandhi). *Huayuwen jiaoxue yanjiu* 文教研究, *8*(2), 27–64.
Lin, M., & Yan, J. (1980). Beijinghua qingsheng de shengxue xingzhi (The acoustic features of the neutral tone in Beijing dialect) 北京话轻声的声学性质. *Fangyan* 方言, *3*, 166–178.
Lin, Y. A. (2010). The De-marked Modification Structure in Mandarin Chinese. *Proceedings of the 22nd North American Conference on Chinese Linguistics*, 17.
Lin, Y., & Li, L. (2017). Chinese words with a light tone in Modern Chinese Dictionary (6th edition)《现代汉语词典》（第6版）轻声词处理问题刍议. *Hebei shi fan da xue xue bao. Zhe xue she hui ke xue ban*, *40*(6), 101–106.
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). Kluwer Academic.
Liu, S., & Samuel, A. G. (2004). Perception of Mandarin Lexical Tones when F0 Information is Neutralized. *Language and Speech*, *47*(2), 109–138.
Ma, S., & Wang, Y. (2015). Xiandai hanyu putonghua qingsheng shuaitui xianxiang ji qi zhengyin wenti (The phenomenon of neutral tone decay and its correction in Modern Mandarin Chinese) 现代汉语普通话轻声衰颓现象及其正音问题. *Journal of Puyang Vocational and Technical College* 濮阳职业技术学院学报, *28*(6), 97–99.
Maidment, J., & Lecumberri, M. (1996). Pitch analysis methods for cross-speaker comparison. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, *4*, 2247–2249.
Matisoff, J. A. (1991). Sino-Tibetan Linguistics: Present State and Future Prospects. *Annual review of anthropology*, *20*(1), 469–504.
Menn, L., & Boyce, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, *25*(4), 341–383. Retrieved March 28, 2021, from https://ezproxy-prd.bodleian.ox.ac.uk:2082/docview/1299114858/fulltextPDF/26E0710FD1C24A3EPQ/1?accountid=13042

Milroy, L. (1980). *Language and social networks*. Basil Blackwell.
Nespor, M. (1986). *Prosodic phonology*. Foris. Retrieved July 30, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://ebookcentral.proquest.com/lib/oxford/detail.action?docID=3049945
Ni, J., Kawai, H., & Hirose, K. (2006). Constrained tone transformation technique for separation and combination of Mandarin tone and intonation. *The Journal of the Acoustical Society of America*, *119*(3), 1764–1782.
Nieminen, T. (2020). *Praat-textgrids: Manipulation of Praat TextGrids* (Version 1.3.1). Retrieved July 6, 2022, from http://github.com/Legisign/Praat-textgrids
Nolan, F. (2003). Intonational equivalence: An experimental evaluation of pitch scales. *Proceedings of the 15th International Congress of Phonetic Sciences*, 4.
Norman, J. (1979). Chronological strata in the Min dialects. *Fang yan*, (4), 268–274.
Norman, J. (1988). *Chinese*. Cambridge University Press.
Ohala, J. J. (1978, January 1). Production of Tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5–39). Academic Press.
Paul, W. (2015). *New perspectives on Chinese syntax*. De Gruyter Mouton.
Peng, J. (2006). *Xiang yu yin yun li shi ceng ci yan jiu. (Analysis of the historical strata of Xiang)* 湘语音韵历史层次研究. Shanghai Normal University. Retrieved August 24, 2021, from http://www.wanfangdata.com.cn/details/detail.do?_type=degree&id=Y932132
Peng, S., Chan, M. K. M., Tseng, C., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a pan-Mandarin system for prosodic transcription. In S. Jun (Ed.), *Prosodic Typology*. Oxford University Press.
Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *The Journal of the Acoustical Society of America*, *66*(2), 363–369.
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Massachusetts Institute of Technology.
Pierrehumbert, J. (1981). Synthesizing intonation. *The Journal of the Acoustical Society of America*, *70*(4), 985–995.
Pierrehumbert, J. (1988). *Japanese tone structure*. MIT Press.
Pittayaporn, P. (2018). Phonetic and systemic biases in tonal contour changes in Bangkok Thai. In M. Giriko & H. Kubozono (Eds.), *Tonal Change and Neutralization* (pp. 249–278). De Gruyter Mouton. Retrieved August 15, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://doi.org/10.1515/9783110567502
Politzer-Ahles, S., Schluter, K., Wu, K., & Almeida, D. (2016). Asymmetries in the perception of Mandarin tones: Evidence from mismatch negativity. *Journal of experimental psychology. Human perception and performance*, *42*(10), 1547–1570.
Poser, W. J. (1984). *The phonetics and phonology of tone and intonation in Japanese*. Massachusetts Institute of Technology.
Potisuk, S., Gandour, J., & Harper, M. P. (1997). Contextual variations in trisyllabic sequences of Thai tones. *Phonetica*, *54*(1), 22–42.
Prom-on, S., Xu, Y., & Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America*, *125*(1), 405–424.
Pulleyblank, D. G. (1986). *Tone in lexical phonology*. D. Reidel Pub. Co ; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers.

