Journal of Phonetics 106 (2024) 101341

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Cross-dialectal perspectives on Mandarin neutral tone

Chenzi Xu [a,b]

[a] University of Oxford, 41 Wellington Square, Oxford OX1 2JF, United Kingdom
[b] University of York, Heslington, York YO10 5DD, United Kingdom

ARTICLE INFO

Article history:
Received 13 September 2023
Received in revised form 7 June 2024
Accepted 9 June 2024
Available online 22 July 2024

Keywords:
Neutral tone
Standard Mandarin
Plastic Mandarin
Generalised additive modelling

ABSTRACT

With an aim to investigate the nature of Mandarin neutral tone through the lens of language variation and change, this study examines the pitch patterns of speech sequences containing neutral tone syllables, i.e. those that do not have any of the four canonical lexical tones and are often overlooked in prior studies of tones, in two Mandarin varieties: Standard Mandarin and Plastic Mandarin spoken in Changsha, China. Using Generalised Additive Mixed Models, the study shows (a) that f0 contours of a sequence of neutral tone syllables following various lexical tones converge in the end at a low pitch in both Mandarin varieties, and (b) that the low pitch target of neutral tone syllables tends to be the same across the two Mandarin varieties. The cross-dialectal comparison favours the phonological account that neutral tone is underlyingly underspecified and attracts the boundary tone. It suggests that the constant pitch target across two Mandarin varieties with distinct lexical tone contours may be attributed to the stable transfer of prosodic structure in the Standard-Plastic variation.

1. Introduction

Prominence in speech is often accompanied by greater articulatory care, whereas weak elements tend to exhibit a broader spectrum of variability. It is frequently observed that an English reduced vowel, schwa /ə/ for example, shows a continuum of vowel qualities that are contextually conditioned (Flemming & Johnson, 2007). Similarly in the suprasegmental domain, it is well established that the acoustic properties of the Standard Mandarin neutral tone, which never occurs in stressed positions, display considerable variability and are largely influenced by its neighbouring contexts (e.g. Chao, 1968; Shen, 1992; Yip, 1980). The high susceptibility of unstressed vowels and neutral tones to coarticulation raises the parallel debate on whether they are phonologically underspecified or targetless (e.g. see Browman & Goldstein, 1992, Flemming, 2009 for views on schwa; Shen, 1992 on neutral tone).

The theoretical discussion on the presence of a target, in addition to explaining the variable surface realisation, is pertinent to understanding how a tone or a vowel may undergo change or remain unchanged. While historical research on tones has predominantly concentrated on the changes of lexical tones, a synchronic comparative analysis of neutral tones offers insights into their potential mechanisms of evolution. Neutral tone is not confined to Standard Mandarin. A few studies have reported on neutral tones in other Mandarin varieties such as Ürümqi Mandarin (Wei, 2011), Taiwan Mandarin (Huang, 2018), and Tianjin Mandarin (Li & Chen, 2019). Neutral tone in other Mandarin varieties can potentially be a test ground of theoretical proposals built upon observations of neutral tones in Standard Mandarin (refer to section 1.4).

This study aims to shed light on the nature of Mandarin neutral tone through the lens of language variation and change, by eliciting and examining comparable neutral tone data from Standard Mandarin and Plastic Mandarin, a newly emerged urban Mandarin variety in Changsha. Such cross-dialectal comparison between two closely related Mandarin varieties under the context of new dialect formation highlights the patterns of continuity and variation in tones, which in turn contributes to the discussion on the representation of neutral tone.

1.1. Tone change in Mandarin variation

Since at least the second millennium BC, the language ancestral to contemporary Chinese varieties on the North China Plain along the banks of the Yellow River has been spreading and evolving. The history of Chinese languages can be characterised by a constant interplay between the

E-mail address: chenzi.xu@ling-phil.ox.ac.uk

https://doi.org/10.1016/j.wocn.2024.101341
0095-4470/

natural diversification accompanying the spread and the institutional drive for uniformity. Largely owing to the expansive plain territory in the north as well as the historical unity and influence, the northern varieties in China are overall quite homogenous, compared to the remarkable diversity among southern varieties. The northern group, that is, the Mandarin group, has been the language of the capital for most of the last millennium (Norman, 1988), and is the largest group in terms of both population and geographical span, exerting a strong influence on other Chinese varieties.

The homogeneity of Mandarin varieties lies in their segmental phonology, along with grammar and lexicon. A defining characteristic of most Mandarin varieties compared to other Chinese groups, for example, is the absence of voiced obstruents in their phoneme inventories. Most regional varieties such as Tianjin, Xian, Jinan, Chengdu, and Nanning Mandarin also have the same tone categories as Beijing, or Standard Mandarin. The phonetic realisation of the four lexical tones, especially their most salient pitch patterns, however, is highly variable among these varieties (e.g. Hou et al., 2023; Li et al., 2017; Wu, 2015). In other words, the tone patterns are not stable and can be a significant cue of regional Mandarin variation.

The role neutral tone plays in Mandarin variation and change is rather obscure given that it is understudied both synchronically and diachronically. The presence of unstressed syllables tends to be universal in Mandarin varieties (Norman, 1988), though differences across dialects are observed including the lower frequency of their use in Taiwan Mandarin (Huang, 2018). Neutral tone is particularly prominent in Beijing Mandarin (Qian, 1985), and the degree and frequency of neutral tone realisation in Beijing can be a linguistic resource in indexing Beijing localness or cosmopolitan personae (Q. Zhang, 2005). While a synchronic comparison of neutral tones among long-existing Mandarin varieties reveals characteristics of neutral tones, the analysis of tonal specifications in neutral tone syllables may be complicated by varied strata of sound change over long stretches of time. The investigation of neutral tone in the process of new Mandarin dialect formation offers a novel and relatively straightforward perspective on neutral tone in Mandarin change over the past two decades. This study probes into a case of a newly emerged Mandarin variety that features distinctive tonal patterns, known as Plastic Mandarin, in southern China where a non-Mandarin local variety predominates.

1.2. Plastic Mandarin

In the city of Changsha, China, Standard Mandarin has gained momentum in the repertoire of a large part of the population which hitherto had spoken the regional vernacular, Changsha, one of the Xiang varieties largely mutually unintelligible with Standard Mandarin (see N. Wu, 2023 for Changsha phoneme inventory and tones). The addition of the standard variety, both the written and spoken form, in Changsha has been accelerated by education and mass media, as well as greater human mobility. The increasingly intensified contact between Standard Mandarin and Changsha was accompanied by the emergence of Plastic Mandarin. Plastic Mandarin has been quickly spreading among younger generations and rapidly reshaping local speech patterns.

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. This term has multiple connotations: first, the pronunciation of *sù* in Changsha, [sʷo], is a homophone of another morpheme meaning 'bad', thus *sù pǔ* indicates 'bad Mandarin'; second, *sùliào* 'plastic' 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.

Although it still remains unknown when this term first appeared in Changsha and who coined it, the earliest mention of this term in academic work to my knowledge occurs in Wu (2005), p. 22, referring to "a variety that combines the features of the local Changsha dialect with features of Standard Mandarin". It 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, but 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 mostly spoken by young people. Despite the absence of authoritative definition or standard, most respondents in Xu et al. (2012)'s study could identify this particular Plastic Mandarin and differentiate it from the general Changsha-accented L2 Mandarin. Furthermore, Jing and Niu (2010) observe that 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 article, Plastic Mandarin refers to the emerged and crystallised Mandarin of the young generations in Changsha.

Plastic Mandarin as a new Mandarin dialect features a core of distinctive pitch patterns. The phoneme inventory of Plastic Mandarin is hardly different from that of Standard Mandarin (Jing and Niu, 2010), though there may be individual variations, especially in the nasal codas. Plastic Mandarin has the same four citation tone categories as Standard Mandarin, but the shapes of tone contours in Plastic Mandarin exhibit systematic variations when compared to their counterparts in Standard Mandarin, as illustrated in Fig. 1, where prototypical time-normalised f0 contours in semitones referenced on speaker mean f0 are shown (Xu, 2022). Plastic Mandarin may have arisen when Mandarin tones adapted the pitch pattern of some corresponding Changsha tones (see Xu, 2022 for Plastic Mandarin citation tones). While Plastic Mandarin manifested an incursion of northern elements, the Mandarin phonemes, into the area, it concurrently signalled an infiltration of southern features, some Changsha prosodic patterns, into Mandarin.

1.3. Standard Mandarin neutral tone

Standard Mandarin is best known as a lexical tone language, although about one-third of all syllables in natural speech (Duanmu, 2000) may in fact not carry any of the four

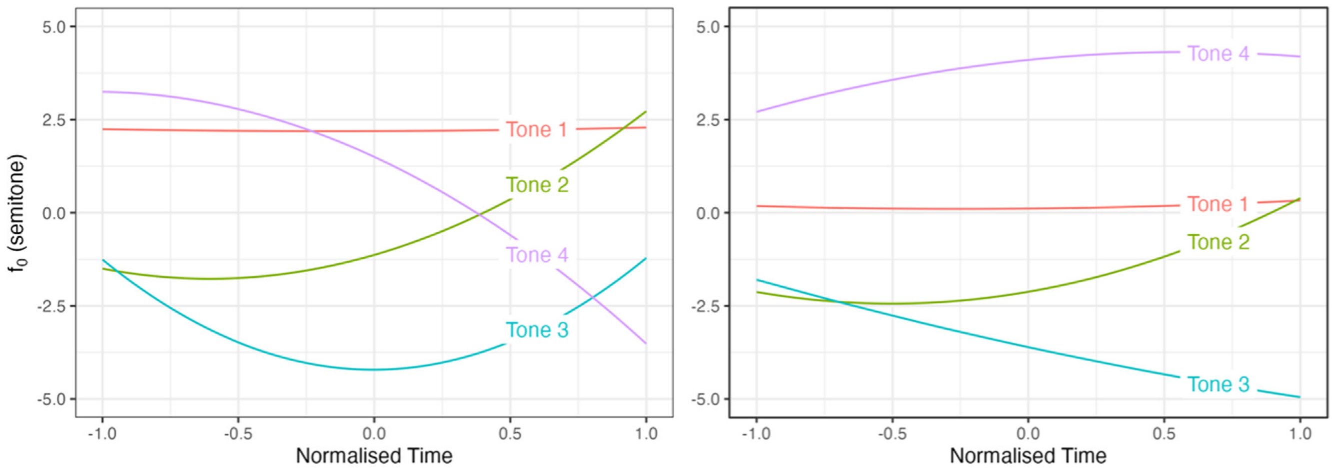


Fig. 1. Prototypical citation tone contours of Standard Mandarin (left) and Plastic Mandarin (right).

canonical lexical tones (Lee & Zee, 2003). 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.

Acoustically neutral tone syllables are often manifested in reduced duration, loudness, and vowel contrasts (Duanmu, 2000). 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 (1a), the back mid unrounded /ɤ/ is realised as a mid-central vowel [ə]; in (1b), an example from Duanmu (2000), the velar nasal is dropped, nasality is coarticulated with the vowel, and a back low unrounded /a/ is realised as a mid-central vowel [ə]. Moreover, the vowel can even be deleted, as in (1c). Sometimes voiceless onsets, especially voiceless unaspirated affricates, in neutral tone syllables may be voiced (Cao, 1986). The four lexical tones in Standard Mandarin are, therefore, sometimes regarded as full tones (e.g. in Chen & Xu, 2006; Tang et al., 2017), with neutral tone being the reduced tone. The degree of reduction tends to depend on speaking rate, formality, focus, emphasis, and so on. In line with the *Pinyin* transcription of Standard Mandarin in many Chinese dictionaries (e.g. Institute of Linguistics CASS, 2016), diacritics denoting tones are absent for neutral tone syllables (in bold) here.

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

Neutral tone syllables mostly occur enclitically whereby their pitch patterns tend to depend heavily on the preceding lexical tone (Chao, 1968; Shih, 1987; Yip, 2002). The phonetic description of neutral tone in Standard Mandarin, however, is not consistent across different studies. Table 1 presents the findings of four influential studies amongst many others (e.g. Dreher & Lee, 1968; Lin & 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 & Zee (2008). While the impressionistic studies only report one pitch value of neutral tone in different contexts, the acoustic studies include the pitch movement details. Although their conclusions disagree over the phonetic details, one common characteristic is that neutral tone in Standard Mandarin is falling or overall lower in pitch than the offset of the preceding tone when preceded by H, LH, HL (T1, T2, T4) tones, but rising or higher in pitch than the offset of the preceding tone when preceded by the L (T3) tone.

A small number of grammatical morphemes including suffixes and particles are the most common examples of syllables bearing neutral tone. These include the nominaliser or possessive particle *de* /tə/, the perfective particle *le* /lə/, the classifier particle *ge* /kə/, the diminutive suffix *zi* /t͡sz̩/, and so on. They predominantly fall on the light end of the light-heavy continuum between neutral tone and lexical tone syllables, considering their acoustic parameters including pitch, duration, and intensity (Huang & Li, 2023). Some of them are obligatory instances of neutral tone, such as *de*, and they are not affiliated with any canonical lexical tone category even when uttered in isolation, though their affiliation may be traced back diachronically (Lin,

Table 1
Phonetic descriptions of neutral tone in previous studies.

Preceding tone		Neutral tone			
		Chao, 1968	Shih, 1987	Chen, 2000	Lee & Zee, 2008
T1 H	˥	half-low	starts high, then falls	mid	mid falling
T2 LH	˧˥	mid	starts high, then falls, but not as low as after H	mid	high falling
T3 L	˨˩˦	half-high	starts fairly low, then rises	half-high	mid-level
T4 HL	˥˩	low	starts fairly low, and falls even lower	low	low falling

2006). Only a limited number of such morphemes exist – Shen (1992) referenced 47 grammatical morphemes. Other neutral tone syllables, in contrast, have a canonical lexical tone when spoken in isolation. While a neutral tone syllable can serve as an allophonic variant of its corresponding full tone syllable in some words, it leads to phonemic distinctions in others as exemplified in minimal pairs (2a) and (2b).

(2)	a.	*dōng xī*	/tʊŋ˥ ɕi˥/	东西	'east and west'
		dōng xi	/tʊŋ˥ ɕi/	东西	'things'
	b.	*dà yì*	/ta˥˩ i˥˩/	大意	'general idea'
		dà yi	/ta˥˩ i/	大意	'careless'

Not all neutral tones are necessarily uniform underlyingly. Shen (1992) proposed a three-way typology (toneless, detonic, and atonic) for neutral tones in Standard Mandarin and Zhang (2022) similarly suggested the categories of intrinsic and derived neutral tones, in which the latter are underlyingly the same as their lexical tone counterparts. The present study focuses on the most frequently occurring and representative neutral tone syllables, such as the possessive marker *de* and the plural marker *men*, thereby providing a baseline for further neutral tone research in Plastic Mandarin, considering the scarcity of studies in this area.

1.4. Approaches to Mandarin neutral tone

Various accounts have attempted to explain the observed variable pitch realisation of neutral tone in Standard Mandarin. Neutral tone has been treated as the fifth lexical tone, a tone sandhi phenomenon, or a product of tonal neutralisation resulting from the interaction of stress (Liu, 2002). The contextual dependency of neutral tone revealed by previous impressionistic observations and earlier acoustic studies has led to only a few attempts on the phonological specification of neutral tone as an independent tone category. Within the framework of autosegmental phonology, Yip (1980) proposes that a neutral tone is specified with a register feature [−upper], and the surface values of a neutral tone are realised through left-to-right feature spreading from the preceding tone. Shen (1992) adopts a similar feature spreading view as Yip (1980), linking a neutral tone to the end component of its preceding tone (i.e. [H] or [L]), but her model differs in that toneless neutral tone syllables such as *de* are not specified with any features.

