Open Library of Humanities

Variable f0 of toneless moras in Suzhou Chinese: a computational analysis

Yuhong Zhu, Hiroshima University, Japan, yuhongzhu@hiroshima-u.ac.jp

Toneless Tone Bearing Units (TBUs) as an analytical device have been widely used in tone and intonation studies, but there is a lack of consensus on how toneless TBUs realize in f0. This study investigates the variable f0 realization of phonologically toneless moras in Suzhou Chinese (Northern Wu) using Naive Bayes classification models. By embedding the toneless moras in different contexts (e.g., between two H tones, preceded by H and followed by L), I have found both intra- and inter-speaker variation across the tonal contexts. Speakers of Suzhou Chinese realized toneless moras in three main ways: insertion of a 'Default L' tone, linear pitch interpolation between the left and right tonal contexts, and spreading of the left tonal context. In addition, I found insufficient evidence for the proposal that toneless TBUs as weak elements in speech realize as a stable target in the mid pitch range (Y. Chen & Xu 2006). The categorical pitch variation of toneless moras is indicative of a phonological model of variable/optional processes, where a single phonological representation (here, the *absence* of tone) can be mapped to distinctive surface forms (Coetzee & Pater 2011; Coetzee & Kawahara 2013).

1 Introduction

Syllables and moras devoid of lexical and/or phonological tones, also known as 'Neutral Tones', have been widely discussed in previous literature on both Mandarin and regional dialects of Chinese (M. Chen 2000; Yip 2002; H. Zhang 2016; Y. Zhang 2021, among others). The source of tonelessness may range from underlying representation to tonal processes such as neutralization or tone sandhi. In addition, toneless prosodic units are demonstrated to realize in various phonetic forms across dialects, and sometimes even within the same variety. The current study investigates the realization of phrase-medial toneless moras in Suzhou Chinese (Northern Wu), which result from being in metrically unparsed and tonally weak positions. Previous research on Suzhou has argued for several types of tonelessness: Ye (1993) and Wang (2011) for toneless functional words, Ling (2011; 2014) for 'neutral tones' in polysyllabic phrases, Zhu (2023a) for toneless moras in light-heavy disyllables. However, it remains unclear whether the observed low/mid pitch of these toneless syllables/moras results from a language-specific 'default' tonal target (Yip 1980; Y. Chen & Xu 2006; H. Zhang 2016), or interaction with intonation towards the end of a prosodic phrase (Takahashi 2019; Roberts 2020).

The primary objective of this study is to tease apart the mechanism of toneless realization in Suzhou, while controlling for the influence of intonation. I focus on phrase-medial toneless moras that are surrounded by varying phonological tones, and analyze the f0 data adopting the computational methods of Shaw & Kawahara (2018); Kawahara et al. (2022): instead of relying on the subjective judgement of fieldworkers when determining whether a certain f0 trajectory is 'level' or 'falling', I fit the toneless mora data to Naive Bayes classification models trained on (real and simulated) tokens representing competing hypotheses (e.g., 'Default L Insertion', 'Linear Interpolation'; see (6)) of toneless realization. The methodology is particularly relevant for the current investigation, as (i). it allows for robust and relatively impartial classification of gradient f0 data into discrete categories; (ii). it analyzes *individual* tokens and captures (potentially categorical) variation in instances of phonetic realization.

Contrary to predictions made by a gradient model of 'target approximation' (the PENTA model; Xu 2005; Y. Chen & Xu 2006; Prom-on et al. 2009; Xu et al. 2022), I demonstrate that toneless moras in Suzhou Chinese have categorical targets that qualify for distinct phonological processes. Furthermore, there is no uniform realization strategy either within speakers or within tonal contexts — across sixteen speakers and three tonal contexts, I have identified three distinctive realizations independently attested in previous tonelessness literature: (i). an inserted low pitch target, also known as 'Default L' (M. Chen 2000; Yip 2002; H. Zhang 2016); (ii). a context-dependent, interpolated pitch (Pierrehumbert 1980; Pierrehumbert & Beckman 1988; Myers 1998; Gussenhoven 2004; Lee & Zee 2008); (iii). a pitch value resulting from spreading of the tautosyllabic toned mora (Pierrehumbert 1980; Yip 1980; Wang 1997). The non-uniform,

non-deterministic realization of toneless moras in Suzhou is indicative of optional/probabilistic phonological processes (Coetzee & Pater 2011; Coetzee & Kawahara 2013), and is comparable to the neutral tone data of Mandarin in M. Zhang et al. (2019) and those of Taiwanese Southern Min in Liu et al. (2021).

The contribution of the current study is two-fold. First, it addresses the ongoing debate of toneless/neutral tone realization in Chinese languages (Y. Chen & Xu 2006; Y. Zhang 2021 for overview) by presenting original fieldwork data of Suzhou Chinese. Here, I focus on toneless moras arising from tone sandhi in phrase-medial position, an understudied tonal aspect of Suzhou. Previous literature has only approached tonelessness in functional words, or phrase-final syllables/moras. Moreover, I adopt the simulation and classification methods by Shaw & Kawahara (2018), M. Zhang et al. (2019) and Kawahara et al. (2022) in assessing the presence/absence of (phonological) tonal targets for the toneless moras. The variable realization in f0 brings further validating evidence for adopting computational tools in evaluating the *lack* of phonological tone/gesture/feature, where an unbiased, robust analysis is often untenable due to the quantity and nature of the data. I demonstrate that the stochastic simulation and classification analyses proposed by Shaw, Kawahara and colleagues provide unique insights into 'noisy' phonetic data, especially when the variability comes from two or more distinctive phonological entities. This is where traditional statistical analyses (e.g. ANOVAs, linear regressions) fall short, as most would rely on grand mean values to some extent (Kawahara et al. 2022; Zhu 2023c for discussion).

The paper is structured as follows. The background section in §2 reviews three topics relevant to the current investigation: tonelessness and its acoustic correlates, distribution of toneless moras in Suzhou, and theoretical proposals assessing 'targetlessness' from noisy phonetic data. §3 discusses the methodology for both the elicitation data and subsequent computational analyses, which is followed by the results in §4. §5 visits several remaining issues and §6 concludes the paper.

2 Background
2.1 Phonetic realization of tonelessness

Tone Bearing Units (TBUs) that are unspecified for (or unassociated with) phonological tones, i.e. toneless TBUs, are hardly a novel concept to analysts working with tone and intonation. Since the inception of Autosegmental Phonology (Leben 1973; Goldsmith 1976), tonelessness has been adopted as a common analytical device to account for tonal behaviors that are otherwise atypical of full lexical tones — transparency to spreading, non-trigger of tonal processes, highly context-dependent pitch, to name a few (Yip 1980 for overview). Despite its extensive coverage in the tone and intonation literature, there seems to be little agreement on the acoustic or articulatory realization of syllables/moras that are underlyingly toneless. While some authors hold the view that context dependency is at the core of tonelessness (and more broadly, *targetlessness*; see Browman & Goldstein 1992), others argue that toneless TBUs assume a (language-specific)

'default' pitch value or articulatory target. This subsection reviews some of the representative work in the tonelessness literature, with a focus on Chinese languages.

Phonetic interpretation of toneless TBUs was developed in parallel among studies on stress and pitch accent languages on the one hand, and studies on lexical tone languages on the other.[1] The seminal work of Pierrehumbert (1980) on English intonation and Pierrehumbert & Beckman (1988) on Tokyo Japanese pitch accent both acknowledge that sparsely-toned languages have phonetic implementation rules, through which syllables and moras not specified for tones acquire their surface pitch. Pierrehumbert and colleagues have proposed two implementation rules: (linear/'sagging') interpolation, where toneless units acquire pitch not through specified tonal targets, but through motor-system-based interpolation; spreading, where toneless TBUs are assigned a specific target from adjacent (often preceding) tones (see also Gussenhoven 2004; Y. Chen & Xu 2006).

Lexical tone languages, which are by definition densely toned, have two main sources of tonelessness: tone deletion/neutralization and morphemes underlyingly unspecified for tone. In the context of Chinese languages, the *neutral tone* (輕聲[2]) in Standard Mandarin/Putonghua is the most well known and well studied toneless phenomenon. Many researchers have made the observation that neutral-toned syllables in Mandarin are largely dependent on the preceding syllables in their pitch value (Chao 1968; Yip 1980; Y. Chen & Xu 2006; Duanmu 2007; Lee & Zee 2008). Nevertheless, the authors do not necessarily agree on the precise pitch values (and sometimes even on the shapes) of neutral tones under different contexts (see Y. Zhang 2021 for summary). Nor have Chinese linguists reached consensus on the symbolic representation (or lack thereof) for neutral tones. Of note are two experimental studies: Y. Chen & Xu (2006) argue that neutral tones in Mandarin are not subject to phonetic implementation rules such as interpolation or spreading, and have a 'weakly-articulated' tonal target resembling a Mid tone. They state that 'neutral-tone syllables do have a target that is independent of the surrounding tones' (Y. Chen & Xu 2006: 47), implying that neutral tones are also assigned a static value. Lee & Zee (2008), while not being explicit regarding the phonological makeup of neutral tones, do focus more on the f0 variability of neutral-toned syllables when placed in different preceding and following tonal contexts.

While several studies have explored tonelessness in regional Chinese dialects, there is likewise disagreement on the exact realization mechanism. Tonelessness is frequently invoked as an analytical device to account for the complex tone sandhi patterns of these dialects, with its (phonetic/phonological) value often described as a 'default L' or 'default M'. Multiple sources

---

[1] Here, I use the terms 'stress and pitch accent languages' and 'lexical tone languages' only for ease of reference. I remain neutral on whether these languages are distinctive types or ones on a more gradient scale (cf. Hyman 2009).