Pulleyblank, E. G. (1978). The Nature of the Middle Chinese tones and their development to early Mandarin 中古汉语声调的本质和到早期官话的演变. *Journal of Chinese Linguistics*, *6*(2), 173–203. Retrieved August 6, 2022, from https://ezproxy-prd.bodleian.ox.ac.uk:2116/stable/23752830

Pulleyblank, E. G. (1991). *Lexicon of Reconstructed Pronunciation in Early Middle Chinese, Late Middle Chinese, and Early Mandarin*. UBC Press. Retrieved August 4, 2022, from https://www.jstor.org/stable/495390?origin=crossref

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria. https://www.R-project.org/

Ramsay, J., & Silverman, B. (1997). *Functional data analysis*. Springer.

Ratliff, M. (2015). Tonoexodus, Tonogenesis, and Tone Change.

Reetz, H. (1996). *Pitch perception in speech: A time domain approach* (No. 26). Institute for Functional Research into Language and Language Use. Amsterdem.

Roettger, T. B. (2019). Researcher degrees of freedom in phonetic research. *Laboratory Phonology*, *10*(1).

Roettger, T. B., & Gordon, M. (2017). Methodological issues in the study of word stress correlates. *Linguistics vanguard : multimodal online journal*, *3*(1).

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication*, *6*(4), 343–352.

RStudio Team. (2022). *RStudio: Integrated development for R*. Boston, MA. Retrieved July 6, 2022, from https://www.rstudio.com/

Ryant, N., Yuan, J., & Liberman, M. (2014). Mandarin tone classification without pitch tracking. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 4868–4872.

Salmons, J. (1990). From Tone to Stress: Mechanisms and Motivations. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, *16*, 282–291.

Sangster, C. M. (2002). *Inter- and intra-speaker variation in Liverpool English: A sociophonetic study*. University of Oxford.

Sayahi, L. (2020). Contact, bilingualism, and diglossia. In *The Handbook of Language Contact* (pp. 51–66). John Wiley & Sons, Ltd.

Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., & Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, *11*(9), 1141–1152.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, *10*(3), 209–232.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology*, *3*, 371–405.

Selting, M. (2007). Lists as embedded structures and the prosody of list construction as an interactional resource. *Journal of Pragmatics*, *39*(3), 483–526.

Shen, J., & Souza, P. E. (2018). On Dynamic Pitch Benefit for Speech Recognition in Speech Masker. *Frontiers in Psychology*, *9*. Retrieved August 15, 2022, from https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01967

Shen, S. (2019). *An acoustical analysis of citation tones in Mandarin with Changsha dialect* 长沙"塑普"单字调分析. Zhongguo gaoxiao renwen shehui kexue xinxi wang. Retrieved July 21, 2022, from https://www.sinoss.net/c/2019-09-06/557055.shtml

Shen, X. S. (1989). *The prosody of Mandarin Chinese*. University of California Press.

Shen, X. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, *18*(2), 281–295.