Besides acquiring a tonal feature from the preceding tone through progressive assimilation, Lin (2006) posits an underlying L tone for neutral tone, given that neutral tones after Tones 1, 2, and 4 are typically falling. The rising neutral tone after Tone 3, an L tone, is accounted for by dissimilation appealing to the Obligatory Contour Principle. Li (2003) also assigns 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, consisting of a neutral tone syllable or a sequence of neutral tone syllables and the preceding full tone syllable regardless of their morphosyntactic structure. In Li (2003)'s account, the pitch target of neutral tone syllables is derived by inserting a post-lexical prosodic word boundary L tone at the right edge of the last neutral tone.

Apart from phonological analyses, some experimental studies have also proposed tonal targets for neutral tone. In van Santen et al.'s (1998) tone sequence approach to speech synthesis, neutral tone has a single mid-level target placed slightly before the end of a syllable and this target shifts from mid to low when following a falling tone. Chen and Xu (2006) also identify a static and mid target for neutral tones and explain that the greater surface variability of a neutral tone is due to the much slower or less robust implementation of the target and a heavy carryover influence of the preceding syllable.

The seminal paper by Chen and Xu (2006) introduced a classic experimental design paradigm by varying the number of consecutive neutral tone syllables in a prosodic word so as to reveal potential articulatory targets and the strength of coarticulatory effects. A few more recent studies such as Li and Chen (2019), Sun and Shih (2021), and Li & Li (2022) adopted similar designs. In addition to the influence of the preceding lexical tone, some anticipatory effects from the following tone have also been reported, such as the assimilatory effect when the following tone is the high falling Tone 4 and the dissimilatory effect when the following tone is the low Tone 3 in Standard Mandarin. Sun and Shih (2021) also show that such anticipatory tonal coarticulation is conditioned on the strength of prosodic boundary. The present study extends the scope of such an f0 analysis by investigating neutral tones in Plastic Mandarin. It also manipulates the number of consecutive neutral tones as well as their adjacent tones, with comparable speech data in Standard Mandarin, allowing for a direct comparison.

1.5. Research questions and hypotheses

From the perspective of language variation, the emergence of Plastic Mandarin may have involved systematically varying the pitch targets of the lexical tones in Standard Mandarin, as demonstrated in Fig. 1: Tone 1 was lowered from a high level into a mid level tone, Tone 2 changed from a mid rising to a low rising tone, in Tone 3 the dipping contour was truncated, flattened, and became a low falling tone, and Tone 4 was turned into a high rising tone from a high falling tone. While the tone change patterns in Tones 1, 2, and 3 are not uncommon in East and Southeast Asia according to the crosslinguistic survey in Yang & Xu (2019), the great resemblance between Tones 1, 2, and 4, especially Tone 4, in Plastic Mandarin and their counterparts in Changsha (Xu, 2022) indicates a significant influence from language contact. How is neutral tone, known for its heterogeneous tone patterns in Standard Mandarin, manifested in Plastic Mandarin? Assuming that Mandarin neutral tone has an underlying tone target, does it conform to the tone variations that reshape a tone system, as observed in all other lexical tones (Tones 1 to 4)? If not, what prevents it from patterning with the other tones? In an environment where Changsha Xiang has been dominating, how might the prosodic systems of the two potential source languages, Standard Mandarin and Changsha, interact in the formation of Plastic Mandarin? The current study offers some unique perspectives into these questions.

Considering the scarcity of studies in Plastic Mandarin, one main goal of this study is to explore its neutral tone patterns. Another goal is to clarify the nature of Mandarin neutral tone, primarily through an acoustic–phonetic analysis of neutral tones in the two Mandarin varieties that potentially served as the input and output in the formation of a new dialect. The specific research questions are:

1) How is a neutral tone or neutral tone sequence realised in various tonal contexts in Plastic Mandarin?
2) What are the similarities and differences in neutral tone between Plastic Mandarin and Standard Mandarin?
3) Is there an underlying pitch target for neutral tone in Standard and Plastic Mandarin?

With the first two research questions being exploratory in nature, several hypotheses can be proposed with regard to the neutral tone patterns in Plastic Mandarin.

First, neutral tones in Plastic Mandarin are similar to their counterparts in Standard Mandarin, with various f0 patterns but approaching a mid-low or low target, as manifested also in other Mandarin varieties such as Tianjin Mandarin and Taiwan Mandarin.

Second, the pitch patterns of neutral tones in Plastic Mandarin resemble those in Changsha, mirroring the potential influence of Changsha on the lexical tones in Plastic Mandarin. Unlike the neutral tones in Standard Mandarin whose pitch patterns were largely predicted by the preceding tone, the pitch pattern of a neutral tone in Changsha is determined by its corresponding full tone regardless of the neighbouring tones: it has a high pitch contour if its full tone counterpart is rising/high level but a low pitch contour if its full tone counterpart is a falling tone (Zhong, 2003).

Third, neutral tones in Plastic Mandarin have a consistent and distinct pitch pattern, acting as the fifth lexical tone pattern in the tone space of Plastic Mandarin.

In light of the underlying tone target of a neutral tone, the hypotheses diverge into two tendencies: if there is an underlying target for neutral tones akin to other lexical tones in Standard Mandarin, neutral tones are highly likely to have undergone a tone change towards a distinct tonal target, patterned with all other tones in this Standard-Plastic variation. On the contrary, if there is no underlying target, the tone change mechanism manifested in other lexical tones may not apply to neutral tones.

2. Methods

This study utilises comparable semi-spontaneous speech recordings in two Mandarin varieties. The current section provides a thorough description of data collection, acoustic processing, and statistical techniques. It also highlights the importance of fieldwork methods in eliciting data from a non-standard variety.

2.1. Experiment design

2.1.1. Eliciting a non-standard accent: The disguised word-guessing game

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 to formal interviews. The more formal the interviews, 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 that illustrated in 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 numerous factors. Hence, a verbal game embedded with carefully designed speech stimuli was devised for the purpose of acquiring controlled yet spontaneous speech.

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. A pair of subjects was presented with a scenario prompt through an incomplete stimulus sentence in written Chinese on the laptop screen. As illustrated in Table 2, the blank representing the missing keyword serves as an implicit in situ WH-question. One participant, the Clue-giver, was given three correct keywords in each scenario and tried to have their partner guess all 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, having received clues from their partner. A transcript excerpt of a conversation between two participants is displayed in Table 3. 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

Table 2
An example scenario prompt in the word guessing game. The three keywords are only available to the Clue-giver.

Scenario Prompt	他们堆的_____很稳当。 tāmen duī-de_____hěn wěndang 3PL built_____very stable
Keywords	雪人,城堡,台阶 xuěrén, chéngbǎo, táijiē **snowman, castle, stairs**

Table 3
An excerpt of the transcript of the conversation between speakers 213 and 214 during the word-guessing game. The Clue-receiver had three attempts, highlighted in light grey, with proposed keyword noun phrase underlined. The actual targeted phrases in this study are in bold.

Speaker ID	Role	Transcript
214	Clue-giver	法国的国王都会住在一个地方。 'One place the kings of France lived in.'
213	Clue-receiver	他们堆的宫殿很稳当。(2a) '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。(2b) 'The castle they **built** is very stable.'
214	Clue-giver	对。 'Yes.'

would confirm the correct answers by verbally filling in the correct keywords in the sentence before moving on to the next scenario. The pair spoke in an uninhibited fashion during the game. There are 16 scenario prompts and 16 × 3 = 48 keywords in total.

One key advantage of this game is that the participants generally find it entertaining and the relaxed and friendly atmosphere sparks naturalistic conversation. Such a technique allows for control over the tokens of requisite phonetic features that yield meaningful comparison on the one hand, and on the other hand, it directs participants' attention away from their own speech and keeps them maximally engaged.

2.1.2. Manipulating the linguistic contexts

This study examined the f0 contours of three sets of phrases embedded in semi-spontaneous natural speech with a varying number of consecutive neutral tone syllables:

(a) the disyllabic [X-***de***] set,
(b) the trisyllabic [X_1-$\mathbf{X_2}$-***men*** (*zhèng*)] set,
(c) the quadrisyllabic [X_1-$\mathbf{X_2}$-***men-de***] set,

where X_2, *men*, *de* are neutral tone syllables (marked in bold). The preceding tone of the neutral tones, X or X_1, was manipulated to be one of the four lexical tones in each Mandarin variety, and X_2 represents the reduplicated syllable of X_1 but in neutral tone. The inclusion of a neutral tone sequence is conducive to assessing the magnitude of the tonal coarticulatory effects. In set (b), the Tone 4 syllable *zhèng* with a high starting pitch in both Mandarin varieties follows the neutral tone syllable *men*, whereas the following tone after *de* in sets (a) and (c) varies randomly (see example keywords with different tones following *duī-de* in Table 2). Combining sets (b) and (c) enables an analysis of how the distinct fourth syllables, *zhèng* and *de*, affect the f0 contours of the [X_1-X_2-*men*] sequences. In set (a), the voicing part of the X syllable was also controlled to explore its potential influence on the f0 contours of *de*.

In addition, focus may affect f0 contours such as the post-focus compression effect in Mandarin (Xu, 1999). In the present study, most targeted phrases were positioned before the expected narrow focus, since the pre-focus f0 is known to be similar to the focus-neutral f0, deprived of compression effect (e.g. Chen & Gussenhoven, 2008). This is demonstrated in examples of targeted sentences (3), taken from a conversation excerpt displayed in Table 3, where the constituents of the proposed answers (underlined) are the new information and naturally in completive focus position (Drubig, 2003). If a wrong word was introduced and rejected, the updated word might also be in contrastive focus, as in (3b).

(3) a. tāmen duī *de* <u>gōngdiàn</u> hěn wěndang
3PL built DE palace very stable
'The <u>palace</u> they built is very stable.' *'<u>gōngdiàn</u>' in focus*

b. tāmen duī de <u>chéngbǎo</u> hěn wěndang
3PL built DE castle very stable
'The <u>castle</u> they built is very stable.' *'<u>chéngbǎo</u>' in focus*

Another key advantage of this game is that participants' attention was mostly fixed on the missing keywords and the pronunciation of the actual targeted phrases with neutral tones, for instance *duī de* in (3), was very natural. The targeted phrases are further introduced and displayed in section 2.4 *Materials*, and see supplementary materials for a full list of targeted sentences.

2.2. Participants

Thirty-five young adults participated in this study. Twenty-one teenagers and young adults contributed to the Plastic Mandarin speech data (sixteen females and five males; mean age = 17.24 years, SD = 0.70 years). They were all students at the Changsha Nanya Middle School and recruited from pre-existing friendship groups. Fourteen of them were born and raised in Changsha; others 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.

Another fourteen young adults (nine females and five males; mean age = 24 years, SD = 1.96 years), recruited in Oxford, United Kingdom, participated in the recording of Standard Mandarin. 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. All participants are English learners and speak English at various levels. No history of speech or hearing disorders was reported. Written informed consent was acquired from subjects prior to recording.

2.3. Recording

Recordings of Plastic Mandarin were conducted in a small quiet room furnished with soundproofing materials in Changsha Nanya Middle School. 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. Recordings of Standard Mandarin were conducted in an audio studio with soft furnishings in the Phonetics Laboratory of the University of Oxford.

Each speaker in a 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 44,100 Hz and 16-bit quantisation. The same set of equipment was used for all pairs of participants.

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

2.4. Materials

Each channel of the audio recordings was extracted separately, and all the targeted sentences were manually annotated, transcribed, and extracted from long audio files that recorded interactions throughout the word-guessing game. The audio recordings and their transcriptions were then processed by the Penn Phonetics Lab Forced Aligner (P2FA) for Mandarin Chinese (Yuan & Liberman, 2008, 2015) to achieve automatic phonetic segmentation. Although P2FA's acoustic models were trained on Standard Mandarin, it can still be applied to Plastic Mandarin since the two varieties are highly related and their segment inventory is fairly similar. The automatic segmentation was subsequently checked and less than 5% of the sentences were noted as 'bad or wrong alignment'. These were manually corrected, following the practice in many studies (e.g. Sun & Shih, 2021; Yuan, 2012; Zue et al., 1990). Utterances with overlapping speech or laughter were not considered in further analysis.

Given the alignment, the relevant sound intervals can be extracted using the time stamps and labels. A custom-written Python script was used to automatically retrieve phrases containing neutral tones from audio recordings of targeted sentences. Then, 50 ms of silence was added to both edges of each extracted sound interval to prevent boundary f0 loss due to windowing in periodicity analysis. These zero-padded speech tokens constitute the audio data for subsequent f0 analysis.

Table 4 and Table 5 present a breakdown of the number of speech tokens by the lexical tone categories of X or X_1 in the datasets of Plastic Mandarin and Standard Mandarin. The slight imbalance in the number of repetitions can be dealt with in the statistical analysis.

Two subsets of phonetically controlled disyllabic sequences, in the form of the clitic *de* construction [X-*de*] as shown in Table 4, were embedded in the stimulus sentences. Each subset features a consistent vowel in the X syllables. Efforts were made to find X syllables with the same onset consonants and within the range of high frequency words. Given the tonotactic gaps, the onset consonants of X syllables in the /ueɪ/ set differs. All X syllables, except for *duì* 'redeemed', are within the top 1400 most commonly used characters (Xiao, 2009). There are 606 Plastic Mandarin tokens and 415 Standard Mandarin tokens in total.

The particle *de* is the most frequent word in Mandarin Chinese, ranked top with regard to both its frequency and dispersion in different registers in the 50-million-word corpus of Xiao (2009, p. 20). It marks modification in pre-nominal strings in various contexts and has several roles including marking possession, adjectives, prepositions, and nominalisation (Li & Thompson, 1981). Given its 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 (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 (Lin, 2010; Sproat & Shih, 1987).

Two groups of four-syllable sequences as shown in Table 5 were mined and acquired from the recordings. The first group contains a trisyllabic noun phrase containing two neutral tone syllables, followed by the Tone 4 particle *zhèng* indicating progressive aspect. Some analysed *zhèng* as an adverb 'just' (Paul, 2015). In this case, *zhèng* does not usually cliticise to the left. These trisyllabic noun phrases are the plural form of frequent kinship terms, in which the first two syllables are reduplicatives. The second syllable X_2 is considered as a duplicate of the first syllable X_1, but it has neutral tone. The second group contains the same trisyllabic phrases with an extra neutral tone syllable *de*, indicating possessive constructions. All these four-syllable sequences appear at the beginning of an eight-syllable sentence. There are 547 Plastic Mandarin tokens and 372 Standard Mandarin tokens.

2.5. Acoustic processing

A custom-written Python script was used to generate f0 estimates (pitch floor and ceiling were set at 75 Hz and 500 Hz;

Table 4
The number of disyllabic tokens containing the neutral tone syllable de in Plastic Mandarin (PM) and Standard Mandarin (SM). Pinyin transcriptions are based on Standard Mandarin.

	T1- *de*	T2- *de*	T3- *de*	T4- *de*	Total
	[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	
PM	68	78	69	82	297
SM	54	50	46	56	206
	[tueɪ tə] *duī-de* 堆的 stacked	[xueɪ tə] *huí-de* 回的 returned	[tʰueɪ tə] *tuǐ-de* 腿的 leg's	[tueɪ tə] *duì-de* 兑的 redeemed	
PM	87	73	60	89	309
SM	56	49	40	64	209

Table 5
The number of polysyllabic tokens containing consecutive neutral tone syllables in Plastic Mandarin and Standard Mandarin. Pinyin transcriptions are based on Standard Mandarin.