[2] Widely used in literature on both Standard Mandarin and regional varieties of Chinese, this term conflates several tonal properties that do not necessarily co-occur: (i). neutralization of lexical tone contrast; (ii). production with lower intensity, shorter duration and compressed pitch range (similar to being 'unstressed'); (iii). lack of tonal specification. The current paper focuses on the third property, i.e., TBUs that are not specified with any tonal feature/gesture. See §2.2 for some disambiguation in Suzhou Chinese.

have cited Shanghai Chinese (Shanghai Wu/Shanghainese) as a representative example of 'default L' toneless syllables in weak positions, where non-initial syllables within a prosodic word surface with a low pitch if not associated with any tone after tone sandhi (Zee & Maddieson 1979; M. Chen 2000: 219; Yip 2002: 23; Y. Chen 2008: 256; H. Zhang 2016: 90). Matthew Chen in his overview also provides further dialectal examples for phonologically toneless TBUs: Zhenhai (M. Chen 2000: 68–69; Rose 1990), Wenzhou (M. Chen 2000: 77), New Chongming (M. Chen 2000: 174), among others. Interestingly, without further elaboration, all regional dialects mentioned in M. Chen's overview seem to have a designated tonal value for their toneless syllables — there seems to be an assumption that syllables in Chinese dialects, regardless of their lexical (under-)specification, *must* bear some phonetic/default tone on the surface (cf. Well-Formedness Condition, Goldsmith 1976; also see SPECIFY in Yip 2002).

A recent study by Takahashi (2019) explicitly calls into question this assumption for Shanghai Chinese: Takahashi explores the possibility where toneless syllables in Shanghai are not specified with any default value, but receive low pitch through interaction with a boundary L (similar to Pierrehumbert & Beckman 1988). Crucially, the author provides instrumental data showing that the pitch value of the same syllable (e.g. a L-toned second syllable) is lower in shorter (disyllabic) words than in longer (trisyllabic, quadrisyllabic) ones. The results are compatible with a boundary L account, where the boundary tone exerts more influence over the first two (fully-toned) syllables when it is closer to them in disyllables (see also Pierrehumbert & Beckman 1988 for similar length-dependent data). A default L account does not explain the lower pitch in disyllables, as the second syllable can be assigned a H tone with no default L to begin with.[3]

The current study investigates the phonetic realization of toneless *moras* in Suzhou Chinese along the same lines. Bearing in mind the recent debate on surface specification of toneless TBUs (or lack thereof), I aim to shed light on two aspects regarding tonelessness in Suzhou through my data: (i). whether a toneless mora has to be tonally specified; (ii). if so, what tonal value (default L, default M, etc.) it may take (see also §2.3 for research questions).

2.2 Toneless moras in Suzhou Chinese

Suzhou Chinese (蘇州話) is spoken at Suzhou City in Jiangsu Province, China. Being a geographical and dialectal neighbour to Shanghai, Suzhou shares many linguistic features with Shanghai Chinese, including a three-way voiced vs. voiceless unaspirated vs. voiceless aspirated laryngeal contrast, similar left-dominant tone sandhi patterns and many common vocabularies.

The term 'tonelessness' (輕聲) has been mentioned in several recent documentations of Suzhou. Wang (2011: 82) notes that non-initial characters of prosodic words are 'similar to neutral tone in Beijing Mandarin, as they can be characters of any lexical tone'.[4] Here, Wang is emphasizing

---

[3] Due to tone sandhi, all disyllables in Shanghai are fully toned.
[4] Author's translation. See also Xie (1982: 245).

the fact that the lexical tone contrast of non-initial characters is *neutralized* due to tone sandhi in Suzhou (fn. 2). Nevertheless, these non-initial syllables are always assigned full tones in the fieldwork transcription as a result of tone sandhi, similar to the Shanghai case. For instance, a '2.4' transcription corresponds to a [L.H] disyllable, where the original lexical tone of the second syllable is neutralized to a high tone.[5] Rather than being TBUs without phonological tones, this reference to 'neutral tone' describes a neutralizing phonological process, and is beyond the scope of this paper.

Another source of 'neutral tone' in Suzhou comes from frequent discussions of functional words in the language: Ye (1993: 7) transcribes all common functional words (e.g., prepositions, sentence particles, affixes) as '21', and comments that they are 'indistinguishable from the 21 in tone sandhi'. Wang (2011: 91), on the other hand, groups functional words and words in fast speech together and argues that 'there is basically no tonal contour; it can be transcribed as 3 no matter what the original tone is'.[6] These accounts show that researchers do not necessarily agree on the phonetic value of tonelessness in Suzhou, which will become relevant in the following section. Apart from this, functional words in Mandarin have also been shown to have different phonetic realizations from toneless TBUs arising from phonological processes such as deletion and tone sandhi (M. Zhang et al. 2019 for a comparison of Mandarin grammatically toneless morphemes vs. toneless syllables after deletion). The current study focuses on how tonelessness in lexical words realizes in pitch, an aspect not fully explored by previous literature on Suzhou.

In Suzhou, prosodic words with initial 'checked' or light syllables demonstrate several special phonological properties, including exceptional disyllabic tone sandhi (Zhu 2023a) and toneless moras (Ø hereafter).[7] An example tonal minimal pair is given in (1) below:

(1)   Tonal minimal pair of a heavy-initial and a light-initial disyllable. See Zhu (2023a) for more examples with pitch data. Moras following tones indicate their association status. In addition, coda glottal stops in light syllables are often deleted in running speech ([ba^L] in 'white flower'), but preserved in phrase-final position.
   a.   /LH/μμ + /H/μμ = Lμμ.Hμμ ([mɛ:^{L}.ho:^{H}] 梅花 'plum flower')
   b.   /LH/μ + /H/μμ = Lμ.HμØμ ([ba^L.ho:^{HØ}] 白花 'white flower')

In the examples above, (1a) is a heavy-heavy disyllable and (1b) a light-heavy disyllable. Noticeably, the two examples are identical in tonal input (/LH/ followed by /H/), but take

---

different surface forms only attributable to the weight difference in the initial syllable — heavy-heavy disyllables have all four moras toned, whereas light-heavy disyllables end with a toneless mora Ø. This phenomenon also differs from that of lexically toneless functional words: instead of a completely toneless morpheme/syllable (e.g., [sɛːᴴ.lãː$^{\emptyset}$] 山浪, morphologically 'mountain-PREP', 'on the mountain'), only one mora of the non-initial morpheme/syllable is unspecified for tone: [hoːᴴ�放], as in (1b). Furthermore, tonelessness only takes place in light-heavy disyllables, as heavy-light and light-light ones are also fully toned. This can be generalized as:

(2)   Toneless mora distribution in Suzhou Chinese disyllables. T: any tone
   a.   Light-heavy: Tμ.TμØμ
   b.   Elsewhere: T.T

Zhu (2023a) accounts for this weight-sensitive tonal pattern by proposing two foot structures for disyllables in Suzhou. Footing in Suzhou is always trochaic (left-dominant), but varies in size under different weight configurations. When a disyllable is light-heavy, a left-aligned, bimoraic trochee built directly on moras (based on Kager 1993; Kager & Martínez-Paricio 2018; Breteler & Kager 2022) parses the trimoraic sequence *non-exhaustively*: linearly, (μ+.μ–)μ, where brackets stand for foot boundaries and plus/minus stand for foot head/dependent respectively. The unparsed final mora is unable to bear a phonological tone (de Lacy 2002; Breteler 2018; Zhu 2023a), leading to its toneless status. All other three disyllabic weight profiles (heavy-heavy, heavy-light, light-light) are fully parsed by a disyllabic trochee and fully toned: (σ+.σ-). In short, TBUs (either directly or indirectly) dominated by a foot must be phonologically toned (cf. SPECIFY), while ones unparsed must remain toneless (*NONFT → T, Zhu 2023c; also see de Lacy 2002; Breteler 2018). Below I show a schematic representation of the two footing scenarios in Suzhou.

(3)   Quantity-sensitive footing in Suzhou Chinese. PrWd: Prosodic Word; Ft: Foot
   a.   A light-heavy disyllable, non-exhaustively parsed by a bimoraic trochee
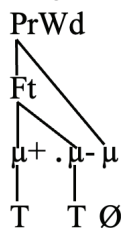
   b.   A heavy-heavy disyllable, fully parsed by a disyllabic trochee
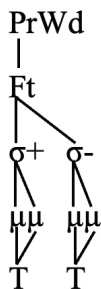
In (3a), the dot between the first and the second moras represents the syllable boundary — syllables are not explicitly drawn here as they are on a separate autosegmental tier, but the prosodic word is nevertheless syllabified. The final/third toneless mora is not parsed by the bimoraic foot, and is directly dominated by the prosodic word. It is unable to license a phonological tone, and is represented by the 'Ø' symbol and the lack of association line. Note that such a footing is anomalous to some extent, as the right foot boundary splits the second heavy syllable in half ('foot straddling'; see Prince 1976; Hayes 1995). In (3b), on the other hand, all moras are parsed under a disyllabic foot, and the disyllable is fully toned.[8] Zhu (2023a) motivates the straddling bimoraic foot from the cross-linguistic observation that '(foot) dependents may not be more complex than heads' (Head-Dependent Asymmetry; Dresher & van der Hulst 1998: 342; Iosad 2013). Simply put, a light-heavy disyllable parsed by a disyllabic trochee leads to the undesirable outcome where the head is lighter (i.e. less complex in weight) than the dependent: $*(\sigma+_\mu.\sigma-_{\mu\mu})$. Suzhou Chinese opts for the bimoraic trochee as a 'repair' strategy, violating Syllable Integrity while obeying the Head-Dependent Asymmetry (see Zhu 2023a for an OT-based analysis).