Shen, X. S. (1991). Question intonation in natural speech: A study of Changsha Chinese. *Journal of the International Phonetic Association*, *21*(1), 19–28.

Shen, X. S. (1992). Mandarin neutral tone revisited. *Acta Linguistica Hafniensia*, *24*(1), 131–152.

Sherr-Ziarko, E. (2015). Word Frequency Effects on Homophonous Words in Mandarin Chinese. *Proceedings of ICPhS*, 5.

Shi, B., & Zhang, J. (1986). Vowel intrinsic pitch in Standard Chinese. *Lund Working Papers in Linguistics*, *29*. Retrieved March 25, 2021, from https://journals.lub.lu.se/LWPL/article/view/16948

Shih, C. (1987). *The phonetics of the Chinese tonal system.*

Shih, C. (1988). Tone and intonation in Mandarin. *Working papers of the Cornell Phonetics Laboratory*, *3*, 83–109.

Shih, C. (2000, November 30). A declination model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 243–268). Springer.

Shih, C., Kochanski, G., Fosler-Lussier, E., Chan, M., & Yuan, J.-H. (2001). Implications of prosody modeling for prosody recognition, 6.

Sievert, C., Parmer, C., Hocking, T., Chamberlain, S., Ram, K., Corvellec, M., Despouy, P., Brüggemann, S., & Inc, P. T. (2021, October 9). *Plotly: Create Interactive Web Graphics via 'plotly.js'* (Version 4.10.0). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=plotly

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366.

Simpson, A. (2001). Definiteness agreement and the Chinese DP. *Language and Linguistics*, *2*(1), 125–156. Retrieved July 16, 2022, from https://dornsife.usc.edu/assets/sites/1494/docs/definiteness_agreement.pdf

Slowikowski, K., Schep, A., Hughes, S., Dang, T. K., Lukauskas, S., Irisson, J.-O., Kamvar, Z. N., Ryan, T., Christophe, D., Hiroaki, Y., Gramme, P., Abdol, A. M., Barrett, M., Cannoodt, R., Krassowski, M., Chirico, M., & Aphalo, P. (2021, January 15). *Ggrepel: Automatically position non-overlapping text labels with 'ggplot2'* (Version 0.9.1). Retrieved July 6, 2022, from https://CRAN.R-project.org/package=ggrepel

Sóskuthy, M. (2017, March 15). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction* [Monograph]. Retrieved July 19, 2022, from https://eprints.whiterose.ac.uk/113858/

Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics*, *84*, 101017.

Sproat, R., & Shih, C. (1987). Prenominal Adjectival Ordering in English and Mandarin. *North East Linguistics Society*, *18*(3). https://scholarworks.umass.edu/nels/vol18/iss3/12

Stanford, J. N. (2008). A sociotonetic analysis of Sui dialect contact. *Language Variation and Change*, *20*(3), 409–450.

Sui, Y. (2016, November). The Interaction of Metrical Structure and Tone in Standard Chinese. In J. Heinz, R. Goedemans, & H. van der Hulst (Eds.), *Dimensions of Phonological Stress* (pp. 101–122). Cambridge University Press.

Sun, Y., & Shih, C. (2021). Boundary-conditioned anticipatory tonal coarticulation in Standard Mandarin. *Journal of Phonetics*, *84*, 101018.

Sundberg, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics*, *7*(2), 71–79.

Svantesson, J.-O. (1991). Hu: A language with unorthodox tonogenesis. In J. H. C. S. Davidson & H. L. Shorto (Eds.), *Austroasiatic languages: Essays in honour of H. L. Shorto* (pp. 67–79). School of Oriental and African Studies , Univ. of London.

Tagliamonte, S. A. (2006, May). *Analysing sociolinguistic variation*. Cambridge University Press.

Takahashi, Y. (2019). The phonological status of Low tones in Shanghai tone sandhi: Default tones or boundary tones? *Language and Linguistics*, *20*(1), 15–45.