	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	70	69	59	262
SM	49	49	43	47	188
	[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	74	75	279
SM	42	47	42	51	182

window length was 40 ms) at a 10-millisecond frame rate within the voiced region of each audio token, employing the *Parselmouth* package (Jadoul et al., 2018) based on Praat's autocorrelation algorithm (Boersma, 1993). The script also detects and flags potential f0 measurement errors by identifying unlikely f0 jumps between adjacent voicing periods, and automatically removes boundary f0 estimates that have a large change from their neighbouring f0 and lay much farther away from the token mean f0. f0 estimates were also manually corrected for artifacts such as octave errors. It should be noted that mostly f0 estimates of modal voice portions were included here for further analysis and f0 measurements were often not available in parts with creaky voice or voiceless consonants. The output of the script was an ordered time series of f0 measurements. To allow inter-speaker comparison, these f0 values in Hz were normalised to semitones, with the speaker mean f0 as the base value for semitone conversion (Andruski & Costello, 2004). This normalisation method was considered the best among many others in preserving phonemic variation as well as sociolinguistic variation, according to Zhang (2018). The speaker mean f0 was established by averaging the mean f0 of each of the four citation tones in a Mandarin variety, using the monosyllabic citation tone data from the same person (for more details on the monosyllabic data, see Xu, 2022). The speaker mean f0 represents the centre of tonal space constructed from a balanced set of all citation lexical tones. For a detailed illustration of the calculation of the speaker mean, please refer to the supplementary materials.

2.6. Statistical technique

The current study used Generalised additive mixed models (GAMMs) that allow the inclusion of a time variable in dynamic f0 patterns to reveal how various phrases differ non-linearly with respect to how f0 unfolds over time. GAMMs have emerged as an ideal tool for dynamic phonetic analysis and have been effectively applied to temporally ordered measurements with autocorrelational structures such as formant tracks (e.g. Sóskuthy et al., 2019) and f0 contours (e.g. Chuang et al., 2021; Sun & Shih, 2021). GAMMs not only address nonlinearities in time series but also enable calibrated inferences in the presence of hierarchical data structure. Before applying GAMMs, time measurements were normalised.

2.6.1. Time normalisation

The corresponding time of each f0 measurement was normalised in a linear fashion within each syllable of each token. In other words, syllable boundaries were considered landmarks, and landmarks were aligned across tokens as well as Mandarin varieties. In this way, the overall f0 contours of the targeted phrases can be compared cross-dialectally, as well as the f0 contours of individual syllables at each position.

The acoustic realisation of the disyllabic sequence [X-*de*] varies from token to token. The initial consonant of *de*, a voiceless unaspirated plosive [t], was sometimes completely voiced between the vowels or even nearly deleted. The f0 contour of these tokens may be smooth and continuous or with a break when there is a voiceless stop closure. Such broken f0 curve is treated as a smooth and continuous curve where the break is interpolated. Therefore, the f0 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 normalised duration of each syllable is roughly based on the average duration of the f0 contours, so that the overall shape of a phrasal f0 contour was not greatly distorted. Table 6 lists the duration information of f0 contours by syllable position in both Mandarin varieties. The scales of [−1, 0] and [0, 1] were motivated for the time normalisation of f0 parts in the first two syllables given the ratio between the average duration of these two parts is approximately 1:1 in [X-*de*] phrases in both Mandarin varieties. Given that the duration of X in Table 6 only covers its rhyme, the neutral tone syllable *de* is on average shorter than the full tone syllables X.

In order to enable visual comparison between [X-*de*] phrases and [X_1X_2] phrases, the scales of [−1,0] and [0,1] were also motivated for the time normalisation of [X_1X_2], despite small duration differences between syllables X_1 and X_2. The scale of [1, 2.35] was motivated for the time normalisation of *men*, given that *men* is generally longer than X_1 and X_2, hence a normalised duration of 1.35 was assigned to *men*. The longer duration of *men* may be attributed to it being a closed syllable with a nasal coda, to potential boundary-related lengthening effects in [X_1X_2-*men* (*zhèng*)] phrases (the duration of *men* preceding *zhèng* is longer than that of *men* preceding *de*), or to both. The duration of *de* in [X_1X_2-*men*-*de*] phrases excluding phrase-final creaky pulses is 119 ms (±81.1 ms standard deviation) in Standard Mandarin and 138 ms (±89.2 ms standard deviation) in Plastic Mandarin. The scale [2.35, 3.3], with a normalised duration of 0.95, was selected for the time normalisation of *de* in such quadrisyllabic phrases across both Mandarin varieties, due to its slightly shorter average duration than *de* in disyllabic phrases.

2.6.2. Generalised additive mixed models (GAMMs)

GAMMs were fitted to the data to statistically assess differences among f0 contours of phrases containing neutral tones. GAMMs represent contours and their differences in both parametric and smooth terms and can detect regions where the contours diverge via the confidence intervals around the estimated difference curve. The parametric terms capture f0 height-related patterns that hold across the entire contour, and the smooth terms model patterns related to the shape of an f0 contour (Sóskuthy, 2021).

Table 6
The mean and standard deviation of the duration of f0 contours by each syllable in Standard Mandarin and Plastic Mandarin. [X_1-X_2-*men*] includes the first three syllables in both [X_1-X_2-*men-zheng*] and [X_1-X_2-*men-de*] phrases. Note that the duration of any voiceless consonant in Syllable 1 is not included.

	Phrase	Syllable 1	Syllable 2	Syllable 3
SM	[X-*de*]	146 (±47.9) ms	145 (±99.6) ms	–
PM	[X-*de*]	147 (±47.2) ms	140 (±82.4) ms	–
SM	[X_1-X_2-*men*]	148 (±56.3) ms	130 (±35.3) ms	198 (±56.1) ms
PM	[X_1-X_2-*men*]	152 (±58.9) ms	128 (±38.7) ms	186 (±58.7) ms

In this study, GAMMs with difference terms (Sóskuthy, 2021) were adopted, whereby a baseline smooth was fitted to the f0 contours of one phrase and other smooths represented the non-linear f0 differences of other phrases to the baseline. 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 models, the dependent variable was the normalised f0, modelled as a smooth function over normalised time and represented by a weighted sum of thin plate regression splines (S. N. Wood, 2003) with a smoothing parameter. Two groups of models were built: the first group featured within-variety f0 contour patterns, and the second group modelled the cross-variety effects. The following models belong to the first group and were fitted to the dataset of each Mandarin variety separately. Please refer to the supplementary materials for the model specifications and selection, with the relevant R code.

In the X-*de* models, the independent variables included a smooth over the normalised time, a smooth over the overall duration measured in seconds, a tensor product interaction (*ti()*) between overall duration and normalised time, to account for the influence of duration on the shape of the f0 trajectories (see Sóskuthy, 2017). The two main predictors, the ordered factors of the tone category of X (i.e. T1, 2, 3, 4) and vowel of X (i.e. /ueɪ/ and /a/), were included as separate parametric terms to capture overall differences between the trajectories, separate difference smooths over the normalised time, as well as separate difference surfaces to the tensor product interaction. Moreover, a random smooth over the normalised time with a special grouping variable was included. This grouping variable created 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*) allows potential speaker-specific trends in the main effects.

In the X_1X_2-*men* models, the syllable following *men* (i.e. *zhèng* and *de*) was manipulated, and included as a parametric term, a difference smooth over the normalised time, and a difference surface to the *ti()* interaction, so was the predictor of the tone category of X_1 (i.e. T1, 2, 3, 4). Other independent variables were: a smooth over the normalised time, a smooth over the overall duration measured in seconds, a *ti()* interaction between overall duration and normalised time, and lastly, a random smooth over the normalised time with a special grouping variable that creates separate levels for each combination of speaker, X_1 tone category and the following syllable.

In the X_1X_2-*men-de* models, the main predictor was the tone category of X_1 (i.e. T1, 2, 3, 4) and was included as a parametric term, a difference smooth over the normalised time, and a difference surface over to the *ti()* interaction. Similar to other models, other independent variables were: 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, and a random smooth over the normalised time with a grouping variable of Speaker to capture nonlinear individual differences.

Whereas fitting separate models for each variety and visually comparing the f0 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 language variety is warranted. Thus a second group of models was built, using the combined dataset of Plastic Mandarin and Standard Mandarin. This group focused on cross-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 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 or X_1 tone category in the previous models, while the rest of specification remained the same, except that the random factor smooth has a two-way grouping variable *Speaker* × *Vowel* in the X-*de* models and *Speaker* × *Following syllable* in the X_1X_2-*men* models.

GAMMs were constructed using the *bam* function of the *mgcv* package (Wood, 2022). When analysing time series data such as f0, the residuals (i.e. the difference between the observed and the estimated values) are generally autocorrelated and the associated *p*-values anti-conservative. Since autocorrelation in f0 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). Fig. 2 illustrates the averaged residuals for all trajectories in the GAMMs with and without the AR1 component for the Standard Mandarin disyllabic dataset. Fig. 2(a) suggests that there is high positive autocorrelation at lag 1. In Fig. 2(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).

Within each group, multiple models were built. After initial model-fitting, diagnostic tests were performed using the gam.*check()* function. In addition to the strategy to deal with the inevitable autocorrelation in the residuals, we inspected the normal distribution and homoscedasticity of the residuals, using the *check_resid()* command from the *itsadug* package (van Rij et al., 2022). Fig. 2(c) presents an example quantile–quantile (QQ) plot for the X_1X_2-*men* model for the 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. 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 Fig. 2(d). The overall shape of predicted f0 contours is very similar across models with Gaussian and *scaled-t* distributions. The discussion of f0 contour comparisons in the following sections is mainly based on models with *scaled-t* distributions. The figures were created using the packages *itsadug* (van Rij et al., 2022) and *tidymv* (Coretta, 2023). All model specifications and results are available in the supplementary materials.

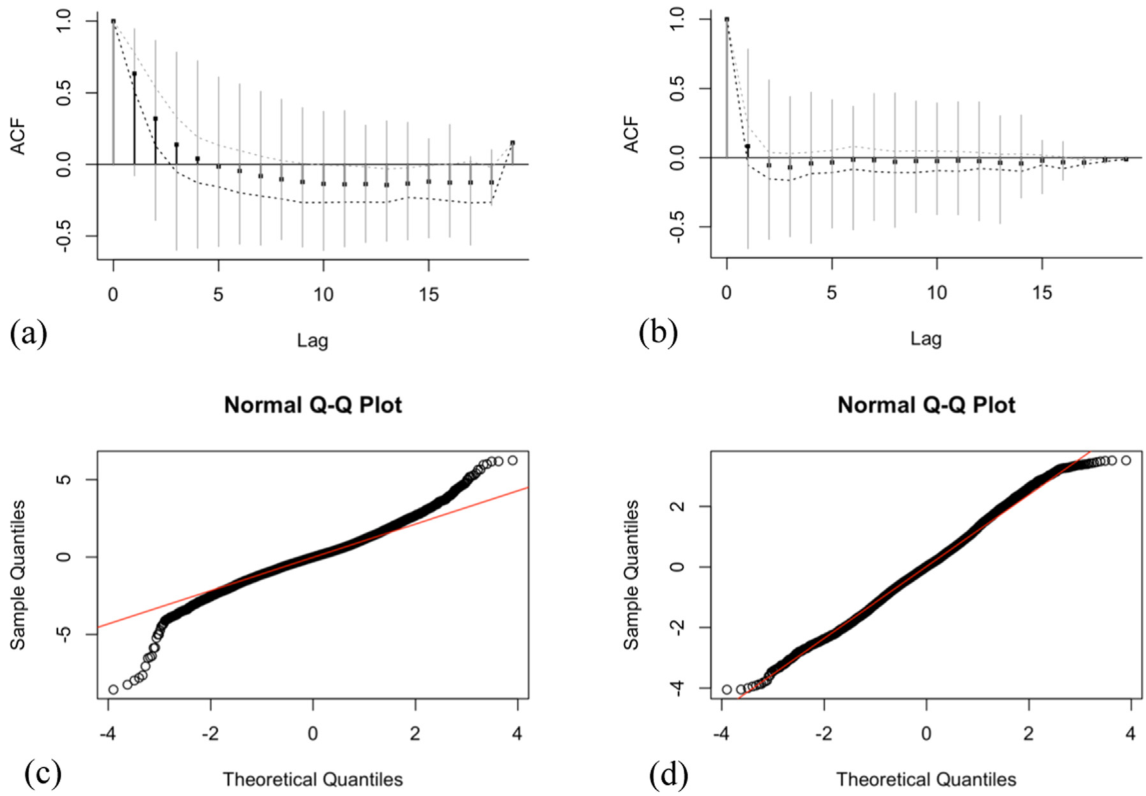


Fig. 2. Diagnostic plot examples. Autocorrelation in the residuals of the models for Standard Mandarin disyllabic tokens: (a) without correction, (b) after correction using the AR1 error model, lag indicates the number of time steps; the quantile–quantile plot of the residuals of the models for Standard Mandarin trisyllabic tokens: (c) a Gaussian model, (d) a scaled-*t* model.

2.7. Reproducibility

The full set of targeted stimuli sentences, f0 data, preprocessing scripts, and various model specifications and results (supplementary materials) are publicly accessible via the Open Science Framework at https://osf.io/3cjqp/.

3. Results

3.1. Neutral tone syllable de in disyllabic sequences

Fig. 3 presents the fitted normalised f0 contours for each of the eight [X-*de*] phrases in the separate models of Standard Mandarin (a) and Plastic Mandarin (b). Recall that the first half of each graph, the normalised time domain [−1, 0], is for f0 contours of the lexical tone syllable X and the second half, the normalised time domain [0, 1], for the neutral tone syllable *de*. By visually 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 tone language.

3.1.1. The lexical tones preceding de

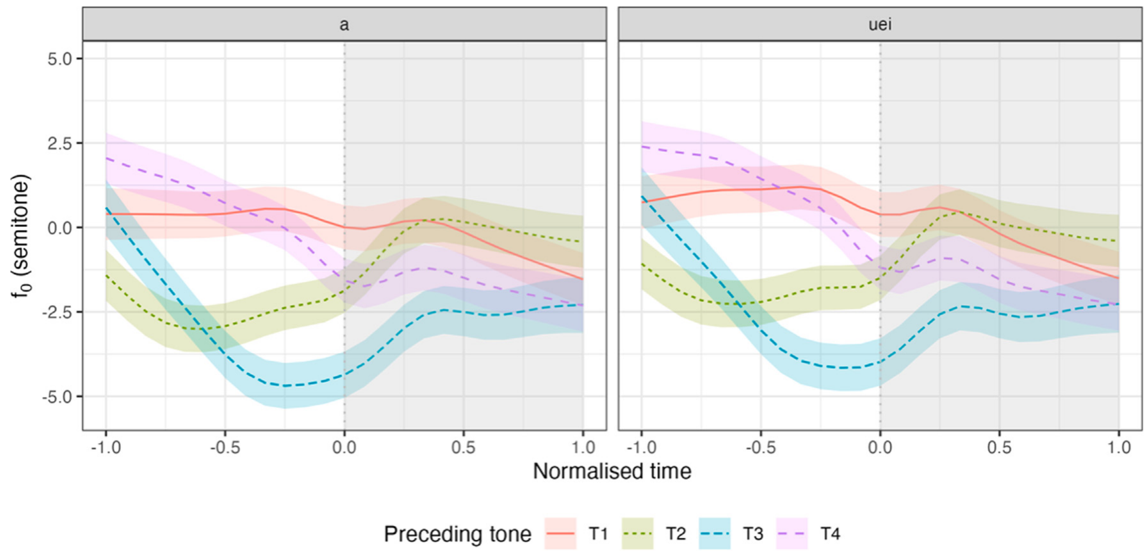
The first halves of the f0 contours shown in Fig. 3 largely resemble the patterns of their corresponding prototypical citation tones in both Mandarin varieties, illustrated in Fig. 1. This reassures that the production of Plastic Mandarin and Standard Mandarin has been successfully elicited from the two groups of speakers.

For Standard Mandarin, Tone 1 here is flat and relatively high compared to Tones 2 and 3, although the estimated normalised f0 is only slightly higher than 0 (i.e. speaker mean) throughout the syllable. The prototypical citation Tone 1, estimated on the same group of speakers, has an average normalised f0 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 syllables in the stressed position of disyllabic sequences largely maintain their 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.