While being a sufficient phonological account, the phonetic evidence for tonelessness presented in Zhu (2023a) is in fact compatible with two different analyses. In **Figure 1** I show the spectrogram and pitch tracking data of the word [ba^L.ho:$^{H\emptyset}$] 白花 'white flower' elicited in a carrier sentence. The light-heavy disyllable ends in a low phonetic pitch, corresponding to the final toneless mora Ø. However, there are two ways to interpret the low pitch in this particular context:

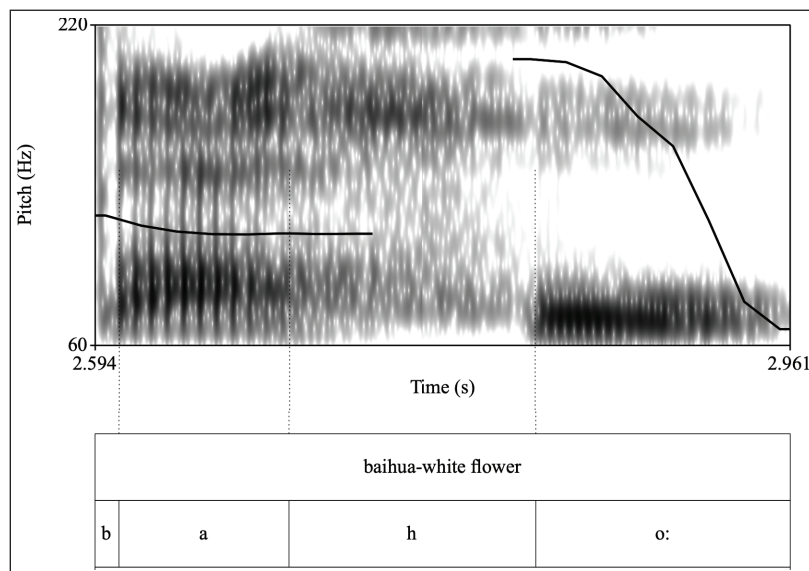**Figure 1:** Spectrogram and pitch data for [ba^L.ho:$^{H\emptyset}$] 白花 'white flower'. Figure from Zhu (2023a).

---

[8] Heavy-light and light-light disyllables will be parsed by the same disyllabic trochee. The representation is omitted here for exposition reasons.

(4)   Interpretations of the low ending pitch in **Figure 1**. See (6) below for a full list of competing hypotheses.
  a.   The final mora is phonologically toneless, but receives a default/phonetic L tone on the surface (Default L Hypothesis)
  b.   The final mora is both phonologically and phonetically toneless, and receives its low pitch from interpolation to a L% boundary tone (assuming that L% does not directly associate with the mora) (Interpolation Hypothesis)

Interpretation (4a) is in line with most previous studies on default L in Chinese dialects: TBUs can be phonologically toneless, but are assigned a (mandatory) low pitch target by phonetic implementation rules (Gussenhoven 2004; Y. Chen & Xu 2006). In contrast, (4b) applies the findings in the autosegmental-intonational tradition in assuming that TBUs can stay unspecified for tones throughout, even in lexical tone languages.[9] The pitch values of 'true' toneless TBUs will then be determined by intonational tones, such as a boundary tone (both Takahashi 2019 and Roberts 2020 have argued for the existence of a boundary L% in Shanghai). A notable example along this vein is given by Roberts (2020), who argues for a complete intonational analysis of Shanghai tone sandhi. According to Roberts, *all* within-word tonal phenomena in Shanghai can be reanalyzed as the interaction between lexical pitch accents and boundary tones, where interpolation accounts for pitch transitions between targets.

   To that end, the disyllables in Zhu (2023a) were elicited in a carrier sentence, but were nevertheless influenced by a phrase boundary L% tone (Zhu 2023c, Chapter 4 for intonational data). It is conceivable that the low pitch may have resulted from interpolation towards the boundary L%, instead of due to a default L. One also cannot fully rule out Tone Spreading as a possible implementation (Pierrehumbert 1980; Pierrehumbert & Beckman 1988), as it does not take place in **Figure 1** possibly due to the presence of an overt L%. The same ambiguity in phonological and phonetic interpretation of tonelessness is also explicitly discussed in Yip (2002: 63). In short, disyllabic prosodic words, widely used in fieldwork of varieties of Chinese,[10] are not well-equipped to test how exactly word/phrase-medial Ø realizes in pitch.

   This paper directly addresses this ambiguity by placing the toneless Ø in question in *word-medial* position of trisyllabic and quadrisyllabic phrases, minimizing the potential effect of intonation.[11] The next subsection reviews recent research on variable phonological process/realization, and contrast it with a 'stable target' model of tonelessness such as that of Y. Chen & Xu (2006). I also

---

[9] As amply discussed in Gussenhoven (2004) and Hyman (2009), being classified as a 'lexical tone language' does not necessarily mean mandatory tone assignment in all TBUs.

[10] Most recent tonal documentations of Suzhou are primarily based on disyllabic words, including Zhu (2023a). As a matter of fact, fieldwork including Xie (1982) and Wang (2011) may have concluded that non-initial syllables of disyllabic words are 'neutral tone' due to the influence of a phrase-final L%.

[11] In comparison, Ling (2011) examines the pitch data of the *end* of polysyllabic prosodic words in Suzhou. See discussion below.

discuss several competing hypotheses regarding the phonetic realization of Ø, summarizing both theoretical generalizations and relevant empirical data in Suzhou and Shanghai.

2.3 Variable phonological realization vs. (stable) target approximation

Research on lexical tone and tonal representation faces two sources of variability from the data: acoustic 'noise' inherent in the signal and linguistically meaningful (e.g. sociolinguistic, phonological) variation. On the other hand, to arrive at a set of phonologically contrastive, symbolic representations, fieldworkers often have to make arbitrary decisions on whether to treat (statistically) different phonetic forms as *categorical* types or *gradient* variation. For instance, Bermúdez-Otero & Trousdale (2012: 694–695) cite a case of external /n#k/ sandhi in Present Day English where both speakers with a gradient reduction pattern and ones with 'bimodal' gestural reduction were found. The task is no easier for tonal languages, as f0, the acoustic basis of phonological tones, is continuous in nature. Here, I discuss two recent approaches on teasing apart variation in the phonetic signal under the tonelessness (neutral tone)/targetlessness context.

One strand of research on phonological theory adopts the view that phonological processes are not necessarily one-to-one in nature, as one phonological form may surface as multiple distinctive phonetic forms that might be sociolinguistically meaningful (Coetzee & Pater 2011; Coetzee & Kawahara 2013). As such, a model of phonological derivation, whether it takes the form of SPE-style rules or Optimality Theoretic constraint interactions, must be capable of mapping a single phonological input to a certain distribution of outputs (see Coetzee & Pater 2011 for examples).

To identify and classify variable surface forms of the same lexical item, earlier sociolinguistic studies mostly make use of the fieldworkers' own judgement (for example, Guy 1991). Manual classification by fieldworkers, however, is undesirable for two reasons. First, the analyst would often have to make a compromise between fewer speakers or fewer items per speaker, as it is impractical to classify a large amount of tokens by hand. In addition, classification by hand brings in the subjectivity of individual fieldworkers and disagreement among them, especially when the (continuous) phonetic data are *ambiguous* between two categories to begin with. The challenge left to a fieldworker, then, is to come up with robust and unbiased criteria to classify the phonetic data.

A recent study by Shaw & Kawahara (2018) presents a computational approach to assess surface specification of potentially 'targetless' vowels in Japanese. The study focuses on the debate on devoiced vowels in Tokyo Japanese, where some scholars believe the vowel is completely deleted during devoicing (i.e. empty V slot) and others argue that devoiced vowels are simply full vowels without vocal fold vibration. Their 'simulation and classification' paradigm consists of two parts. A set of 'targetless' vowel articulations was stochastically sampled, where the relevant articulator (here, Tongue Dorsum) takes a *linear interpolation* trajectory between two full vowel targets (similar to linear interpolation of toneless syllables in Pierrehumbert 1980; Pierrehumbert

& Beckman 1988). The authors refer to this data set as the 'simulated' set. Secondly, a Naive Bayes classifier (per lexical item per speaker) was trained on both the 'simulated' set and a set of full vowel trajectories (where no deletion was expected). The trained Naive Bayes model was then capable of evaluating devoiced vowel trajectories on a *token-by-token* basis, and would give a probability of 'targetless/deleted vowel' vs. 'full but unvoiced vowel' for each devoiced token. Although not explicitly set out to explore the possibility of variable phonological realizations, Shaw and Kawahara have found noticeable variation across both speakers and lexical items — some speakers heavily preferred to keep a full articulatory target with no voicing, while others produced most of their devoiced tokens with linear interpolation (i.e. no vowel target at all). The Naive Bayes classifier (and more generally, most machine learning classifiers) is shown to achieve token-by-token evaluation on a relatively impartial basis.[12]

The simulation and classification paradigm has then been extended to the study of Neutral Tones in Mandarin Chinese (M. Zhang et al. 2019) and pitch accent neutralization in Japanese (Kawahara et al. 2022), the former of which is more pertinent to the current study. As f0/pitch trajectories are similar in nature to the EMA data recorded in Shaw & Kawahara (2018), a Naive Bayes classifier with similar 'simulated' vs. 'full tone' training input should also be able to assess whether a Neutral Tone syllable in Mandarin is (articulatorily) toneless or has some sort of default tonal value. The results of M. Zhang and colleagues have identified two distinct *types* of Neutral Tones in Mandarin: The realization of underlyingly toneless morphemes was overwhelmingly linear interpolation — they were classified as having no tone at all. On the other hand, there was cross-speaker variation among Neutral Tones created by phonological neutralization processes, as some speakers produced tokens indistinguishable from full lexical tones (i.e. 'default tone') while others had varying degrees of reduction in their Neutral Tone tokens. The toneless moras I focus on resemble the latter type, as they come from underlyingly toned morphemes that lose their tones during tone sandhi.