Talkin, D. (1995). A robust algorithm for pitch tracking (RAPT). In W. B. Kleijn & K. K. Paliwal (Eds.), *Speech coding and synthesis* (pp. 495–518). Elsevier. Retrieved August 28, 2022, from https://www.ee.columbia.edu/~dpwe/papers/Talkin95-rapt.pdf

Tang, P., Xu Rattanasone, N., Yuen, I., & Demuth, K. (2017). Acoustic realization of Mandarin neutral tone and tone sandhi in infant-directed speech and Lombard speech. *The Journal of the Acoustical Society of America*, *142*(5), 2823–2835.

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (2019). Acquisition of weak syllables in tonal languages: Acoustic evidence from neutral tone in Mandarin Chinese. *Journal of Child Language*, *46*(1), 24–50.

Taylor, P. (2000). Analysis and synthesis of intonation using the Tilt model. *The Journal of the Acoustical Society of America*, *107*(3), 1697–1714.

Thurgood, G. (1999). *From ancient Cham to modern dialects: Two thousand years of language contact and change*. University of Hawai'i Press.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica*, *19*(2), 333–363.

Ting, P.-H. (1996). Tonal evolution and tonal reconstruction in Chinese. In C.-T. J. Huang & Y.-h. A. Li (Eds.), *New Horizons in Chinese Linguistics* (pp. 141–159). Springer Netherlands.

Titze, I. R. (1989). On the relation between subglottal pressure and fundamental frequency in phonation. *The Journal of the Acoustical Society of America*, *85*(2), 901–906.

Třísková, H. (2017). De-stress in Mandarin: Clitics, cliticoids, and phonetic chunks. In *Key Issues In Chinese as a Second Language Research*. Routledge.

Trudgill, P. (1986). *Dialects in contact*. Blackwell.

Trudgill, P. (2004). *New-dialect formation: The inevitability of colonial Englishes*. Edinburgh University Press. Retrieved April 13, 2023, from https://www.jstor.org/stable/10.3366/j.ctv2f4vkzd

Tsu-lin, M. (1970). Tones and prosody in Middle Chinese and the origin of the rising tone. *Harvard Journal of Asiatic Studies*, *30*, 86–110.

van Rij, J., Wieling, M., Baayen, R. H., & Rijn, H. van. (2022, June 17). *Itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* (Version 2.4.1). Retrieved July 6, 2022, from https://CRAN.R-project.org/package=itsadug

van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

van Santen, J., Shih, C., & Möbius, B. (1998). Intonation. In R. Sproat (Ed.), *Multilingual text-to-speech synthesis: The Bell Labs approach* (pp. 141–190). Kluwer Academic Publishers.

Wang, J. (1997). The representation of the neutral tone in Chinese Putonghua. In J. Wang & N. Smith (Eds.), *The representation of the neutral tone in Chinese Putonghua* (pp. 157–184). De Gruyter Mouton.

Wang, Y. (2018). *Changsha "suliao" putonghua yanjiu 长沙"塑料"普通话研究A study of Changsha Plastic Mandarin*. Shanghai Jiao Tong University. Shanghai. Retrieved August 31, 2022, from https://wap.cnki.net/touch/web/Dissertation/Article/10248-1019635302.nh.html

Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *The Journal of the Acoustical Society of America*, *106*(1), 458–468.

Wei, Y. (2011). Wulumuqi hanyu fangyan qingsheng de yuyin xingzhi ji youxuanlun fenxi (An acoustic and OT analysis of the neutral tone in Ürümqi Mandarin) 乌鲁木齐汉语方言轻声的语音性质及优选论分析. *Journal of Southwest Agricultural University (Social Science Edition) 西南农业大学学报(社会科学版)*, *9*(1), 115–118.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics*, *23*, 349–366.

Wickham, H., & RStudio. (2020, May 19). *Modelr: Modelling Functions that Work with the Pipe* (Version 0.1.8). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=modelr

Wickham, H., & RStudio. (2021, April 15). *Tidyverse: Easily Install and Load the 'Tidyverse'* (Version 1.3.1). Retrieved July 6, 2022, from https://CRAN.R-project.org/package=tidyverse

Wieling, M., Tomaschek, F., Arnold, D., Tiede, M., Bröker, F., Thiele, S., Wood, S. N., & Baayen, R. H. (2016). Investigating dialectal differences using articulography. *Journal of Phonetics*, *59*, 122–143.