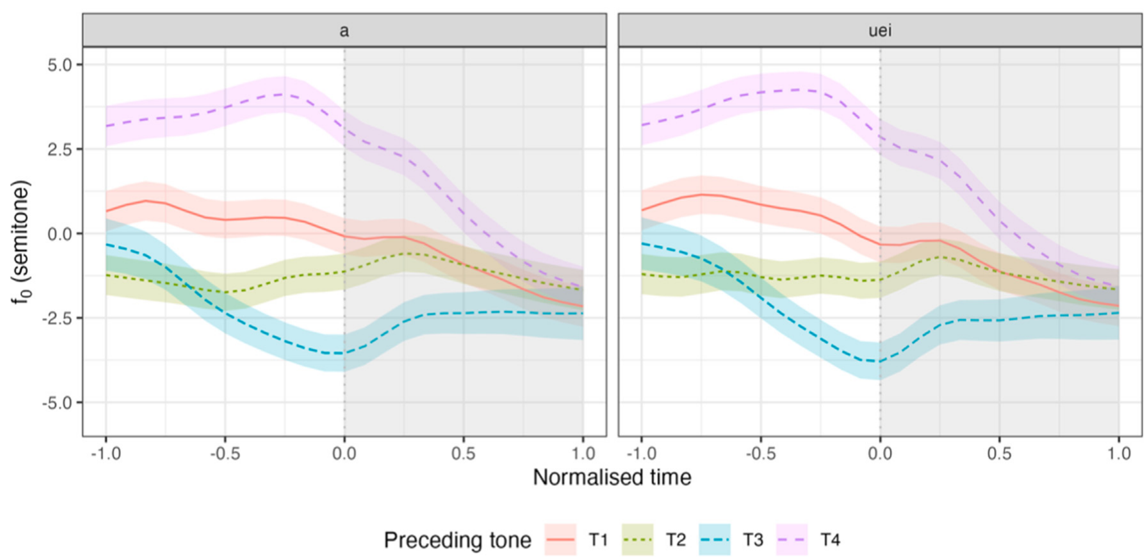
For Plastic Mandarin, the high rising Tone 4 almost reaches as high as 5 semitones above the speaker mean, consistent with its citation form. Other tones have slight differences compared to their citation form: Tone 1 here slightly falls, Tone 2 tends to be level, and the overall f0 range of Tone 3 is shifted higher than that of its citation form.

3.1.2. f0 contours of neutral tone syllable de

The fitted f0 contours of *de* in Fig. 3 exhibit various patterns. The end of an f0 contour in Fig. 3(a) and Fig. 3(b), though, tend



(a) Standard Mandarin /(t)a tə/ and /(C)uei tə/



(b) Plastic Mandarin /(t)a tə/ and /(C)uei tə/

Fig. 3. Fitted f0 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 (duration: 250 ms). The colourful shades indicate 95% pointwise confidence intervals. The dotted line at 0.0 indicates syllable boundary. The normalised time domain [−1, 0] shows f0 contours of the lexical tone syllable X and [0, 1] the neutral tone syllable *de* (in grey background).

to cluster in the approximate f0 range of [−3, 0] and [−3, −1] respectively. The converging effect in Plastic Mandarin tends to be visually more obvious than that in Standard Mandarin. The neutral tone patterns in Plastic Mandarin do not resemble those in Changsha depicted in Zhong (2003).

The difference curves are useful to check whether the f0 differences at the end of *de* among different groups are statistically significant. Fig. 4 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, the absolute value of estimated f0 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 the last 4% of *de* in the comparison of the f0 difference between T1-*de* and T4-*de* phrases. This suggests that the f0 contours converge at the very end of *de* for these pairs of phrases. The f0 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 have a large f0 excursion at the end of the syllable, thus it may take a longer time for contours of *de* with a larger excursion of the f0 onset to converge.

In Standard Mandarin, only the estimated f0 difference towards the end of *de* between T1-*de* and T2-*de*, T1-*de* and T3-*de*, and T3-*de* and T4-*de* phrases is not statistically significant. The results of both Mandarin varieties show that there is a converging trend for different contours but the f0 variance is still large at the end of the neutral tone syllable *de*.

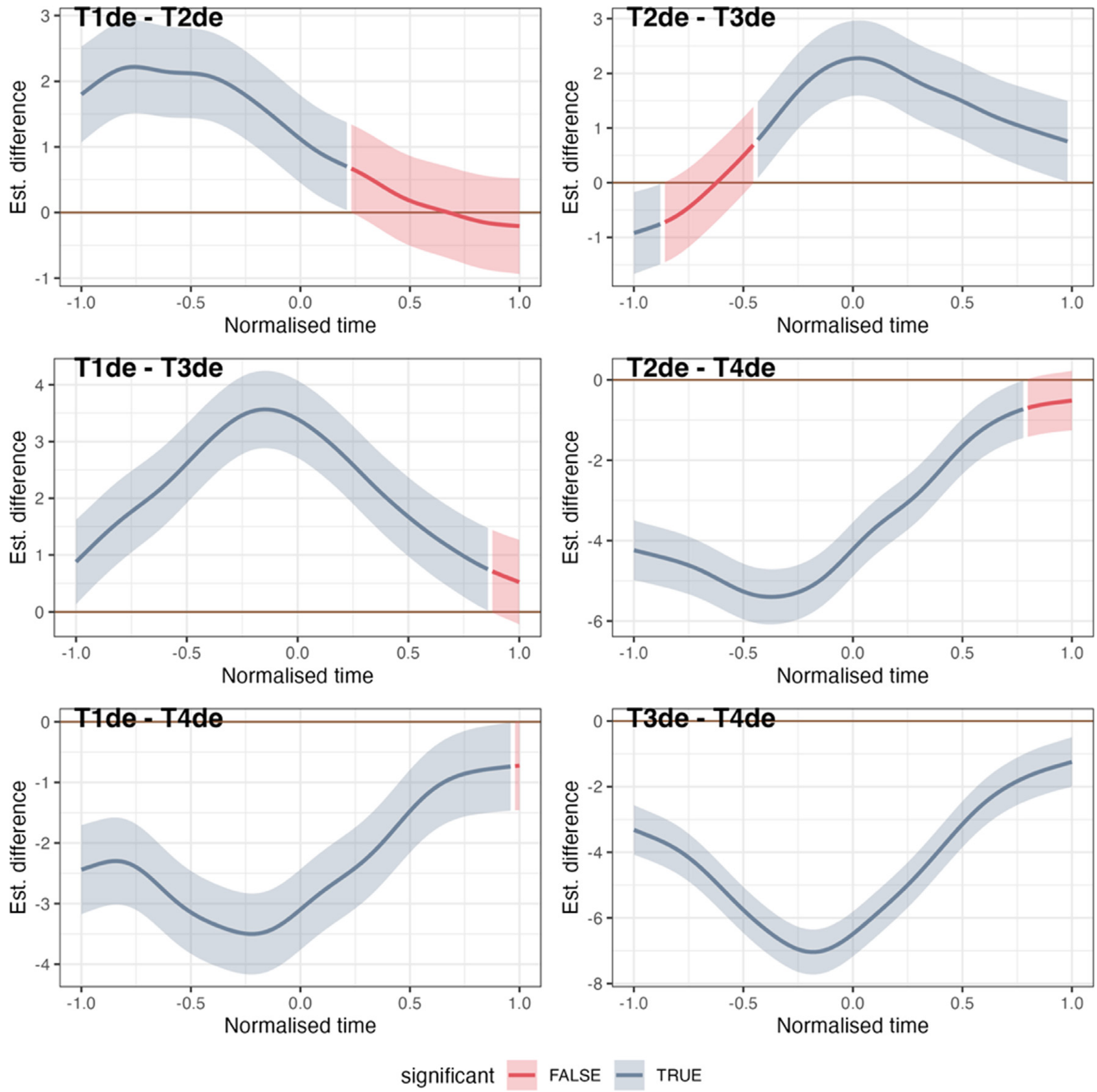


Fig. 4. The estimated difference f0 trajectories between each two groups of the disyllabic Plastic Mandarin tokens /ta tə/ with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.3. The duration effect

While the falling Tone 3 in Standard Mandarin in Fig. 3 may correspond to the low variant of Tone 3 in the literature, the estimated f0 curve of [T3-*de*] phrases reveals a carryover effect of Tone 3 on the neutral tone — the canonical dipping shape of Tone 3 seems to continue into the neutral tone syllable and is completed at about 30% into the second syllable. The end of such an 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 f0 perturbation after a stop consonant.

In order to understand the estimated f0 curve between the normalised time domain [0, 0.3], some original tokens were chosen at random for further examination. Fig. 5 demonstrates three tokens of Standard Mandarin *dǎde* with different lengths produced by three speakers. Token (a) has a 60 ms gap (42% of the duration of *de*) without f0 measurement, while Token (b) and (c) have a continuous f0 transition from the first syllable to the second syllable. In Fig. 5(a) the initial f0 point of *de* is about 28 Hz (2.3 semitones) higher than the offset f0 point of the first syllable. In Fig. 5(b), f0 starts to rise after 0.2 s and *de* also has an onset f0 that is 21 Hz (3.84 semitones) higher than the f0 trough of the first syllable. In Fig. 5(c), the starting f0 of *de* follows the f0 fall in the first syllable, and it continues to go lower during the majority of the voicing portion of the consonant [t̪]. The three example tokens have demonstrated a diverse realisation in terms of both segments and f0 contours. They might, however, reveal the general case that the starting f0 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. Interestingly, the carryover rising tail after a T3 occurred in the Plastic Mandarin T3-*de* tokens too, although the citation form of T3 in Plastic Mandarin is low falling.

Notice also that the first syllable *dǎ* in Fig. 5(a) and (b) is longer than that in (c). The shorter the first syllable, the more carryover effect is reflected in the f0 contour of *de*. Tokens

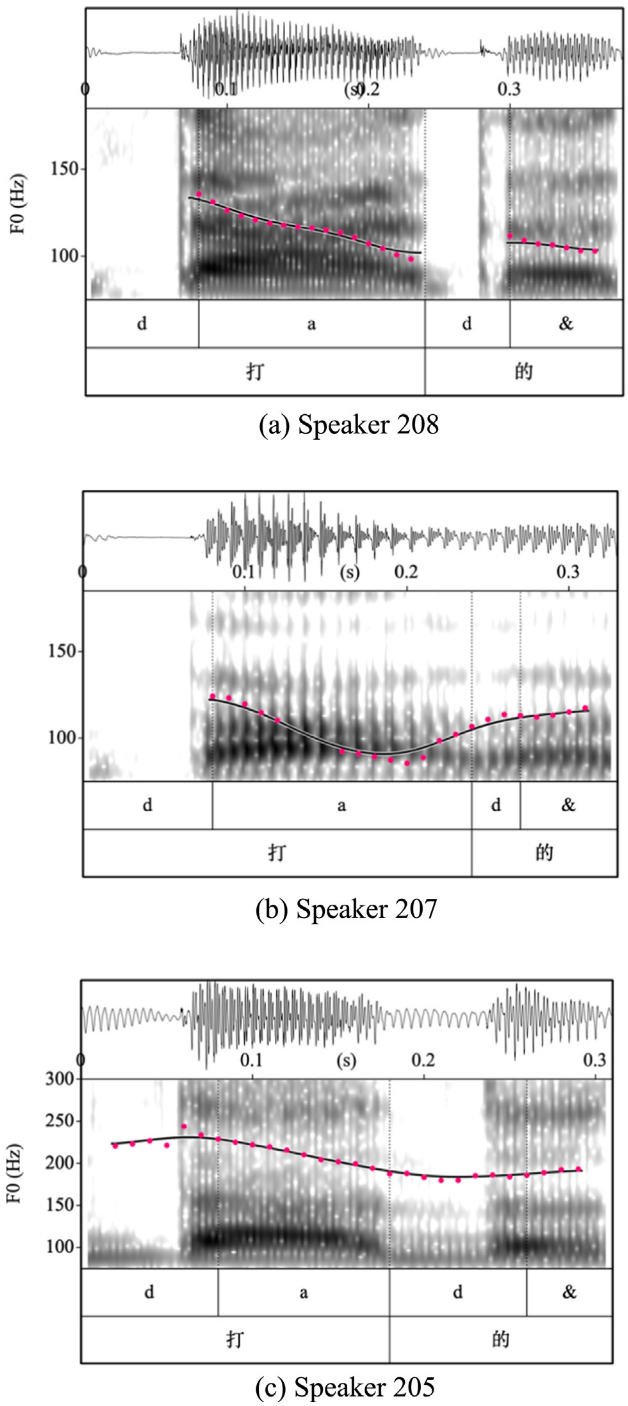


Fig. 5. The waveform, spectrogram (0–5000 Hz), and f0 measurement of three tokens of *dǎde* 'made' by three speakers. TextGrid alignment is automatically generated by P2FA. Black line is the smoothed f0 curve. Pink dots indicate f0 measurement. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(b) and (c) are shorter overall, and their f0 contour of *de* is different from that of token (a), which has a negative slope (f0 gradually decreases).

The model has included a continuous variable of overall duration. Fig. 6(a) illustrates the nonlinear interaction surface of overall duration and normalised time for *dǎde* tokens in Standard Mandarin with a GAMM heatmap, created using the *fvisgam* function of the *itsadug* R package (van Rij et al., 2022). Fig. 6(b) shows two example f0 contours estimated with different total duration: 0.25 s and 0.32 s. From Fig. 6, we can see that in longer tokens, the f0 of *de* is predicted to decrease, reflected by the darker green colour towards 1 in Fig. 6(a) and falling grey curve in Fig. 6(b). Conversely, the f0 of *de* in shorter tokens is predicted to rise.

The duration effect in Plastic Mandarin, as illustrated in Fig. 7, is based on the analysis of tokens *dàde* 'big'. For the longer tokens, f0 drops more in the *de* part, especially in the normalised time range from 0.5 to 1. This is shown in the Fig. 7(a), where 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 f0 change. The f0 height at the end of *de* in these two tokens is directly reflected in Fig. 7(b), and the estimated f0 difference is approximately one semitone.

3.2. Neutral tones in trisyllabic sequences

The trisyllabic phrases were composed of a reduplicative kinship term followed by the plural particle *men*, in which both the duplicate syllable X_2 and *men* are considered as neutral tone syllables. Fig. 8 presents the fitted normalised f0 contours for each of the four phrases in the separate models of Standard Mandarin (a) and Plastic Mandarin (b).

3.2.1. The opaque influence of the following syllable

Fig. 8 separates the estimated f0 contours according to the following syllable of *men*: a neutral tone syllable *de* or Tone 4 syllable *zheng*. A Tone 4 syllable in both Mandarin varieties starts at a high f0 in citation form. The Tone 4 syllable is, therefore, the best choice among other lexical tones to investigate whether the following tone influences the f0 contour of a neutral tone syllable. In other words, given its high starting f0, Tone 4 may elicit an evident anticipatory assimilation effect as reported in Chen & Xu (2006), or interpolation effect if the hypothesis is that the f0 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. The model summary statistics about the parametric term of the following syllable and its relevant smooth terms in the AR1 corrected model with *scaled-t* distribution were extracted using the *summary()* command in *mgcv* package, to check whether they are different from zero. In Standard Mandarin, the parametric term ($t(1) = -0.298$, $p = 0.77$) and the difference *ti()* term ($F(1, 5.04) = 1.43$, $p = 0.17$) are not statistically significant, but the smooth term ($F(1, 5.86) = 2.90$, $p = 0.005^{**}$) is statistically significant. Similarly in Plastic Mandarin, the parametric term ($t(1) = -0.92$, $p = 0.36$) and the difference *ti()* term ($F(1, 3.77) = 0.60$, $p = 0.75$) are not statistically significant, but the smooth term ($F(1,6.14) = 2.21$, $p = 0.03^{*}$) is statistically significant. This suggests that the expected differences are not in f0 height but in f0 shape.