The above studies have all approached phonetic data under the assumption that targetlessness, just as underspecification for features, may contribute to surface variation. In addition, these studies have treated the variation in phonetic forms as being *linguistically meaningful*: the results may point at further directions of sociolinguistic investigation, or at least call for the implementation of one-to-many phonological models. The Parallel Encoding and Target Approximation (PENTA) model developed by Yi Xu and colleagues, on the other hand, holds an opposite view on both aspects. First, it makes an explicit assumption that 'each syllable is assigned an underlying pitch target specified in terms of not only height but also slope' (Xu et al. 2022: 379) — there is no room for underspecifying targets. Secondly, contours or transitions between specified pitch targets are caused by an 'asymptotic approach' towards each target, and

---

[12] Interestingly, as discussed in Shaw & Kawahara (2018: 501), the researchers can include *built-in* biases in a Naive Bayes model depending on the hypotheses they aim to test. The authors chose not to do so by assigning equal prior probabilities (0.5 vs. 0.5) to the two possible classes.

acoustically distinct surface forms are 'mostly a by-product of physical inertia' (Xu et al. 2022: 380), which is not linguistically meaningful.

The strongest evidence Y. Chen and Xu present against the targetless analysis is the finding that the alleged unspecified segments/features often realize in ways 'that cannot be readily explained by interpolation' (Y. Chen & Xu 2006: 48) — regarding the case of neutral tones, neither interpolation nor spreading could adequately account for the pitch data they gathered. Instead, they argue that neutral tones, just as other full tones, *do* have an intended pitch target, which is often not fully reached due to them being a '*weak element*' in speech and having a short duration. The seemingly context-dependent pitch of neutral tone syllables, then, is simply a direct outcome of incomplete target approximation.

Noticeably, two methodological choices may have made the results of Y. Chen & Xu (2006) difficult to interpret. Firstly, when reporting the f0 values of neutral tones under different conditions, the authors refer to the *averaged* f0 trajectory over three repetitions from four speakers. Each test condition varying in tonal context, focus and/or speaking rate was therefore a mean of twelve instances by different speakers. As amply demonstrated by Shaw & Kawahara (2018) and Kawahara et al. (2022), categorical variation of the same underlying form under the same context (e.g., repetitions of the exact same utterance) may exist even within an individual speaker. As such, it is not difficult to picture a scenario where a mix of high-pitched neutral tones (resulting from, say, being preceded by a phonological H) and low-pitched ones have been averaged to a 'mid' trajectory no different from Chen and Xu's data.[13] In addition, in order to elicit naturalistic strings of more than one consecutive neutral tones, Y. Chen & Xu (2006) have used both reduplicated morphemes (e.g., *māma*, 'mother') and underlyingly toneless morphemes (e.g., *men*, pl.) in their material. Neutral tones resulting from phonological processes (e.g., reduplication) and from being underlyingly unspecified are shown to demonstrate different pitch distributions (cf. M. Zhang et al. 2019). Therefore, combining more than one type of toneless syllables in Y. Chen & Xu (2006) may have also brought unwanted variation and lead to mixed results.

Despite being a contentious approach, the treatment of neutral tone as a 'weakly-articulated' tone also finds some support in documentations of Wu Chinese dialects. Zee & Maddieson (1979: 120, 125), while adopting an L-insertion rule for toneless (third) syllables in Shanghai Chinese, also concede that 'the pitch of the third syllable is closer to a phonetic Mid level'. The recent phonetic account of Suzhou by Wang (2011: 91) also argues that toneless morphemes, regardless of their original lexical tone, can be always transcribed as '3' on the Chao tone scale. Similarly, Ling (2011: 1260) makes a remark that toneless syllables at the end of polysyllabic phrases 'tend to concentrate around the mid level of the tonal pitch range'. The analysis of Zee and Maddieson

---

[13] There also have been ample studies within Chinese languages showing variable realization of the same 'neutral' tone. See M. Zhang et al. (2019), Y. Zhang (2021) for Mandarin, and Liu et al. (2021) for Taiwanese Southern Min.

is consistent with that of Y. Chen (2008: 256) — that non-initial syllables are neutralized to a low tone in Shanghai, and the somewhat Mid realization can be seen as the articulatory undershoot of the 'weak' low tone.[14] In addition, Wang (2011: 91) also entertains the concept of 'weakly articulated' ('读得轻') in his description of toneless grammatical morphemes of Suzhou.

Bearing in mind the discussion so far, the current study focuses on teasing apart the potential pitch variability of tonelessness by looking at *token-by-token* realization of toneless moras in Suzhou Chinese. As discussed in Kawahara et al. (2022: 107), investigation on individual pitch trajectories is particularly useful for distinguishing two scenarios of tonal realization: a mix of full 'default L' tones and completely toneless/targetless TBUs on the one hand, and a homogeneous set of 'weakly-articulated' neutral tones (as argued by Y. Chen & Xu 2006; Y. Chen 2008) on the other. In addition, I aim to look into a singular source of tonelessness in Suzhou: that resulting from footing and tonal licensing constraints ((3a); see also de Lacy 2002; Breteler 2018). By including only tonal context and individual speaker as the independent variables, We are able to determine whether a single phonological form (or, more accurately, the *absence* of phonological form) assumes variable surface forms in different conditions. Before outlining the methods of the current study in the next section, I provide the Research Questions and specific hypotheses below.

(5)   Research Questions
   a.   Do toneless moras in Suzhou Chinese realize with a static surface pitch or categorically distinct forms?
   b.   What pitch value(s) do the toneless moras surface as?

Where (5a) concerns with whether the realization is static or variable, (5b) focuses on the exact content of realization. Regarding (5b), I may summarize the following hypotheses from previous studies:

(6)   Hypotheses of toneless realization
   a.   H1/Default L Hypothesis: Toneless moras in Suzhou surface with a low pitch similar to a phonological L tone (M. Chen 2000; H. Zhang 2002)
   b.   H2/Interpolation Hypothesis: Toneless moras in Suzhou remain targetless and surface with interpolated pitch dependent on their surrounding tones (Pierrehumbert 1980; Pierrehumbert & Beckman 1988)
   c.   H3/Spread Tone Hypothesis: Toneless moras in Suzhou surface with a pitch value similar to the tone to their left (Pierrehumbert 1980)
   d.   H4/Weakly-articulated Hypothesis: Toneless moras in Suzhou surface with a weakly-articulated mid pitch, with an intermediate value between L and H (Y. Chen & Xu 2006)

---

[14] It is important to note that Y. Chen (2008) distinguishes the 'weak' low tone from lexical tones that are 'strong' (which includes *lexical* L), 'in that it allows more influence from the preceding lexical tone and it takes longer time to attain its ideal low tone target' (Y. Chen 2008: 256). That is, the crucial distinction lies in the *weak articulation* of the former, while its exact f0 value is less of a concern.

A few clarifications: H1 and H4 are often not well distinguished in the literature and conflated as a single category or pre-determined, 'default' tones. Here, I consider a toneless realization to be weakly articulated (H4) when it can be robustly distinguished from a *lexical* L tone — the exact phonetic value being irrelevant, whereas a 'Default L' (H1) tone is acoustically indistinguishable from a lexical L (cf. Ye 1993: 7). In addition, a toneless TBU undergoing Tone Spreading (H3) may become indistinguishable from other hypotheses under certain tonal contexts (for instance, a Spread Tone LØ → LL is identical to a Default L LØ → LL process). I discuss this issue in §3.1.1.

3 Methods

3.1 Materials

The elicitation materials of this study came from a larger research project investigating the interaction between tone/tonelessness and intonation in Suzhou Chinese. In order to tease apart the realization of toneless moras, the current study included three sets of pitch trajectory data: phrases with medial toneless moras ('Toneless' set), phrases with medial L tones ('L Tone' set), and a set of pitch trajectories emulating linear interpolation between phonological tones ('Simulated' set; see below). The former two data sets were obtained from recorded speech of native speakers of Suzhou Chinese (§3.2), while the last was stochastically simulated using methods outlined in Shaw & Kawahara (2018); Kawahara et al. (2022). Note that the Simulated data set was only created as training input to the Naive Bayes classifiers, and does not otherwise appear in the results.