Wilke, C. O. (2020, January 12). *Ggridges: Ridgeline Plots in 'ggplot2'* (Version 0.5.2). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=ggridges

Winter, B. (2013, August 26). *Linear models and linear mixed effects models in R with linguistic applications*. Retrieved June 15, 2022, from http://arxiv.org/abs/1308.5499

Winter, B. (2015). The other N: The role of repetitions and items in the design of phonetic experiments. *Proceedings of the 18th International Congress of Phonetic Sciences*, 5.

Winter, B., & Grawunder, S. (2012). The phonetic profile of Korean formal and informal speech registers. *Journal of Phonetics*, *40*(6), 808–815.

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, *1*(1), 7–18.

Wood, S. (2022, March 29). *Mgcv: Mixed GAM computation vehicle with automatic smoothness estimation* (Version 1.8-40). Retrieved July 6, 2022, from https://CRAN.R-project.org/package=mgcv

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95–114.

Wood, S. N. (2006, February 27). *Generalized Additive Models : An Introduction with R.* Chapman and Hall/CRC.

Wu, F., & Kenstowicz, M. (2015). Duration reflexes of syllable structure in Mandarin. *Lingua*, *164*, 87–99.

Wu, J. (2015). *Tonal Bilingualism: The case of two closely related Chinese dialects.* Universiteit Leiden. Utrecht.

Wu, Y., Adda-Decker, M., & Lamel, L. (2020). Mandarin lexical tones: A corpus-based study of word length, syllable position and prosodic position on duration. *Interspeech 2020*, 1908–1912.

Wu, Y. (2005, January 31). *A synchronic and diachronic study of the grammar of the Chinese Xiang dialects.* De Gruyter Mouton.

Xiao, N., & Li, M. (2018, May 14). *Ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for 'ggplot2'* (Version 2.9). Retrieved July 5, 2022, from https://CRAN.R-project.org/package=ggsci

Xiao, R. (2009). *A frequency dictionary of Mandarin Chinese: core vocabulary for learners.* Routledge. Retrieved May 19, 2022, from http://www.loc.gov/catdir/toc/ecip0823/2008030172.html

Xu, C. (2018). *Characterising Plastic Mandarin: A quantitative modelling approach.* University of Oxford. Oxford, UK.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association*, *33*(2), 165–181.

Xu, X., Chen, Z., & Peng, X. (2012). Dialect enregisterment during the popularization process of Putonghua — A sociolinguistic study on Changsha "plastic" Putonghua 普通话普及过程中新方言的"登(enregisterment)"-长沙塑料普通话的社会语言学调查研究. In *Beijingshi yuyan wenzi gongzuo yanjiu luncong* 北京市语言文字工作研究论丛（*Collection of linguistic studies in BeiJing).* Shoudu shifan daxue chubanshe. Retrieved August 31, 2022, from https://read01.com/zQRjEP.html

Xu, Y. (1997). Contextual Tonal Variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*, 51–105.

Xu, Y. (2009). Timing and coordination in tone and intonation—An articulatory-functional perspective. *Lingua*, *119*(6), 906–927.

Xu, Y., & Prom-on, S. (2014). Toward invariant functional representations of variable surface fundamental frequency contours: Synthesizing speech melody via model-based stochastic learning. *Speech Communication*, *57*, 181–208.

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *J. Acoust. Soc. Am.*, *111*(3), 1399–1413.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*(4), 319–337.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197.

Yang, C., & Xu, Y. (2019). Cross-linguistic trends in tone change: A review of tone change studies in East and Southeast Asia. *Diachronica*, *36*, 417–459. Retrieved October 3, 2020, from http://www.homepages.ucl.ac.uk/~uclyyix/yispapers/Yang_Xu_Diachronica_accepted.pdf

Yip, M. (1980). *The tonal phonology of Chinese* (Thesis). Massachusetts Institute of Technology. Retrieved June 2, 2022, from

https://dspace.mit.edu/handle/1721.1/15971
Accepted: 2009-01-23T14:35:48Z

Yip, M. (1989). Contour tones. *Phonology*, *6*(1), 149–174.