We then plotted the estimated difference curve between the two sets under each condition of X_1 tone category with the associated 95% pointwise confidence interval. Fig. 9 demonstrates the difference curve between the *zheng* set and the *de* set. We can see that in the normalised time range [1.5, 2] in both varieties, the estimated f0 difference (concave up

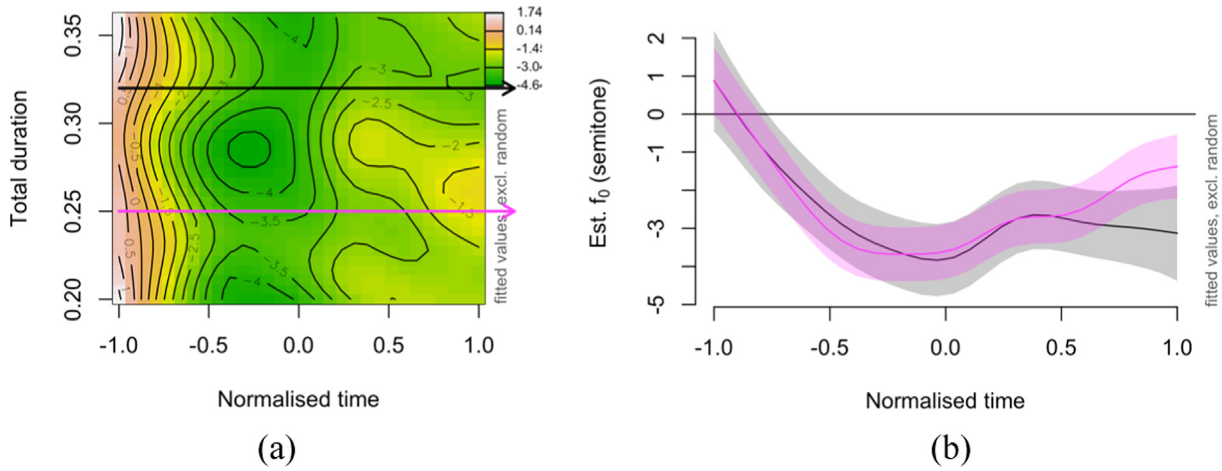


Fig. 6. The effect of duration across the normalised time estimated by the GAMM for Standard Mandarin disyllabic tokens of *dǎde*. (a) the interaction surface between overall duration and normalised time; The 0.2 to 0.8 quantile of the whole range of total duration is shown. f0 in semitone is denoted by colour grade, and regions of equidistant change (here, Δ 0.5 semitones) are denoted by black lines. (b) two estimated f0 contours with two different total duration: 0.25 (pink) and 0.32 (black). The corresponding duration values are marked by arrows in (a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

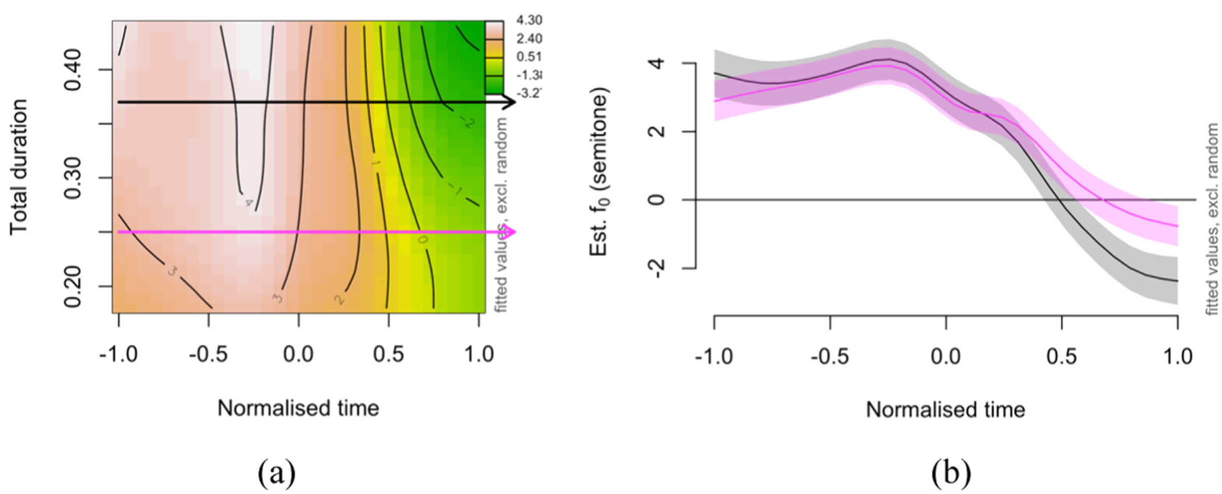


Fig. 7. The effect of duration across the normalised time estimated by the GAMM for Plastic Mandarin disyllabic tokens of *dàde*. (a) the interaction surface between overall duration and normalised time; The 0.1 to 0.9 quantile of the whole range of total duration is shown. f0 in semitone is denoted by colour grade, and regions of equidistant change (here, Δ1 semitones) are denoted by black lines. (b) two estimated f0 contours with two different total duration: 0.25 (pink) and 0.37 (black). The corresponding duration values are marked by arrows in (a). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

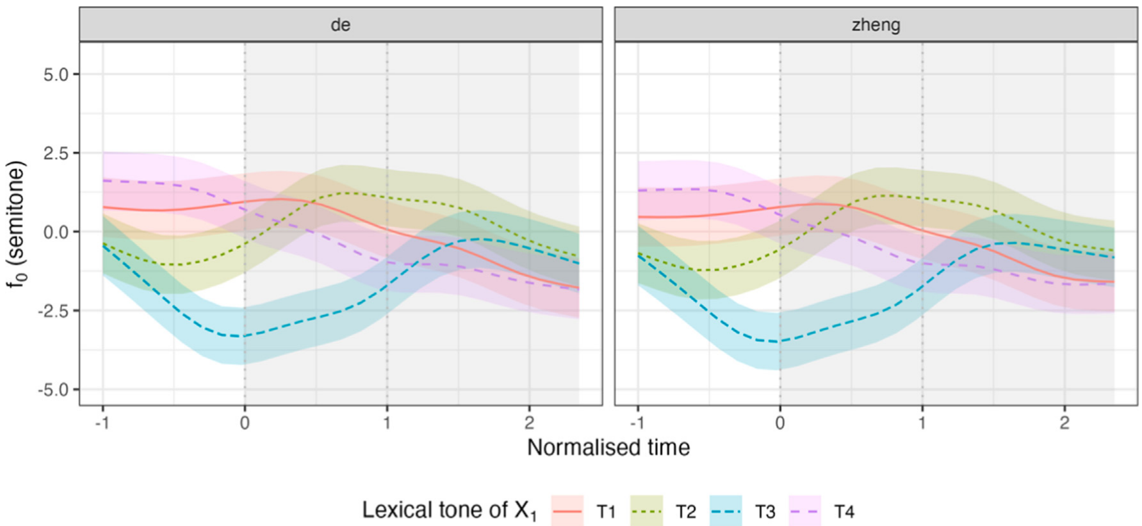
shape in the difference curve) is slightly larger than that in [1, 1.5], which reveals that the f0 may fall slightly lower then rise at the end in the *zheng* set, so as to gain a positive velocity to reach the upcoming high tone. However, no regions in any of the plots excludes zero (not statistically significant). We also built an alternative model examining the interaction effect of the preceding lexical tone X_1 and the following tone of *men* (see supplementary materials for model details) and none of the difference curves by each tone category of X_1 have any regions indicating statistically significant difference.

To conclude, we did not have sufficient statistical evidence supporting the significant differences in f0 contours in the two sets. This reveals that the following Tone 4 has little impact on the f0 contour of these trisyllabic phrases in both Mandarin varieties.

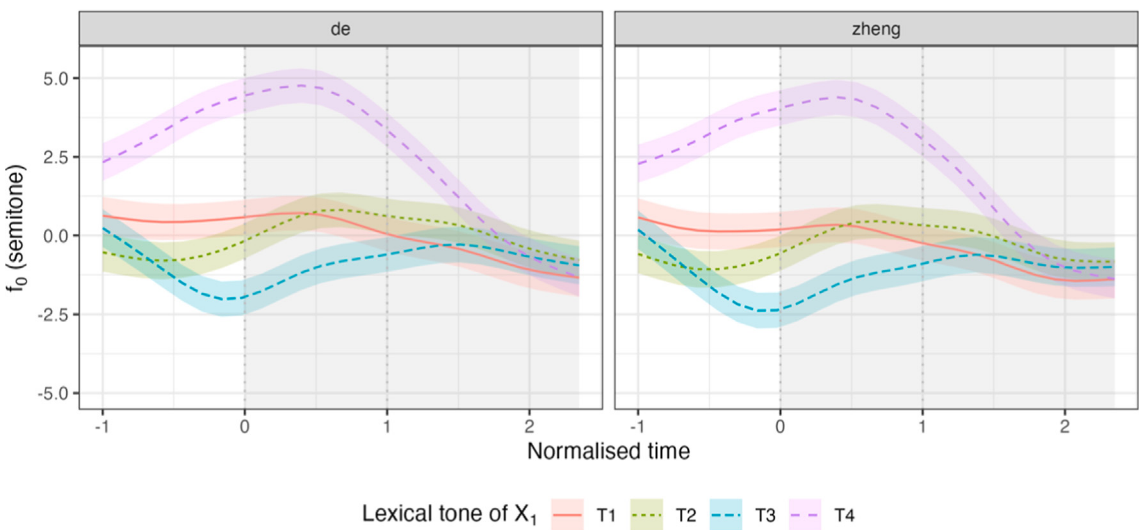
3.2.2. The duplicate syllable in neutral tone

In both Mandarin varieties, the f0 patterns of the duplicate syllable X_2 in the medial position of the trisyllabic sequences are different from those of *de* in the disyllabic sequences, 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.

In the normalised time range [−1,1] for X_1X_2 in Standard Mandarin in Fig. 8(a), we can clearly find four canonical citation tone patterns: a high falling pattern (T4), a mid-to-high rising pattern (T2), a fall-rise dipping pattern (T3), and a (relatively) high level pattern (T1). The high falling pattern here does not fall as much as the citation Tone 4. Similarly in Fig. 8(b), the tone patterns in [−1, 1] mostly show the prototypical citation tone patterns of Plastic Mandarin: a high rising pattern (T4), a mid level pattern (T1), and a mid rising pattern (T2). 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, f0 in T3-initial phrases *nainai-men* ‘grannies’ falls in the first syllable, then rises in the second syllable, although it falls to a lower



(a) Standard Mandarin



(b) Plastic Mandarin

Fig. 8. Fitted f0 contours of four [X_1X_2-men] phrases, grouped by the following syllable of men, by two different groups of Mandarin speakers (duration: 460 ms). The colourful shades indicate 95% pointwise confidence intervals. The dotted lines at 0 and 1 indicate syllable boundary. The normalised time domain [−1, 0] shows f0 contours of the lexical tone syllable X_1, and [0, 1] and [1, 2.35] for the neutral tone syllables X_2 and *men* (in grey background).

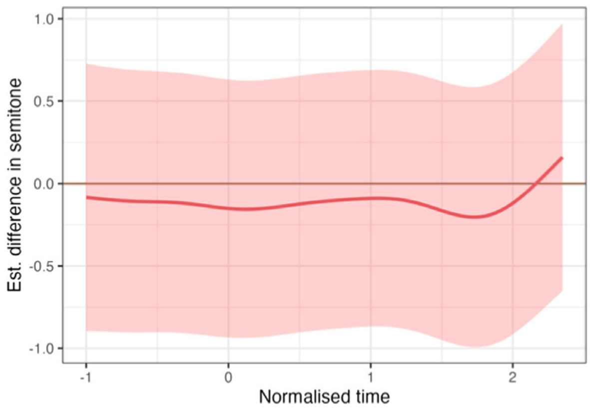
f0 in Standard Mandarin than in Plastic Mandarin. In fact, except for the T4-initial phrases *meimei-men*, the estimated f0 contours of the other phrases in Plastic Mandarin resemble those of the corresponding phrases in Standard Mandarin.

The above observations suggest that the reduplication tends to provide increased duration to realise the f0 contour of the lexical tone of the base syllable X_1. In other words, the lexical tone of X_1 has much influence on the f0 contour of the following neutral tone syllable X_2, and such influence continues into the second neutral tone syllable *men*.

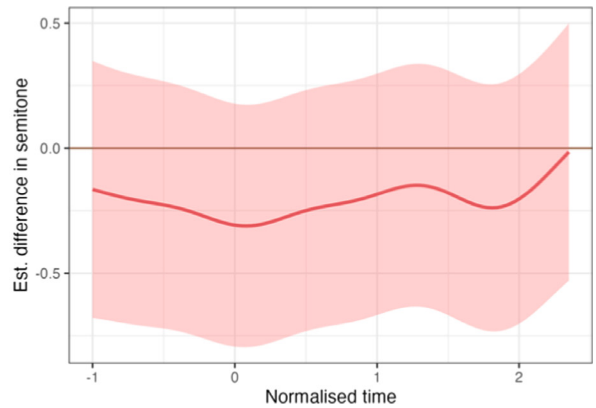
3.2.3. The converging contours of men

Figs. 10 and 11 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 7% of all the difference curves, the normalised time domain [2.11, 2.35], the differences in f0 between any two groups are statistically not significant. For Plastic Mandarin, in the last approximately 12% of all difference curves, the normalised time domain [1.94, 2.35], the differences in f0 between any two groups are statistically not significant. This suggests that different f0 contours converge towards the end of the neutral tone syllable *men* in both Mandarin varieties.

Figs. 12 and 13 show the distribution, kernel density estimation of the final f0 point of *men* over all tokens in Standard Mandarin and Plastic Mandarin respectively. A uniform Gaussian distribution was fitted to the ending f0 points in each dataset. The Gaussian model suggests an average ending f0 of −1.25 semitones (σ = 2.55 semitones) across all [X_1X_2-*men*] phrases in Standard Mandarin, and an average ending f0 of −1.00 semitones (σ = 2.05 semitones) in Plastic Mandarin.



(a) Standard Mandarin



(b) Plastic Mandarin

Fig. 9. The estimated overall f0 difference trajectory in the *zheng* set and the *de* set in each Mandarin variety, with the associated 95% pointwise confidence intervals. The entire area is not significantly different from 0.