3.1.1 Toneless data set

The Toneless set consisted of trisyllabic or quadrisyllabic phrases with their third mora being toneless: recall in (3a) that the final (third) mora of a light-heavy disyllable in Suzhou is toneless. In order to control for the potential interaction between the toneless mora Ø and the boundary L%, the light-heavy disyllables in the current study were immediately followed by fully-toned morphemes of various phonological tones, such that Ø is surrounded by phonological tones on both sides.[15] Three tonal contexts were included in the elicitation: one where Ø was surrounded by two H-toned moras ([Tμ.HμØμ.Hμ…]; HØ.H for short), one with preceding H and following L (HØ.L), and one with preceding L and following H (LØ.H). Note that I did not include a 'LØ.L' tonal context as it could not differentiate three out of the four hypotheses: Default L insertion (H1), Interpolation (H2) and Spread Tone (H3). For LØ.L items, the prediction for all three hypotheses would be a low-level pitch. I return to this matter in §6.

Corresponding to the four hypotheses in (6), we may expect the Ø mora to surface in pitch shapes listed in **Table 1**.

---

[15] Both H% and L% are available in Suzhou, but H% only appears in limited sentential contexts. It is therefore easier to elicit items with phonological H as the right context compared to ones with inotnational H% on the right. See Zhu (2023c), Chapter 4.

	H1 Default L	H2 Interpolation	H3 Spread Tone	H4 Weakly-articulated
HØ.H	Fall	High level	High level	Fall to mid
HØ.L	Steep/early fall	Smooth/late fall	High level	Fall to mid
LØ.H	Low level	Rise	Low level	Rise to mid

Table 1: Predictions of pitch shape under different hypotheses and tonal contexts.

Note that the predictions in **Table 1** describe the pitch shape of the *first two* moras of the trimoraic tonal context (the 'HØ' portion of HØ.L, for instance): as Ø is consistently found in the second mora of the second syllable (…T<u>Ø</u>.T…), measuring the entire second syllable instead of only Ø alleviates the potential complications with mora demarcation (i.e., segmenting the first/ second mora of a [TμØμ] heavy syllable). In addition, I refrain from including the right context pitch trajectory into the analysis because unlike the EMA data in Shaw & Kawahara (2018), f0 tracking may become discontinuous if the right context syllable starts with a voiceless onset (cf. Zhu 2023c: 104, fn. 35).

Under H1 (Default L), the three tonal contexts will be equivalent to 'HL.H', 'HL.L' and 'LL.H' respectively: Ø is always supplied with a default low pitch regardless of the context. H2 corresponds to the interpretation that Ø is truly targetless and assumes linearly interpolated pitch — high level between two Hs, falling between H and L and rising between L and H. Note that although similar, the HØ.L context under H1/Default L and H2/Interpolation may differ in the *steepness* of the fall: an inserted L tone under H1 may translate to an earlier/steeper falling pitch (cf. Remijsen 2013). Pierrehumbert (1980) also proposes a Spread Tone (H3) interpretation of Ø, leading to level trajectories in all three contexts, whose pitch height is dependent on the left context. Lastly, according to the target approximation of Y. Chen & Xu (2006), Ø is consistently realized with an intermediate mid pitch, leading to contours in all three contexts.

3.1.2 L tone data set

The L Tone set, on the other hand, contained phonologically L-toned moras instead of Ø in the same prosodic position, followed by fully-toned morphemes — it functioned as a direct test for H1, the Default L hypothesis. In other words, trisyllabic and quadrisyllabic phrases with the tonal contexts HL.H, HL.L and LL.H enable direct comparison with the Toneless set (HØ.H, HØ.L and LØ.H). **Table 2** shows example phrases for each tonal context from both the Toneless and L Tone sets. Note that all tones in square brackets are *surface* forms with relevant tone sandhi applied (Shi & Jiang 2013, Zhu 2023a for data on Suzhou disyllabic tone sandhi).

A note on the position of target syllables is needed. F0 in Suzhou running speech typically converges towards a low pitch target, with the effect increasingly obvious the longer the

expression becomes (Ling 2011; 2014).[16] Due to the nature of disyllabic tone sandhi in Suzhou, a large amount of L tone words contained target syllables in the *initial* position (e.g. HL followed by L; see also **Table 2**, fn. a), unlike the Toneless words, which always contained a peninitial [TØ] syllable. This has led to the undesirable outcome where syllables in the L tone data set might have had an overall higher f0 than ones in the Toneless data set. While it is difficult to carry out perfectly parallel comparisons between L tone and Toneless words (e.g., a phrase-peninitial HL vs. a phrase-peninitial HØ) without making the elicitation items unnatural-sounding, it is important to acknowledge the potential confound of comparing syllables at different phrase positions during the classification analysis.

3.1.3 Simulated data set

The Simulated data set did not contain pitch trajectories measured from fieldwork recordings. Instead, it consisted of trajectories generated from a stochastic sampling process representing how toneless Ø *would have surfaced* under a linear interpolation hypothesis (H2). Again, because these are hypothetical tokens representing one possible strategy to realize Ø, they are not discussed as part of the results. Due to space constraints, I only give a simplified description of the simulation process here (see the detailed steps for operationalization and simulation using Discrete Cosine Transform in Shaw & Kawahara 2018 and Zhu 2023c).

	Weight profile	Context	Example phrase
Toneless set	Tμ.TμØμ.Tμ… (light-heavy…)	HØ.H	[ba^L.**ho:^{HØ}**.sæ:^{H}.tʰã:^{H}] 白蝦燒湯 'white shrimp soup'
		HØ.L	[ɦo^L.**sã:^{HØ}**.wɛ:^{L}.tsã:^{H}] 學生會長 'student organization leader'
		LØ.H	[po^H.**sɿ:^{LØ}**.tsoʔ^H] 八仙桌 'Baxian table' (furniture)
L tone set	TμTμ.TμLμ. Tμ… (heavy-heavy…)	HL.H	[sɛ:^{H}.**tɕʲʏ:^{HL}**.poʔ^H] 三九八 'three-nine-eight'
		HL.L	[**sʉ:^{HL}**.sɿ:^{L}.ho:^{H}] 水仙花 'narcissus'^{a}
		LL.H	[**dɛ:^{L}**.ɦʷã:^{H}.səu^H] 蛋黃酥 'egg yolk tart'

Table 2: Two elicitation data sets. 'Weight profile' column shows the weight and tonal configuration (T = any tone), 'Context' shows the tonal contexts, while 'Example phrase' gives an example of each context with phonetic transcription, Chinese orthography and gloss. Bolded syllables in the transcription stand for the crucial windows where f0 data were extracted and analyzed (e.g., [**ho:^{HØ}**] for the HØ.H context).
^{a}Due to the /HL/ tone sandhi in Suzhou, it is extremely rare to have non-initial HL tones in everyday words and phrases. One example phrase included was [sæ:^{H}.hɛ:^{HL}.tɑ:^{L}] (燒海帶, 'cooked kelp' or 'to cook kelp'; morphological structure does not affect sandhi), while the majority of HL.L words had an initial HL tone.

---

[16] This could be attributed to either a low 'neutral tone' target as Y.Chen and Xu proposed, or an intonational L% tone. The distinction is not crucial to the current discussion.

First, I measured the mean f0 of the left and right tones for each tonal context (e.g., preceding H and following L for HØ.L) per speaker. The mean values of these surrounding tones served as the baseline for linear interpolation — a straight line connecting the left and right context f0 stood for an idealized, 'perfect' interpolation. Pitch variability observed in the naturalistic data was then incorporated by: (i). measuring the standard deviations of the mean and slope for each Toneless context (e.g., all HØ.L tokens) by speaker; (ii). applying the standard deviation values during the sampling process. This step created pitch trajectories with varying means, slopes and contour shapes.[17] The resulting Simulated tokens were based on linear interpolation (H2) but nevertheless resembled 'noisy actuations of phonological goals' (Shaw & Kawahara 2018: 490) — in the current case, the actuations of tonelessness. **Figures 2** give a visual demonstration of extracted Toneless trajectories (HØ.H, HØ.L, LØ.H) of one speaker and Simulated tokens (H-.H, H-.L, L-.H) created from the Toneless sets.

Recall that I focus on presenting the pitch trajectories of the first two (tautosyllabic) moras of the trimoraic window as the right context tone belongs to a separate syllable. For instance, the top left panel shows pitch trajectories of all [HØ] syllables followed by an H tone by Speaker 01 (grey lines; left/right contexts shown as orange circles). On the other hand, the top right panel demonstrates what a H-to-H (stylized as 'H-.H') interpolation for Speaker 01 would look like: a majority of high level trajectories along with a small number of slightly rising or falling ones. A comparison between the left and right columns clearly shows that individual realizations of Ø are not always dependent on the right context. This is best shown in the 'LØ.H' figure (bottom left), where the majority of tokens remained low level despite a high right context in the next syllable.

In addition, note that several Simulated trajectories in the top right panel can be characterized as having 'accidental' tonal targets: just like the 'accidental' vowels discussed in Shaw & Kawahara (2018: 498–499), there are Simulated trajectories *resembling* phonological HL tones simply due to noise present in the phonetic data. When viewed as individual tokens, these falling trajectories would not differ categorically from true /HL/μμ tones in Suzhou Chinese, despite the fact that they were actually sampled from a 'targetless' Interpolation hypothesis. In other words, non-phonological variation in the acoustic signal could nevertheless lead to categorically distinctive tokens (here, high level vs. falling tones). An analytical tool capable of capturing both the overall distribution and individual realizations is the key to understanding the nature of tonelessness (see also §3.4).

While the L tone set constitutes a direct test for H1/Default L, comparison between the Toneless and Simulated sets helps to tease apart whether Ø in Suzhou surfaces as pitch interpolation (H2). Methods for evaluating H3 (Spreading) and H4 (Weakly-articulated) are less straightforward, and will be discussed in §3.4.