Yip, M. (2001). Tonal features, tonal inventories and phonetic targets. *UCL Working Papers in Linguistics*, (13), 161–188.

Yip, M. (2002). *Tone*. Cambridge University Press.

Yip, P. C. (2000). *The Chinese lexicon: A comprehensive survey*. Routledge.

Yiu, C. Y.-m. (2013, December 13). *The typology of motion events: An empirical study of Chinese dialects*. De Gruyter Mouton.

Yuan, J. (2012). The effects of speaking rate and intonation on the duration of tones in Mandarin Chinese. *Proceedings of Speech Prosody*, 3.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *The Journal of the Acoustical Society of America*, *123*(5), 3878–3878.

Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin Broadcast News Speech. *Speech Communication*, *65*, 67–74.

Yuan, J., & Liberman, M. (2015). Investigating consonant reduction in Mandarin Chinese with improved forced alignment. *ISCA Archive*, 2675–2678. http://www.isca-speech.org/archive/interspeech_2015

Yuan, J. (1960). *Hanyu fangyan gaiyao (A summary of Chinese dialects)*. Wenzi gaige chubanshe
OCLC: 1262159540.

Zhang, J. (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically Based Phonology* (pp. 157–190). Cambridge University Press. Retrieved June 18, 2022, from
http://ebookcentral.proquest.com/lib/oxford/detail.action?docID=266559

Zhang, J. (2014). *A sociophonetic study on tonal variation of the Wúxi and Shànghai dialects*. LOT, Netherlands Graduate School. Utrecht
OCLC: 903803909.

Zhang, J. (2018). A comparison of tone normalization methods for language variation research. *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*. Retrieved February 17, 2021, from https://www.aclweb.org/anthology/Y18-1095

Zhang, K. (1998). Lüe tan xiandai hanyu qingsheng ci de guifan (Brief discussion on the standard of neutral tone syllables in Modern Chinese) 略谈现代汉语轻声词的规范. *Journal of South China Normal University (Social Science Edition)*, (2), 128–130. Retrieved June 3, 2022, from
https://ezproxy-prd.bodleian.ox.ac.uk:3265/kcms/detail/detail.aspx?dbcode=CJFD&filename=HNSB199802024&dbname=CJFD9899&uid=WEEvREcwSlJHSldSdmVpbisvQWp4LzlxRTNpb1A1OXB1QXlpOXFwN3NQbz0%3D%249A4hF_YAuvQ5obgVAqNKPCYcEjKensW4IQMovwHtwkF4VYPoHbKxJw!!

Zhang, M. (2023, March 3). *You Can't Stress This Enough: The Interaction of Tone and Stress in Changsha Xiang*.

Zhang, Y. (2022, January 7). *Neutral Tone in Mandarin: Representation and Interaction with Utterance-level Prosody* (Thesis). University of Cambridge
Accepted: 2022-01-07T02:29:39Z.

Zhao, E., & Wu, Y. (2020, October 16). *Over 80 percent of Chinese population speak Mandarin*. People's Daily Online. Retrieved April 26, 2023, from http://en.people.cn/n3/2020/1016/c90000-9769716.html

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, *37*(2), 231–247.

Zhong, Q. (2003). Changsha hua de qingsheng 长沙话的轻声(Light tone in Changsha dialect). *Fangyan* 方言, (3), 255–264. Retrieved May 4, 2023, from http://www.cqvip.com/qk/81953x/200303/9351513.html

Zhou, C. (2018, September 1). *Beijing hua yuyin yanbian yanjiu* 北京话语音演变研究*(Phonetic changes in Beijing Mandarin)*. Peking University Press.

Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: Timit and beyond. *Speech Communication*, *9*(4), 351–356.