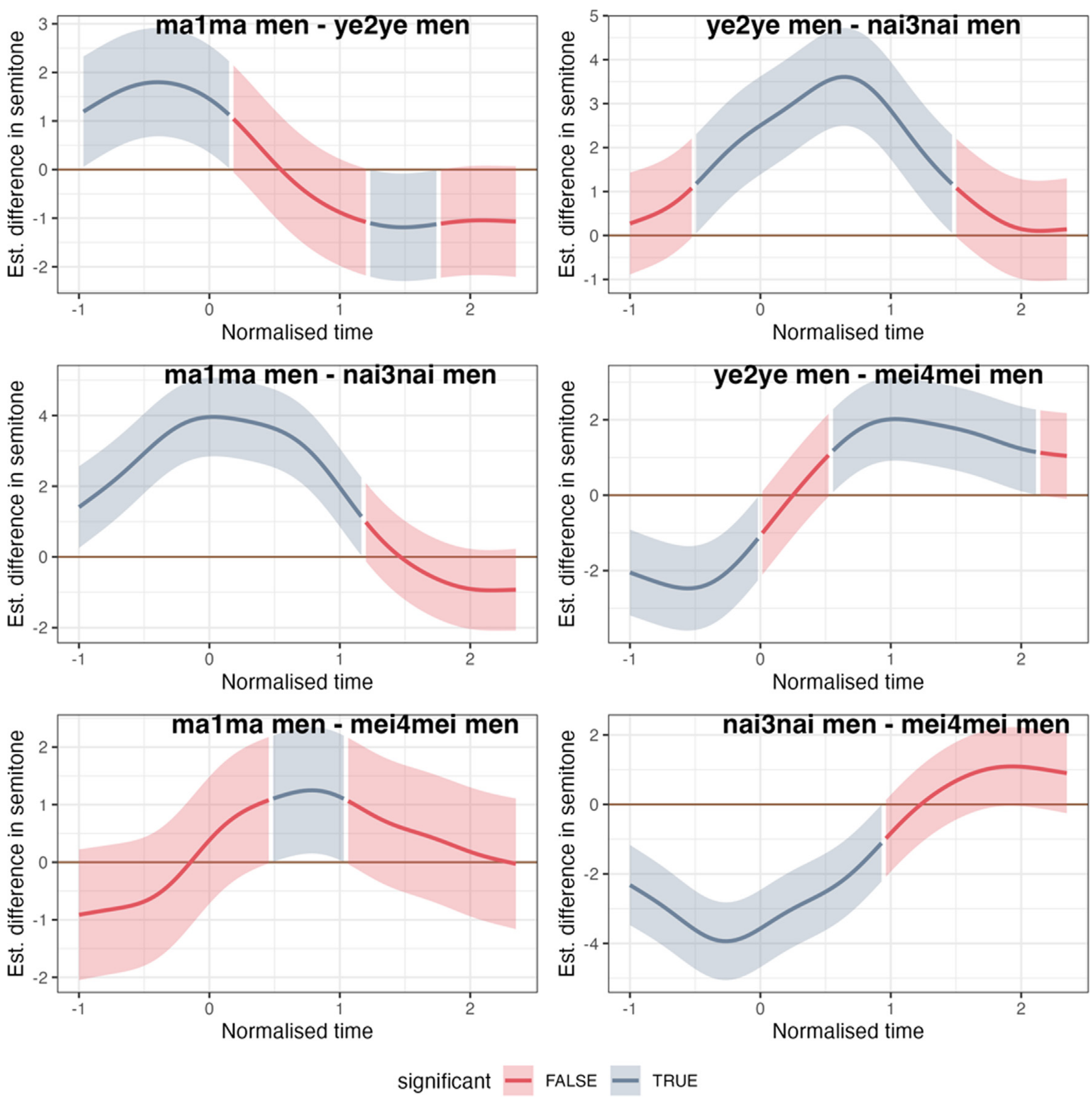


Fig. 10. The estimated difference f0 trajectories between each two groups of the trisyllabic Standard Mandarin tokens in the *de* set, with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red. The number in the pinyin transcription indicates X_1 tone category. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

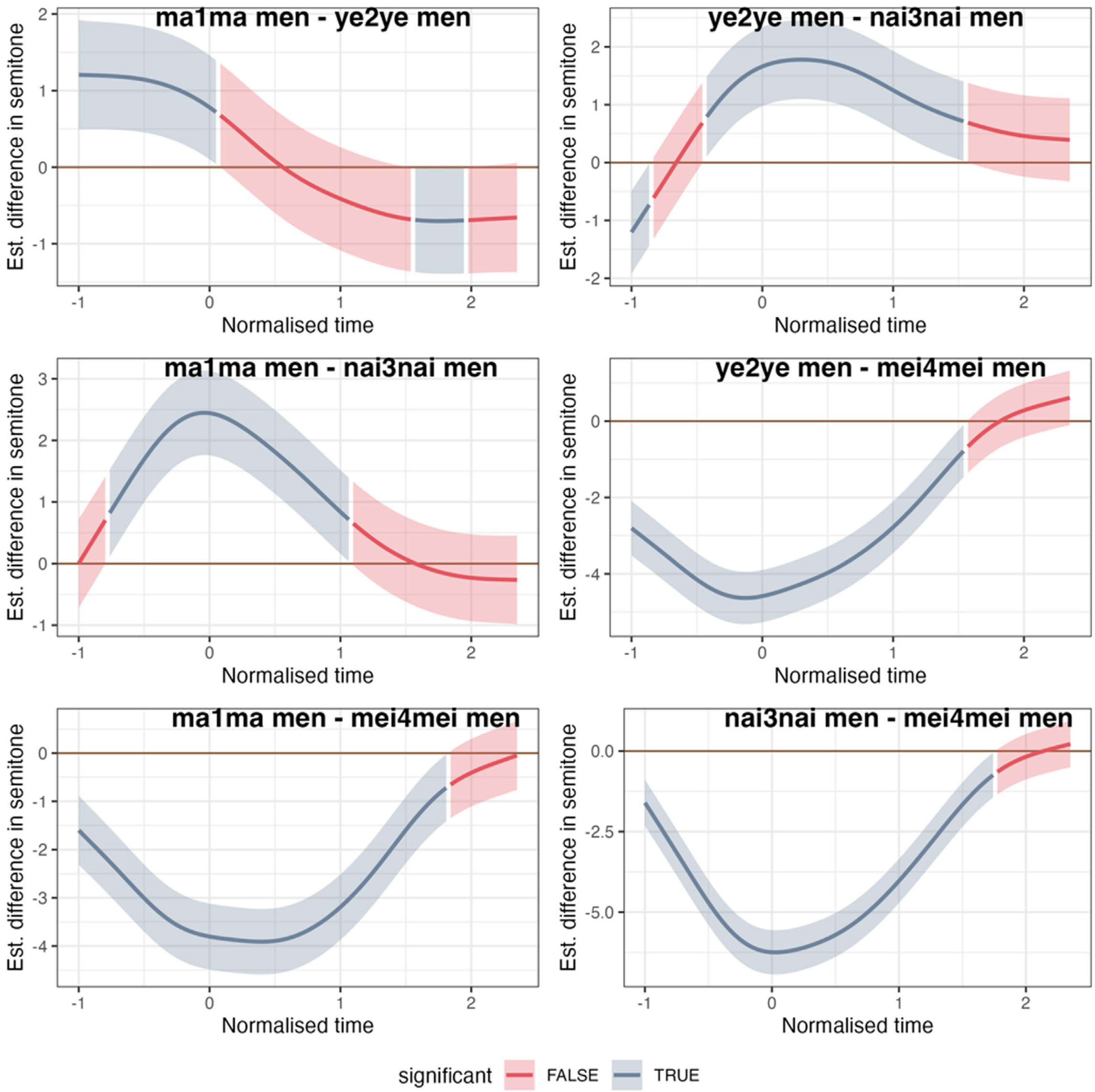


Fig. 11. The estimated difference f0 trajectories between each two groups of the trisyllabic Plastic Mandarin tokens in the *zheng* set, with the associated 95% pointwise confidence intervals. The areas where the difference is not significantly different from 0 were marked in red. The number in the pinyin transcription indicates X_1 tone category. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Neutral tones in quadrisyllabic sequences

This section continues to examine the subset of [X_1X_2-*men*] phrases, those followed by *de*, [X_1X_2-*men*-*de*], in which three consecutive neutral tone syllables follow a lexical tone syllable X_1. From the previous section, we have learnt that the four different f0 contours converge at the end of *men* in both Mandarin variety. This means that the starting f0 of different contours for the last neutral tone syllable *de* is similar. Will the f0 contours of *de* be flat if the contours have already converged at a common pitch target?

Fig. 14 shows the model predictions for the four quadrisyllabic phrases in each Mandarin variety. In each sub-figure, the shape of all f0 contours of *de*, in the normalised time domain from 2.35 (marked by grey dotted reference line) to 3.3, is similar – the f0 contours continue to fall, and their confidence bands overlap with one another. Such a homogeneous f0 pattern of *de* forms a contrast with the diverse f0 patterns of the same syllable *de* when it is the second syllable in the disyllabic sequences. This suggests that, in both Mandarin varieties, the lexical tone of X_1 has little impact on the f0 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 f0 point of *de*. Fig. 15 displays the kernel density estimations of f0 at the end of *de* in Standard Mandarin and Plastic Mandarin respectively. A uniform Gaussian distribution was fitted to the ending f0 points in each dataset. The Gaussian model suggests an average ending f0 of −2.19 semitones (σ = 2.54 semitones) across all [X_1X_2-*men*-*de*] phrases in Standard Mandarin, and an average ending f0 of −2.31 semitones (σ = 2.21 semitones) in Plastic Mandarin. In both Mandarin varieties, the mean ending f0 of *de* is lower than that of *men*, which signals that consecutive neutral tones allow ample time to reach a lower pitch target, similar to the duration effect of a longer neutral tone syllable.

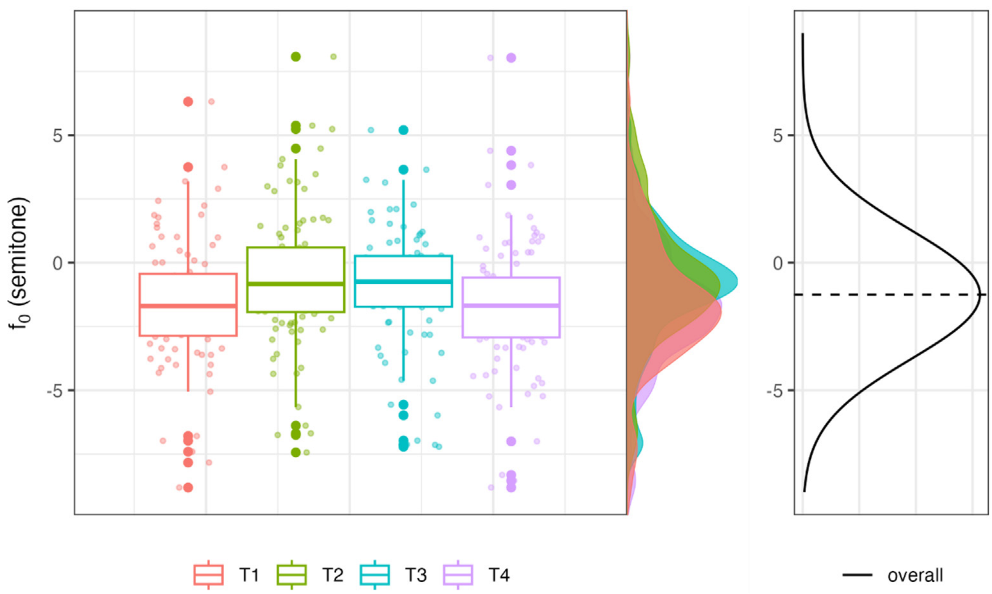


Fig. 12. The distribution, kernel density estimation of ending f0 of *men* in trisyllabic Standard Mandarin tokens and the fitted Gaussian distribution.

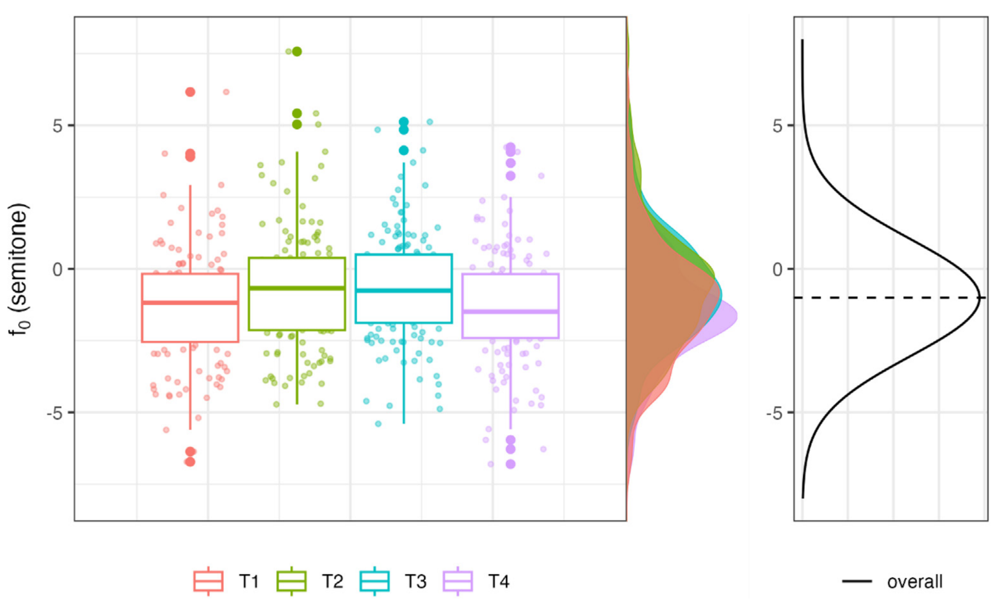


Fig. 13. The distribution, kernel density estimation of ending f0 of *men* in trisyllabic Plastic Mandarin tokens and the fitted Gaussian distribution.

3.4. Cross-dialectal comparison

Through GAMMs built on the combined dataset of two Mandarin varieties, this section mainly investigates whether the differences between Standard and Plastic Mandarin in the observed converging f0 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?

Figs. 16 and 17 present the estimated f0 trajectories of the four groups of trisyllabic and quadrisyllabic phrases in both Mandarin varieties respectively, and the estimated f0 difference trajectories between pairs of the same phrase in the two Mandarin varieties. In both figures, Plastic Mandarin contours have much overlap with Standard Mandarin contours at the end of the normalised tokens.

For trisyllabic phrases, Fig. 16(b) suggests that the f0 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 f0 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 f0 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*.

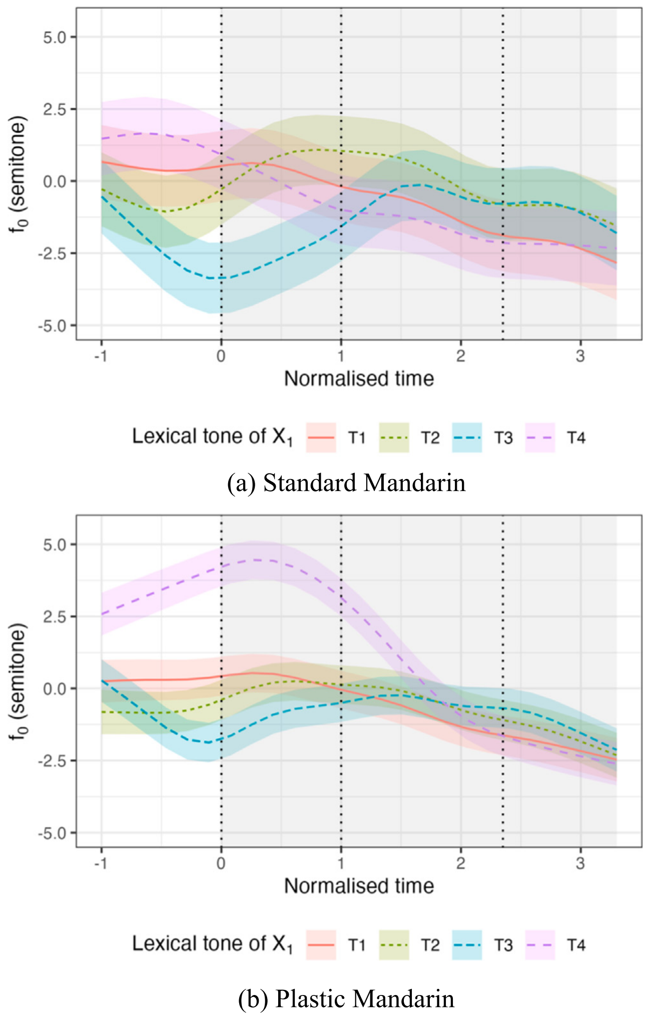


Fig. 14. Fitted f0 contours of four [X_1X_2-*men-de*] phrases by two different groups of Mandarin speakers (duration: 540 ms). The colourful shades indicate 95% pointwise confidence intervals. The vertical dotted lines at 0, 1, and 2.35 indicate syllable boundaries. The normalised time domain [−1, 0] shows f0 contours of the lexical tone syllable X_1, and [0, 1], [1, 2.35] and [2.35, 3.3] for the neutral tone syllables X_2, *men*, and *de* (in grey background).

In Fig. 17(b), the f0 contours of both *mama-men de* and *yeye-men de* phrases (bottom panel) in Plastic Mandarin are, still, not statistically significantly different from the corresponding phrases in Standard Mandarin in their entirety. In the top panel for quadrisyllabic phrases *nainai-men de* and *meimei-men de*, areas marked in red, where the f0 differences between Standard Mandarin and Plastic Mandarin are not statistically significant, have included the whole syllable of *de*. Indeed for all four phrases, the f0 contours of the *de* part, in the normalised time domain [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 such neutral tone syllables in Standard Mandarin and Plastic Mandarin.