---

[17] Discrete Cosign Transform coefficients representing mean, slope and contour shape of each trajectory token were sampled from a normal distribution. See Shaw & Kawahara (2018); Zhu (2023c) for further detail.
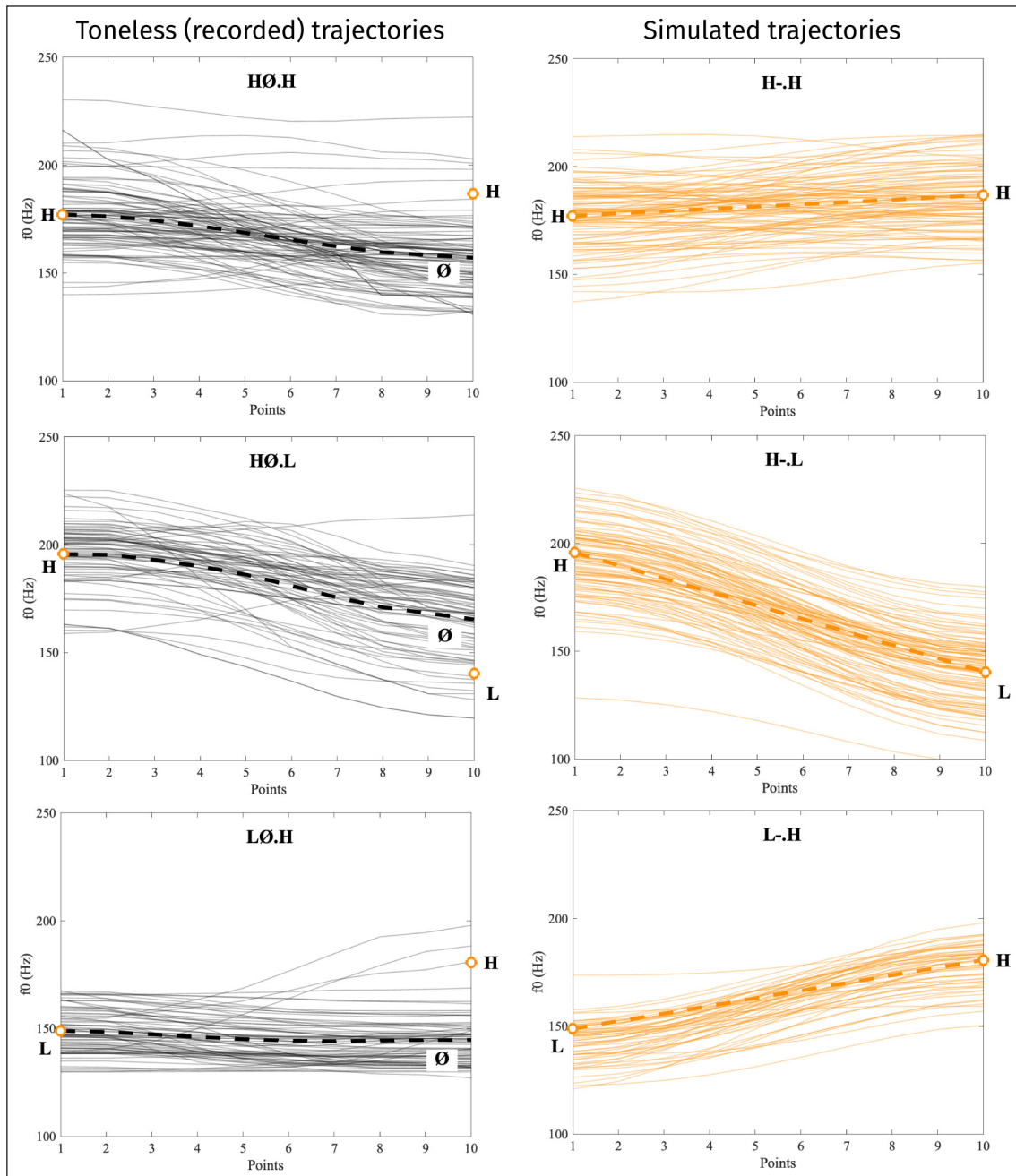
Figure 2: Comparison of Toneless and Simulated trajectories. Left column: raw pitch trajectories the Toneless data set (HØ.H, HØ.L, LØ.H) by Speaker 01. Right column: Simulated trajectories H-.H, H-.L and L-.H. Orange circles stand for the left/right tonal contexts. Dashed lines on the left column stand for the averaged trajectories of all Toneless tokens; Dashed lines on the right stand for the linear interpolation lines.

3.1.4 Segmental and morphosyntactic information of the word list

The word list for the elicitation task contained 37 HØ.H words, 19 HØ.L words and 47 LØ.H words in the Toneless set. The L tone set had 15 LL.H, 10 HL.H[18] and 12 HL.H words. An additional set of 22 words of the HH.H pattern (i.e., /H/μμ followed by another H, [Hμμ.Hμ…]) were added to the word list, both as a distractor and a baseline for H level tones (§3.4). In total, there were 162 trisyllabic or quadrisyllabic words, each repeated three times in a pseudo-randomized list, yielding 486 tokens per speaker. Segmental information (e.g., vowel quality of the target syllable, preceding and following syllable onsets) was not actively controlled for, as the primary goal was to generate data sets containing as many eligible naturalistic expressions as possible.

One reviewer has raised the issue of micro-prosodic perturbations and their potential influence on the pitch data — that f0 can be influenced by both onset voicing and vowel height (also known as *CF0, VF0*; Whalen & Levitt 1995; Kingston 2007; 2011; see, in particular, Luo et al. 2016; Kirby 2018; Shi 2020; Yu 2022 for data of tonal languages). While I acknowledge that there was no control for potential micro-prosodic effects, the crucial analytical window of the toneless data — the second syllable of each polysyllabic phrase, either HØ or LØ — *undergoes a neutralizing tone sandhi.* Most of the relevant CF0/VF0 studies to date have studied carefully-controlled (near-)minimal pairs in isolation or within certain carriers. In the context of tonal languages, the study stimuli often take the shape of monosyllables of contrasting tones/onsets/vowels. To the best of my knowledge, none has explored what effects CF0/VF0 would have on syllables undergoing tone sandhi, and the conclusions of previous micro-prosodic studies may not be immediately transferable to the current case. Further complicating the situation is the *inconsistency* of potential CF0/VF0 effects across different studies, as many authors have pointed out that the influence of micro-prosody may vary by (i). specific language; (ii). individual speaker; (iii) certain lexical tone, or a combination of the above (Luo et al. 2016 being an examplar explicitly addressing the 'inconsistent consonantal effects'). Thus, while the issue of micro-prosody is an important topic of discussion, it invites uncertainties that are beyond the scope of the current paper.

Lastly, some discussion on the morphosyntactic composition of the word list is in order. As widely studied by tonologists of Wu Chinese, tone sandhi of many Wu varieties can be divided into 'Broad Form' (廣用式) and 'Narrow Form' (窄用式) (Qian 1992: 622–624 and H. Zhang 2016: 90 for general discussion; studies discussing Broad vs. Narrow sandhi in Suzhou include Xie 1982; Qian & Shi 1983; Wang 2011, among others). Details aside, Broad Form is found in most noun phrases and demonstrates the 'left-dominant' tone sandhi patterns shared by many

---

[18] Due to disyllabic tone sandhi, the HL.H context does not occur within one sandhi domain, hence the relatively low number of eligible items.

Wu varieties (Yue-Hashimoto 1987; Duanmu 1995); Narrow Form, on the other hand, is typically found in Verb-Object or Adverb-Verb phrases, and has been analyzed as a non-application of left-dominant sandhi (followed by tone reduction).[19]

The current Toneless and L tone data sets are created under the assumption that left-dominant (Broad Form) tone sandhi does in fact apply within the trisyllabic/quadrisyllabic phrases, or at least within the relevant domains where I extracted the f0 data. This is indeed the case: the majority of the elicitation items were noun compounds and idiomatic phrases. For quadrisyllabic phrases, the most common sandhi structure was (σσ)(σσ) — two disyllabic tone sandhi domains. Trisyllabic noun phrases largely follow the (σσ)(σ) structure — a disyllabic modifier plus a monosyllabic head; see (7b). Occasional Verb-Object or Adverb-Verb phrases (e.g., 燒湯 [sæːᴴ.tʰãːᴴ], literally 'to cook soup'; see **Table 2**, HØ.H context) were only included to create specific preceding and following tonal contexts (in this case, a H tone following Ø). In short, I ensured that all phonetic data came from polysyllabic phrases with 'intended' tonal patterns. Below I give several example words with relevant morphosyntactic and tonal information. A full list of all elicitation words can be found in the supplementary files submitted to an OSF page.[20]

(7) Example words with morphosyntactic and tonal information. Brackets indicate morphosyntactic boundaries/tone sandhi domains.

    a. 北方省份 [poᴴ.fãːᴸᴼ.sãːᴴᴸ.vənᴸ] 'northern provinces'

       (po.fãː)   (sãː.vən)

       northern   province

       A Modifier-Head noun compound with two disyllabic sandhi domains. This is the most common word type in the elicitation list.

    b. 橘子水 [tɕʲəᴴ.tsɨːᴸᴼ.sʉːᴴᴸ] 'orange juice'

       (tɕʲə.tsɨː) (sʉː)

       orange   water

       A Modifier-Head noun compound. Third syllable forms its own sandhi domain, creating a HL tone immediately following LØ (LØ.H context).

    c. 早檢查 [tsæːᴴᴸ.tɕɪːᴴᴸ.zoːᴸ] 'do the checkup early'

       (tsæː) (tɕɪː.zoː)

       early   checkup

       An Adverb-Verb phrase. First syllable forms its own sandhi domain (intentional 'Narrow Form' sandhi). Two HL tones in succession give the HL.H(L) context.