4. Discussion

4.1. Approaching a low pitch target

This study provides compelling acoustic evidence that the f0 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 f0 contours converge at the end of neutral tone sequences towards a pitch level that is lower than the speaker mean. Likewise, the converging effect of f0 contours of different groups of phrases is strongly present in trisyllabic phrases and quadrisyllabic phrases in Standard Mandarin. The 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 f0.

Fig. 18 recaps the fitted Gaussian distributions of ending f0 in the trisyllabic and quadrisyllabic tokens in the two Mandarin varieties. At the end of the quadrisyllabic tokens, the two Gaussian distributions of ending f0 have a large overlap. The GAMMs analysis shows that f0 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.

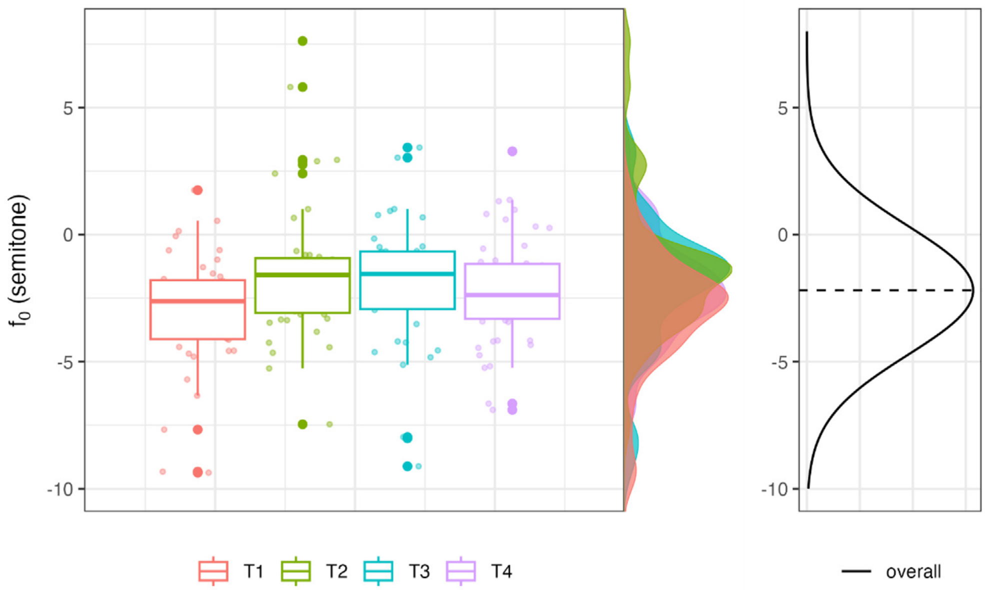
4.2. Interpretations of the low pitch target

Although there is systematic variation in the citation form of lexical tones between Standard Mandarin and Plastic Mandarin (see Fig. 1), the pitch target associated with neutral tones, nevertheless, remains stable and constant in this synchronic variation. Where does the target come from?

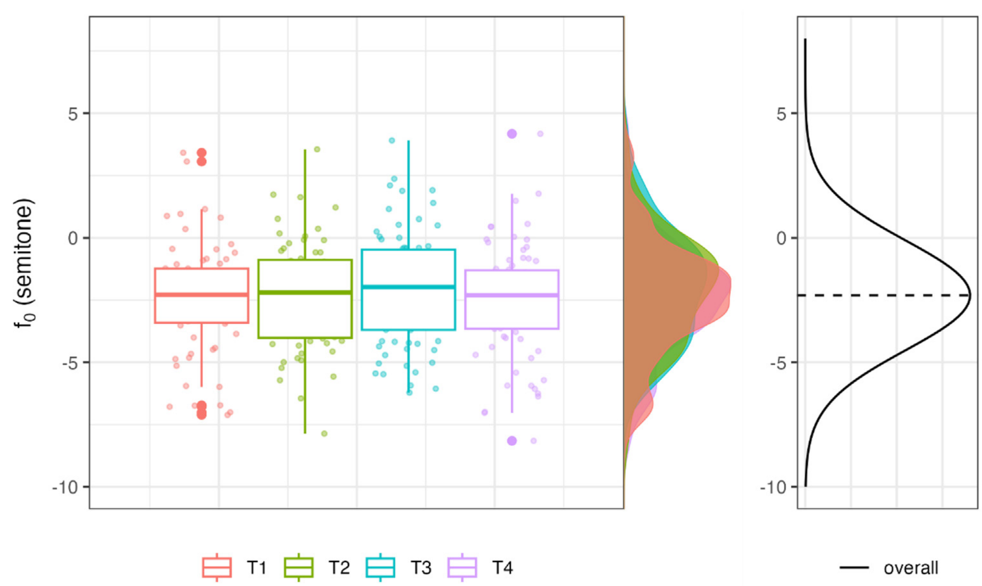
The data show that the f0 contours of neutral tone sequences [X_1X_2-*men*] are greatly influenced by the preceding lexical tone, but hardly influenced by the following lexical tone. The converging effects of neutral tone f0 contours irrespective of the preceding and following tones rule out the claims that neutral tone is merely an interpolation between its preceding and following tone or a complete spreading from its preceding tone.

4.2.1. The underlying low tone hypothesis

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 f0 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 (see Fig. 14). If all neutral tones were L tones, one might 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 dotted black line in Fig. 19(1). It is also highly likely to trigger some tone sandhi when three consecutive L tones were specified, subject to the Obligatory Contour Principle H. Lin (2006) invoked in his underlying L tone proposal. Yet our data displayed a falling trend in all three consecutive neutral tone syllables, despite the variability in the slopes. The alternative boundary tone hypothesis illustrated in Fig. 19(2) seems more likely given our f0 data. Another possible explanation for the continuous fall of f0 could be the weak articulatory strength



(a) Standard Mandarin



(b) Plastic Mandarin

Fig. 15. The distribution, kernel density estimation of ending f0 of *de* in quadrisyllabic tokens and the fitted Gaussian distribution in the two Mandarin varieties. In (a), one potential outlier with f0 less than -10 semitones was removed in the plot.

in the production of neutral tones (Y. Chen & Xu, 2006), as described in section 4.2.3. This perspective of phonetic implementation does not rule out any of these hypotheses.

4.2.2. Attracting the boundary low tone

A more promising account is that a neutral tone does not have any underlying tone, but attracts a low boundary 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-men] phrases are given in (4), in which the focused words are marked in bold. In (4a), *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 regarded 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*.

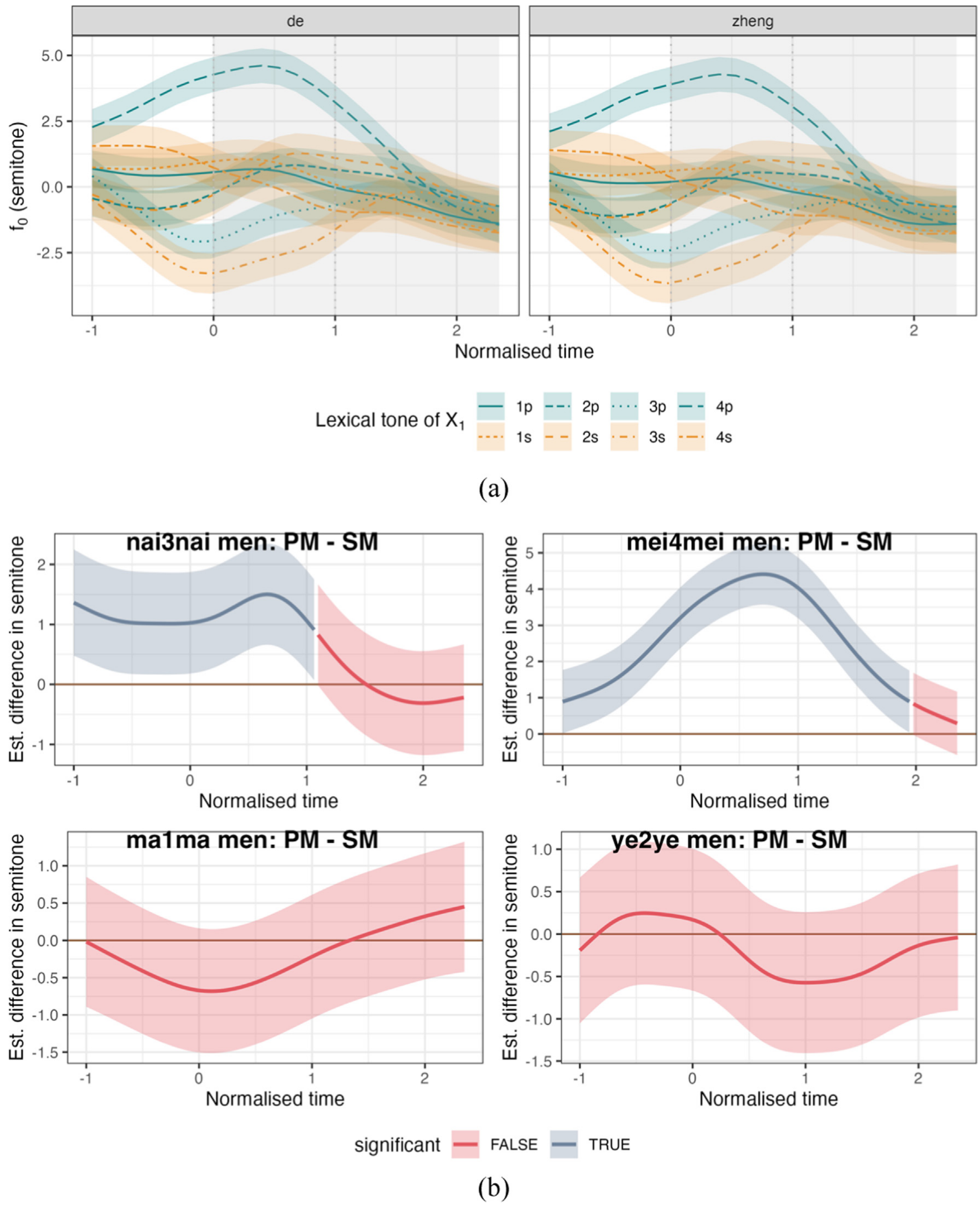
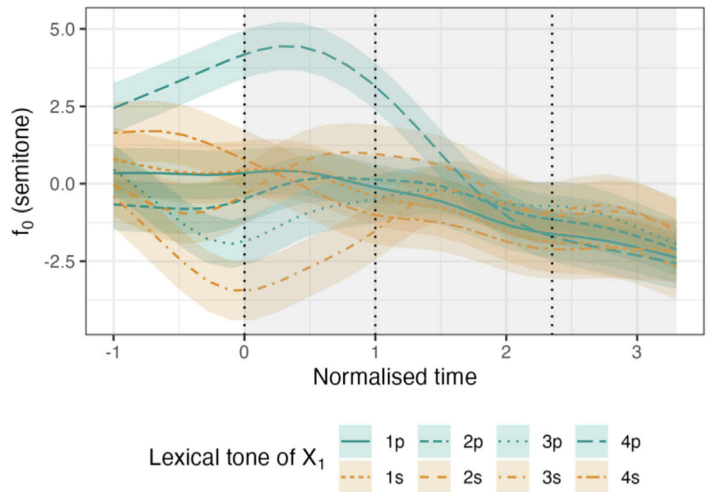


Fig. 16. (a) Fitted f0 trajectories of four groups of trisyllabic phrases in both Mandarin varieties (duration: 460 ms). For the lexical tone of X_1, numbers indicate tone categories, ‘p’ indicates Plastic Mandarin, and ‘s’ indicates Standard Mandarin. (b) The estimated f0 difference trajectories between 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

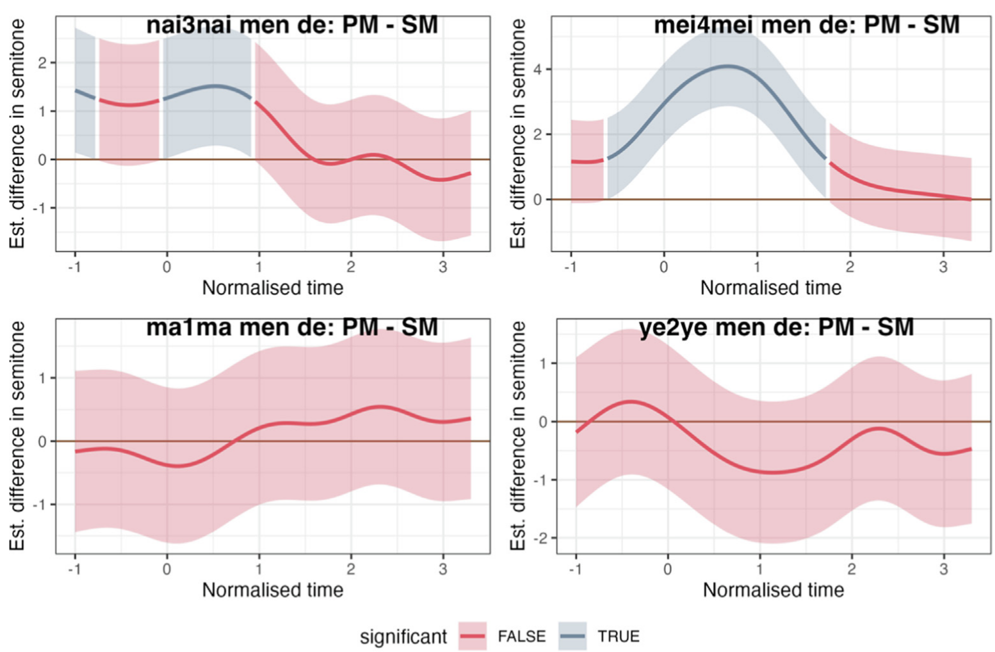
(4) a. mèimei-men zhèng tán-zhe **jítā**.
sister-PL PROG play-DUR guitar
‘Sisters are playing guitars.’

b. mèimei-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 (Nespor, 1986; Selkirk, 1986), in particular the Phonological Phrase and the Intonational Phrase, in Bengali. Takahashi (2019)



(a)



(b)

Fig. 17. (a) Fitted f0 trajectories of four 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 f0 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 (5) 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 (5a) the noun phrase *mèimei-men* precedes a predicate, in (5b) *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 Fig. 19(2).

(5)	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. Acquiring a boundary tone is a post-lexical

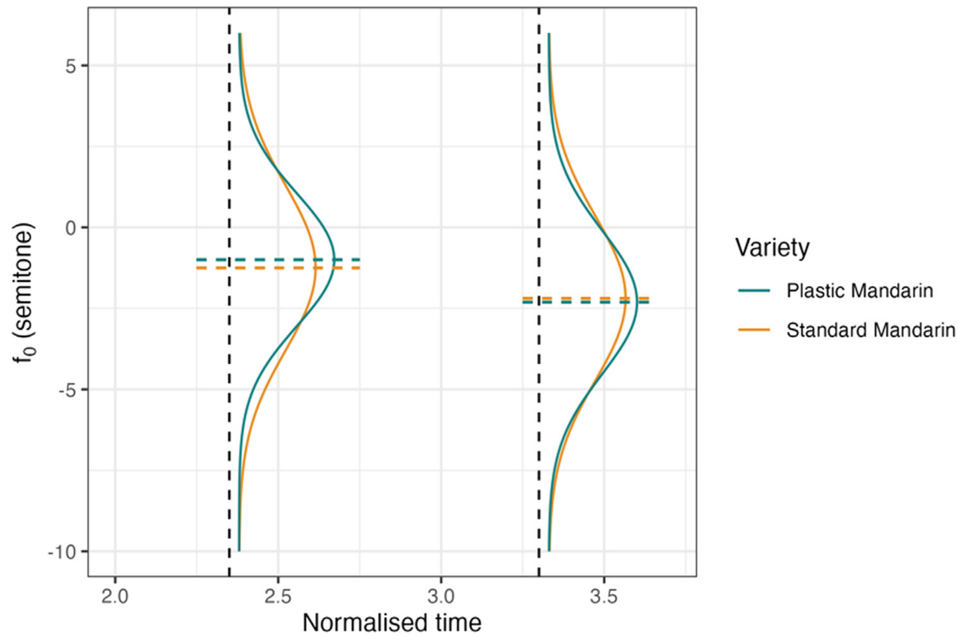


Fig. 18. The fitted Gaussian distribution of ending f0 of trisyllabic (normalised time = 2.35, marked by vertical dotted line) and quadrisyllabic (normalised time = 3.3, marked by vertical dotted line) tokens in the two Mandarin varieties. The horizonal dotted lines indicate the means of the corresponding Gaussian distributions.

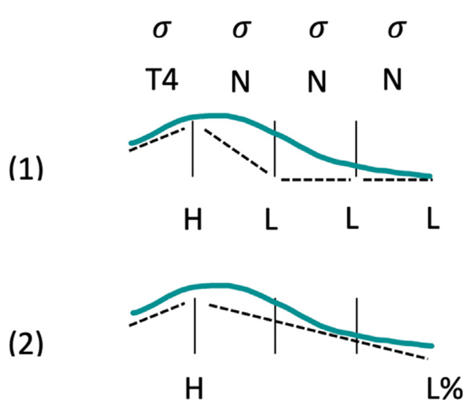


Fig. 19. The schematic illustration of hypothetical pitch targets for a neutral tone sequence, high rising Tone 4 (T4) in Plastic Mandarin followed by three neutral tone syllables (N). Vertical bars demarcate syllables. Letters H and L indicate the underlying tone target of a syllable. L% represents 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 shape of prototypical pitch contour for a [T4-N-N-N] sequence. (1) Neutral tone has an underlying L target (if no obligatory contour or any tone sandhi); (2) neutral tone is underspecified and attracts the L target at the end of the sequence.

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. If neutral tone is inherently the same as the other lexical tones, how to explain its deviation (i.e. its tone target does not vary) from this 'plastic' tonal variation illustrated in Fig. 1? Previous literature has highlighted substantial divergences between neutral tones in Changsha and Standard Mandarin. 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 tones 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 the interrogative particles at the end of a question are most likely to have a higher tone H% instead of a low tone in Mandarin. A. Li & Li (2022) demonstrated that the neutral tone at the end of an intonation phrase is realised as L(ow) in declarative intonation but M(id) in question intonation in Standard Mandarin.

The boundary account is, nevertheless, not conclusive. Firstly, the phonological account alone 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 f0 of a single neutral tone in disyllabic phrases was also subject to higher variability than that of a neutral tone sequence. Secondly, even if a neutral tone sequence has an extended duration, 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 various neutral tones, controlling for pragmatic contexts such as focus and newness. Fourthly, this study only concerns the potential boundary effects on neutral tone syllables. The effects on full lexical tones at prosodic boundaries remain to be explored in future research.

4.2.3. Pitch-related prosodic preplanning

The slow rate of approximating the low pitch target in neutral tones was interpreted as weak articulatory strength (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 (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 f0 heights and duration at edges of prosodic units. Yuan and Liberman (2014) showed strong correlation between the f0 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 f0 slopes. No correlation between the length of the inter-pause unit and the final f0 height was found. Liberman and Pierrehumbert (1984) suggested that utterance-final f0 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 f0 heights in addition to less steep f0 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 f0. 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 f0 height may be more dependent on the duration of a prosodic unit (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 f0 lowering is similar between the two Mandarin varieties.

4.3. Cross-dialectal insights on tonal variation

Apart from Standard and Plastic Mandarin, Li and Chen (2019) and Huang (2018) also observed a mid-low target for neutral tone sequences in Tianjin Mandarin and Taiwan Mandarin respectively. The convergence of the f0 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 indicates 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 groupings in the phrases containing neutral tone syllables are the same between the two Mandarin varieties.

4.4. Implications for neutral tone typology

The f0 patterns of the duplicate syllables X_2 tend to differ from the low-pitch target approaching patterns identified for *de* and *men*. This might be the case that reduplication has a different intonation mechanism and that a copied syllable X_2, together with the X_1, forms a prosodic domain of realising the underlying 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*Hφ), 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 the 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, in light of the neutral tone typology proposed in Shen (1992) and Y. Zhang (2022). A next step of this research is to examine the pitch contours of a wide variety of neutral tone syllables.

5. Conclusions

This study examines speech sequences with neutral tone syllables in two Mandarin varieties, Standard and Plastic Mandarin. There is no canonical f0 contour for a neutral tone syllable in either Mandarin variety. The f0 contour shape varies greatly when a neutral tone syllable is surrounded by different tones, and the preceding lexical tone is a main source of variability. This study suggests that a neutral tone or a sequence of consecutive neutral tones tends to approach a low pitch target associated with a prosodic boundary in both Mandarin varieties and that neutral tone syllables such as the possessive particle *de* and the plural particle *men* may be underlyingly underspecified.

Despite the greater surface variability than other lexical tones, neutral tones maintain some constant or pertinacious aspects of prosody in this Standard–Plastic variation. Unlike canonical lexical tones that differ systematically in contour across the two Mandarin varieties, neutral tones tend to be prosodically conditioned in a similar fashion. It suggests that patterns of tonal variation are constrained by prosodic structure.

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. 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 Mandarin change.

CRediT authorship contribution statement

Chenzi Xu: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Acknowledgements

This work was supported by the Birla Graduate Studentship, Faculty of Linguistics, Philology, and Phonetics, University of Oxford. My special thanks go to Professors John Coleman, Aditi Lahiri, Elinor Payne, Brechtje Post, and Paul Foulkes, Dr Amelia Gully, and Dr Vincent Hughes for their valuable suggestions. I would also like to thank the editor and the anonymous reviewers for their comments and suggestions. Any errors are mine.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2024.101341.

References

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.

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. https://doi.org/10.1017/S0025100304001690.

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. https://doi.org/10.21437/SpeechProsody.2016-234.

Boersma, P (1993). *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound* (14).

Browman, C. P., & Goldstein, L. (1992). "Targetless" schwa: An articulatory analysis. In *Gesture, Segment, Prosody* (pp. 26–67). Cambridge University Press. https://doi.org/10.1017/CBO9780511519918.003.

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.

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), Article 1. https://doi.org/10.5334/labphon.214.

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*. https://doi.org/10.1016/j.wocn.2019.100912 100912.

Chen, M. Y. (2000). In *Tone Sandhi: Patterns across Chinese dialects*. https://doi.org/10.1017/CBO9780511486364.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics, 36*(4), 724–746. https://doi.org/10.1016/j.wocn.2008.06.003.

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. https://doi.org/10.1159/000091406.

Chuang, Y.-Y., Fon, J., Papakyritsis, I., & Baayen, H. (2021). Analyzing phonetic data with Generalized Additive Mixed Models. In *Manual of Clinical Phonetics*. Routledge.

Coretta, S. (2023). *tidymv: Tidy Model Visualisation for Generalised Additive Models* (3.4.2) [Computer software]. https://stefanocoretta.github.io/tidymv/index.html.

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. https://doi.org/10.1017/S0025100306002544.

Dreher, J. J., & Lee, P. (1968). Instrumental Investigation of Single and Paired Mandarin Tonemes. *Monumenta Serica, 27*, 343–373.

Drubig, H. B. (2003). Toward a typology of focus and focus constructions. *Linguistics*, 1–50.

Duanmu, S. (2000). *The phonology of Standard Chinese*. Oxford University Press.

Fitzpatrick-Cole, J. (1996). Reduplication meets the phonological phrase in Bengali. *Linguistic Review, 13*(3–4), 305–356. https://doi.org/10.1515/tlir.1996.13.3-4.305.

Flemming, E. (2009). The Phonetics of Schwa Vowels. In D. Minkova (Ed.), *Phonological weakness in English: From Old to present-day English* (pp. 78–95). Palgrave Macmillan.

Flemming, E., & Johnson, S. (2007). Rosa's roses: Reduced vowels in American English. *Journal of the International Phonetic Association, 37*(1), 83–96. https://doi.org/10.1017/S0025100306002817.

Hayes, B., & Lahiri, A. (1991). *Bengali intonational phonology.* 50.

Hou, X., Zhao, L., & Chodroff, E. (2023). Intermingling tone systems: The relationship of Nanning Mandarin to Nanning Cantonese and Standard Mandarin. In *The Proceedings of the 20th International Congress of Phonetic Sciences*, 1935–1939.

Hu, Y., & Xu, B. (Eds.). (1995). *Xiandai hanyu (modern Chinese)* (Chong ding ben, di 6 ban). Shang hai jiao yu chu ban she.

Huang, J., & Li, A. (2023). Qing sheng yu fei qing sheng zhi jian qing zhong de lian xu tong guan xi. *Chinese Teaching in the World, 3*, 369–387.

Huang, K. (2018). Phonological identity of the neutral-tone syllables in Taiwan Mandarin: An acoustic study. *Acta Linguistica Asiatica, 8*(2), Article 2. https://doi.org/10.4312/ala.8.2.9-50.

Institute of Linguistics CASS. (2016). *Xiandai hanyu cidian (Modern Chinese dictionary)* (7th ed.). The Commercial Press. https://www.cypressbooks.com/proddetail.php?prod=9787100124508.

Jadoul, Y., Thompson, B., & de Boer, B. (2018). Introducing Parselmouth: A Python interface to Praat. *Journal of Phonetics, 71*, 1–15. https://doi.org/10.1016/j.wocn.2018.07.001.

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. https://doi.org/10.16499/j.cnki.1003-5397.2010.04.005.

Keating, P., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 112–156.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society, 1*(1), 97–120. https://doi.org/10.1017/S0047404500006576.

Lahiri, A., & Plank, F. (2022). Phonological phrasing: Approaches to grouping at lower levels of the prosodic hierarchy. In I. B. E. Dresher & H. van der Hulst (Eds.), *The Oxford History of Phonology* (pp. 134–162). University Press. https://doi.org/10.1093/oso/9780198796800.001.0001.

Lee, W.-S., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association, 33*(1), 109–112.

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.

Li, A., & Li, Z. (2022). Prosodic realization of tonal target and F0 peak alignment in Mandarin neutral tone. *Language and Linguistics.* 語言暨語言學*, 23*(1), 47–81. https://doi.org/10.1075/lali.00099.li.

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. https://doi.org/10.1007/s10831-019-09194-4.

Li, Q., Chen, Y., & Xiong, Z. (2017). Tianjin Mandarin. *Journal of the International Phonetic Association, 1–20*. https://doi.org/10.1017/S0025100317000287.

Li, Z. (2003). *The phonetics and phonology of tone mapping in a constraint-based approach* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/17651.

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.

Lin, H. (2006). Mandarin neutral tone as a phonologically low tone. *Journal of Chinese Language and Computing, 16*(2), 121–134.

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. In *Proceedings of the 22nd North American Conference on Chinese Linguistics*, 17.

Liu, L. (2002). 20 shiji hanyu qingsheng yanjiu zongshu 20世纪汉语轻声研究综述 ([An overview of research on the Chinese neutral tone in the 20th century). *Yu wen yan jiu* 语文研究*, 3*, 43–47. https://doi.org/10.3969/j.issn.1000-2979.2002.03.011.

Nespor, M. (1986). Prosodic phonology. *Foris* https://ezproxy-prd.bodleian.ox.ac.uk/login?url=https://ebookcentral.proquest.com/lib/oxford/detail.action?docID=3049945.

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

Paul, W. (2015). New perspectives on Chinese syntax. In *New Perspectives on Chinese Syntax*. De Gruyter Mouton. https://doi.org/10.1515/9783110338775.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Massachusetts Institute of Technology.

Qian, X. (1985). Lun Putonghua de qingshengci he erhuaci 论普通话的轻声词和儿化词. *Shenzhen da xue xue bao. She hui ke xue ban, 3*, 74–82.
Sangster, C. M. (2002). *Inter- and intra-speaker variation in Liverpool English: A sociophonetic study.* University of Oxford.
Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology, 3*, 371–405. https://doi.org/10.1017/S0952675700000695.
Shen, X. S. (1992). Mandarin neutral tone revisited. *Acta Linguistica Hafniensia, 24*(1), 131–152. https://doi.org/10.1080/03740463.1992.10412273.
Shih, C. (1987). The phonetics of the Chinese tonal system. *Bell Laboratories Technical memorandum.*.
Simpson, A. (2001). Definiteness agreement and the Chinese DP. *Language and Linguistics, 2*(1), 125–156.
Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction* [Monograph]. York. https://eprints.whiterose.ac.uk/113858/.
Sóskuthy, M. (2021). Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics, 84*. https://doi.org/10.1016/j.wocn.2020.101017 101017.
Sóskuthy, M., Hay, J., & Brand, J. (2019). Horizontal Diphthong Shift in New Zealand English. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)*. International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia. https://www.semanticscholar.org/paper/HORIZONTAL-DIPHTHONG-SHIFT-IN-NEW-ZEALAND-ENGLISH-S%C3%B3skuthy-Hay/cd1bd700686b3d1270be5536e5881e8946ba57ab.
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.
Sun, Y., & Shih, C. (2021). Boundary-conditioned anticipatory tonal coarticulation in Standard Mandarin. *Journal of Phonetics, 84*. https://doi.org/10.1016/j.wocn.2020.101018 101018.
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. https://doi.org/10.1075/lali.00028.tak.
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. https://doi.org/10.1121/1.5008372.
van Rij, J., Wieling, M., Baayen, R. H., & Rijn, H. van. (2022). *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* (2.4.1) [Computer software]. https://CRAN.R-project.org/package=itsadug.
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.
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.
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. https://doi.org/10.1016/j.wocn.2016.09.004.
Wood, S. (2022). *mgcv: Mixed GAM computation vehicle with automatic smoothness estimation* (1.8-40) [Computer software]. 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. https://doi.org/10.1111/1467-9868.00374.
Wu, J. (2015). *Tonal Bilingualism: The case of two closely related Chinese dialects*. Universiteit Leiden.
Wu, N. (2023). Changsha Xiang Chinese. *Journal of the International Phonetic Association, 1–15*. https://doi.org/10.1017/S0025100323000075.
Wu, Y. (2005). A synchronic and diachronic study of the grammar of the Chinese Xiang dialects. *De Gruyter Mouton*. https://doi.org/10.1515/9783110927481.
Xiao, R. (2009). *A frequency dictionary of Mandarin Chinese: Core vocabulary for learners*. Routledge http://www.loc.gov/catdir/toc/ecip0823/2008030172.html.
Xu, C. (2022). *Investigating the tonal system of Plastic Mandarin: A cross-varietal comparison* [http://purl.org/dc/dcmitype/Text, University of Oxford]. https://ora.ox.ac.uk/objects/uuid:df4f01ec-a867-4d96-89f8-6a17f1de0d82.
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. https://read01.com/zQRjEP.html.
Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics, 27*, 51–105.
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.
Yip, M. (1980). *The tonal phonology of Chinese* [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/15971.
Yip, M. (2002). *Tone*. Cambridge University Press.
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. https://doi.org/10.1121/1.2935783.
Yuan, J., & Liberman, M. (2014). F0 declination in English and Mandarin Broadcast News Speech. *Speech Communication, 65*, 67–74. https://doi.org/10.1016/j.specom.2014.06.001.
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.
Zhang, J. (2018). A comparison of tone normalization methods for language variation research. In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*. PACLIC 2018, Hong Kong. https://www.aclweb.org/anthology/Y18-1095.
Zhang, Q. (2005). A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society, 34*(3), 431–466. https://doi.org/10.1017/S0047404505050153.
Zhang, Y. (2022). *Neutral Tone in Mandarin: Representation and Interaction with Utterance-level Prosody* [Thesis, University of Cambridge]. https://doi.org/10.17863/CAM.79717.
Zhong, Q. (2003). Changsha hua de qingsheng 长沙话的轻声 (Light tone in Changsha dialect). *Fangyan 方言, 3*, 255–264.
Zue, V., Seneff, S., & Glass, J. (1990). Speech database development at MIT: Timit and beyond. *Speech Communication, 9*(4), 351–356. https://doi.org/10.1016/0167-6393(90)90010-7.