---

[19] A well-known example in Shanghai Chinese is 炒飯, with two possible pronunciations depending on morphological structure: [tsʰoᴹ.vɛᴴ], (tsʰo_A vɛ_N)_NP ('fried rice', Broad Form sandhi); [tsʰoᴹ.vɛᴸᴹ], ((tsʰo_V)(vɛ_O))_VP ('to fry rice', Narrow Form sandhi). Brackets correspond to tone Sandhi domains. For discussion, see H. Zhang (2016: 46–49).

[20] https://osf.io/nawk5/.

3.2 Participants

The elicitation data came from my fieldwork conducted in 2022. I recruited 16 native speakers of Suzhou Chinese, all of whom were born and raised in Suzhou City and did not leave the city for a significant period before college. There were 8 males and 8 females (gender was self-reported), with ages ranging from 30 to 70 at the time of fieldwork (mean = 50.4). Due to the ongoing pandemic and international travel restrictions, fieldwork recordings were acquired in a hybrid manner. I recorded 2 speakers in person (both males, aged 31 and 30) in the United States, while the remaining 14 speakers (located in Suzhou, China) were recruited via the Internet and instructed to record using their own smartphones (see §3.3). No participant reported any speaking or hearing disorders. Each Suzhou speaker signed an informed consent script before elicitation and was paid 15 USD or 100 CNY for their participation.

3.3 Procedure

The elicitation was conducted in a hybrid manner, and the procedure differed slightly depending on the mode. For the two in-person sessions, the participants recorded their speech in a relatively quiet room using a Shure SM10A-CN head-worn microphone and a Zoom H4N PRO digital recorder at a sample rate of 44100Hz. For the 14 remote Suzhou speakers, I instructed them to record their own speech in a quiet room using a smartphone. Occasional higher sampling rate recordings (typically 48000Hz) were downsampled to 44100Hz. All recordings were normalized to –1.0dB and noise-reduced using Audacity.

The elicitation session consisted of three parts: a sociolinguistic survey with questions about the participants' life in Suzhou city and their experience with the local language, the word list reading task, and a set of sentences in Suzhou. The sociolinguistic survey helped the speakers get accustomed to being recorded and also code-switch from Mandarin Chinese, the dominant language for younger speakers. During the reading task, speakers read from a PDF file word tokens described in §3.1 in a carrier sentence [kã:^{HL} ___ pə^H.ŋəu^{LH}.tʰɪn^H] (講___拔我聽, 'Say ___ for me.').[21] The PDF file also contained written instructions for remote participants, and all participants completed the reading task without any confusion.

3.4 Analysis

After recording, the word list data were annotated using Praat (Boersma & Weenink 2022). f0 data in Hertz[22] were extracted using the ProsodyPro Praat script, with each trajectory token

---

[21] Elicitation words and carrier sentences were written in Traditional Chinese script to accommodate older speakers who primarily used Traditional Chinese. Younger participants on the other hand had no issue with reading Traditional Chinese.

[22] One reviewer asked why the pitch data were not converted to a less speaker-dependent scale such as semitone or Bark. I refrained from cross-speaker normalization as (i). this would make the results less comparable to studies

containing 10 time-normalized data points (Xu 2013). For words in the Toneless and L tone sets, f0 trajectories of both the second and third syllables were extracted. The trajectories of the second syllables (HØ, LØ of the Toneless set and HL, LL of the L tone set) were the main focus of the analysis, while the third syllables were used to acquire the right tonal context values (following H or L) as discussed in §3.1.

In order to obtain token-by-token evaluations regarding the realization of Ø in Suzhou, I follow Shaw & Kawahara (2018) and M. Zhang et al. (2019) and implemented Naive Bayes classification models. The model details are as follows.

For each speaker and each tonal context, a Naive Bayes classifier was trained on L tone (representing H1/Default L) and Simulated (representing H2/Interpolation) trajectory tokens. I sampled the same number of Simulated tokens as L tone ones to ensure that the training was not biased towards either category. The resulting classifier, when presented with a Toneless token, was able to judge how likely it was sampled from the L tone vs. Simulated set. I give a concrete example below. To explore the realization of Ø between two H tones (i.e., HØ.H) for Speaker 01, a classifier was trained on equal HL.H (L tone) and H-.H (Simulated) tokens. All HØ.H trajectory data of Speaker 01 were then individually fed to the classifier, yielding posterior probability values for how closely each HØ.H token resembled HL.H (H1/Default L) or H-.H (H2/Interpolation) — if a posterior probability of 0 stands for high model confidence for a 'Default L' token, 1 then represents a probable Interpolation/'targetless' token.[23]

The resulting posterior probability values, when plotted on a histogram, show the distribution of a speaker's toneless realization under a given tonal context: while it is possible to observe a majority of probabilities leaning towards either H1/Default L or H2/Interpolation (a peak at either end of the probability scale), there may also be a case where the distribution peaks at *both* probabilities — Ø can realize categorically yet optionally as *either* L or linear interpolation. Such a 'double-peak' pattern has been observed in vowel articulation (Shaw & Kawahara 2018), pitch accent neutralization (Kawahara et al. 2022), and more importantly, in the realization of Mandarin neutral tones (M. Zhang et al. 2019). One additional possibility was referred to as the 'Reduced hypothesis' by Shaw & Kawahara (2018), which depicts a distribution peaking at 0.5 probability: the model classifies the majority of Toneless tokens as neither L-toned (H1) nor linear interpolation (H2), but an 'in-between' value. This scenario corresponds well to the Weakly-articulated hypothesis (H4) argued by Xu and colleagues, where Ø is predicted to take an intermediate, 'in-between' pitch value. In summary, Naive Bayes models trained by L tone and Simulated data sets are capable of testing three hypotheses in (6): H1, H2 and H4. I refer interested readers to Shaw & Kawahara (2018: 503) for a visualization of the possible outcomes discussed so far.

---

such as M. Zhang et al. (2019) and Kawahara et al. (2022); (ii). this would potentially collapse some cross-speaker variation, which was part of my main research focus.

[23] Which end of the probability scale represents which category is inconsequential, as it only depends on how the model was coded.

The Spread Tone hypothesis (H3) is a special case that invites further elaboration. As shown in **Table 1**, H3 gives a prediction identical to H2/Interpolation under the HØ.H tonal context, and one identical to H1/Default L under the LØ.H context — When surrounded by two H tones, a majority of high level Ø tokens may either result from linear interpolation or spread tone. Similarly, when Ø is preceded by (tautosyllabic) L and followed by H, it may receive a low pitch either by 'Default L' insertion or rightward spreading. Consequently, HØ.L is the only tonal context where H3 predicts a tonal shape distinct from all other hypotheses: (high) level vs. different types of contours. Therefore, an additional classification analysis was performed for the HØ.L context, the methods of which are described below. For each speaker, I first trained a separate Naive Bayes classifier based on *slope values* of /HL/μμ and /H/μμ tones[24] from the same speaker. This 'slope' classifier was able to evaluate whether a Toneless token is a contour (H1, H2 or H4) or a level tone (H3). After identifying the level (Spread Tone) HØ.L tokens and removing them from further analysis, the remaining tokens were submitted to a HL.L (H1/Default L) vs. H-L (H2/Interpolation) classifier created by the methods discussed above. In short, for the HØ.L context in specific, a 'slope' classifier was first trained to identify level (Spread Tone) tokens. Non-level HØ.L tokens were then subjected to the classification analysis no different from HØ.H and LØ.H contexts.

For all Naive Bayes classifiers, the input data were partitioned into a 70%/30% train vs. test split. Resubmitting the test portion data yielded over 90% model accuracy on average — the classifiers were able to accurately identify pre-assigned classes 90% of the time. To ensure that the classification results were not biased by a particular set of Simulated tokens, the results I report were aggregated over 10 simulation iterations. That is, for each speaker and tonal context, 10 classification models were trained based on the same L tone tokens and 10 randomly-sampled Simulated sets. The posterior probability for each classified Toneless token was then an average of these 10 iterations.

4 Results

Since the main focus of this study is the potential cross-speaker and cross-context variation of Ø realization in Suzhou Chinese, the presentation of results will be separated by tonal contexts where I report the Naive Bayes classification output by speaker. In addition, I also include pitch plots for several qualitatively distinct patterns to highlight the extent of variation across speakers and contexts.

One cautionary note before I present the classification results: due to a mix of remote recording difficulties, excessive background noise and tracking failures, there were several speakers with far fewer pitch trajectories eligible for analysis than others (Speaker 09, for instance, only had 37 HØ.L tokens extracted from their 19*3 = 57 word list). For the sake of completeness, I chose to

---

[24] Tokens of these two lexical tones were extracted from HL.H and HH.H words respectively; see also §3.1.

include all speakers regardless of the number of their eligible trajectories. Nevertheless, one may remain skeptical about the robustness of classification models when the corresponding training set size is too small.

4.1 HØ.H — toneless mora between two H tones

I begin by presenting the classification results of HØ.H tokens. Recall that individual Toneless tokens for each speaker was evaluated by a Naive Bayes classifier trained on two categories: HL.H, a phonological L tone between two H, and H-.H, a high level pitch between two H. While the former represents H1/Default L for toneless realization, the latter could be attributed to either H2/Interpolation or H3/Spread Tone. Lastly, the Weakly-articulated hypothesis (H4) can be represented as an intermediate probability between the two classes: the HØ.H token does not strongly resemble either HL.H or H-.H, but realizes with an 'in-between' pitch value. In **Figure 3** I show the aggregated plot of the histograms for the 16 speakers.
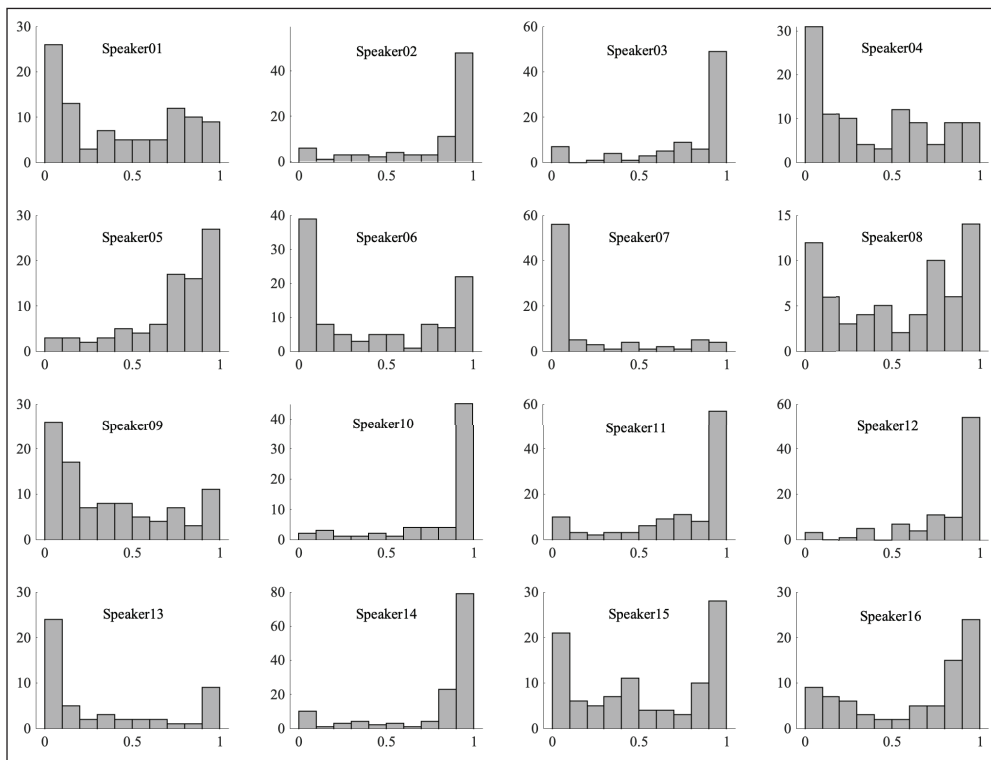
Figure 3: Histograms of HØ.H tokens. 0 = Default L; 1 = Interpolation or Spread Tone.

As discussed in §3.4, the two ends of each histogram stand for one of the two classified categories: here, The x axis stands for the posterior probability of H-.H classification. '1' on the probability scale represents a high model confidence for either H2/Interpolation or H3/Spread Tone, while '0' stands for the alternative 'Default L' realization (H1). As can be seen in the figure, there was

considerable variation between speakers under the same tonal context. For instance, Speaker 02 had a peak at the high probability end in their histogram, indicating that most of their HØ.H tokens were high levels resulting from either Interpolation or Spread Tone. On the contrary, Speaker 07 had their frequency peak at the opposite end, showing a majority of trajectories compatible with a Default L insertion analysis. To demonstrate how the Naive Bayes classifiers assigned categories to each Toneless token, consider the following pitch trajectory plots for Speakers 02 and 07.

Instead of plotting individual tokens, all classification plots contain smoothing spline trajectories of each class (e.g., H1, H2) during the Naive Bayes classification process, with shadings representing the 95% confidence intervals. To illustrate the distribution of classified labels (i.e., how many tokens were classified as H1, H2, etc.), different line colors represent the competing categories, while the thickness of the lines stands for the frequency of each category — that is, a thicker line indicates that a larger number of tokens were identified by the model as the corresponding class. A full list of raw f0 trajectories by speaker and tonal context can be found on the OSF project page.[25]

Two aspects of **Figures 4a** and **4b** deserve further elaboration. First, one would need to decide on the threshold probability values for each category. I consider (i). a token with a posterior probability higher than 0.6 as a possible Interpolation (H2) or Spread Tone (H3) classification; (ii). one with lower than 0.4 probability as Default L (H1); (iii). tokens ranging between 0.4–0.6 as the 'intermediate' realization between HL.H and H-.H, i.e. the 'Weakly-articulated' tokens (H4). Secondly, several speakers had very few tokens of certain classes (for instance, Speaker 02 had 6 out of her 84 HØ.H tokens identified as H4). The scarcity of certain tokens has potentially contributed to higher variability of their corresponding classes. As a result, some of the plots contain thin lines with wide confidence interval bands, as can be seen in **Figure 4a** regarding both H1 (blue) and H4 (black).
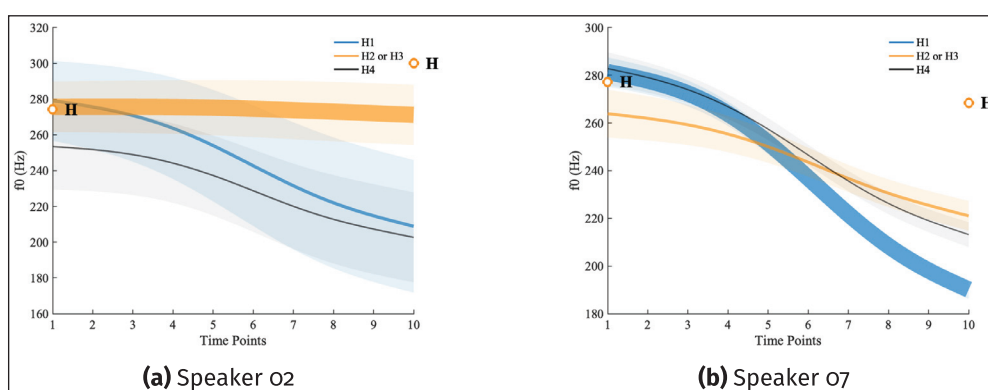

Figure 4: Classification plots for two speakers, HØ.H tokens.

The two classification plots show the notable contrast between Speakers 02 and 07 in realizing their toneless moras between two H tones: the majority of Speaker 02's HØ.H tokens were high levels, with very few falling (H1/Default L) trajectories — hence the thick orange line. On the other hand, Speaker 07 realized most of her HØ.H tokens with a falling HØ syllable, despite the following H tone to the right (right orange circle in **Figure 4b**). For Speaker 07, even the tokens classified as Interpolation or Spread Tone (orange) took the form of a 'milder' fall when averaged. Also noteworthy is the observation that neither of the speakers had a large number of H4/Weakly-articulated tokens: when judged individually, toneless trajectories were rarely classified as 'intermediate'/'Weakly-articulated'. If, according to Y.Chen & Xu (2006) and Wang (2011: 91), toneless moras were to have a stable M target, we would have expected a majority of falling pitch trajectories ending at an intermediate value between H and L, and being classified as H4 (0.4–0.6 posterior probability). Tokens classified as H4/Weakly-articulated, however, were the least frequent type in both cases as indicated by the line thickness.

One may still argue in favor of the weakly-articulated hypothesis in that the central tenet of Y. Chen & Xu (2006) is that neutral tone/tonelessness has a *stable* articulatory target — the actual value of the target is less relevant, as Y. Chen (2008) has explicitly characterized the Shanghai tone sandhi as neutralizing to a (stable) Low target. Several participants in my data also demonstrated qualitatively distinct but systematic Ø realizations under the same tonal context, calling into question the assumption of having one stable target. Speaker 06, as the histogram in **Figure 3** has shown, had a 'double peaks' or bimodal pattern, indicating that the Naive Bayes classifier identified a mix of *both* H1 (high falling) and H2/H3 (high level) tokens. This can be observed in **Figure 5**: of the 103 HØ.H tokens submitted for classification, 53 were identified as H1 (blue), and 38 were H2/H3 (orange). Similar within-speaker variability was also observed in Shaw & Kawahara (2018) for devoiced vowel articulation in Japanese and M. Zhang et al. (2019) for neutral tone realization in Mandarin. The current Suzhou data provide further evidence for variable phonological processes (Coetzee & Pater 2011; Coetzee & Kawahara 2013), where Ø was not mapped to a single output form, even for the same speaker and tonal context.

4.2 HØ.L — toneless mora with preceding H and following L

I now turn to the tonal context HØ.L. It is the only context where one can distinguish H3/Spread Tone from the rest of the possibilities (**Table 1**): while H1/Default L, H2/ Interpolation and H4/ Weakly-articulated predict different types of falling trajectories, HØ syllables under a Spread Tone hypothesis should realize as high levels regardless of their right context L. As discussed in 3.4, I first trained a Naive Bayes classifier for each speaker with the purpose of identifying level tokens.[26] The remaining falling trajectories were then submitted to a HL.L vs. H-.L classifier to distinguish between H1, H2 and H4.

---

[26] As the classification between level and falling tokens was rather clear-cut with few intermediate posterior probability values, I simply classified any trajectory with a higher than 0.5 /H/ tone probability as 'level' for our purposes.
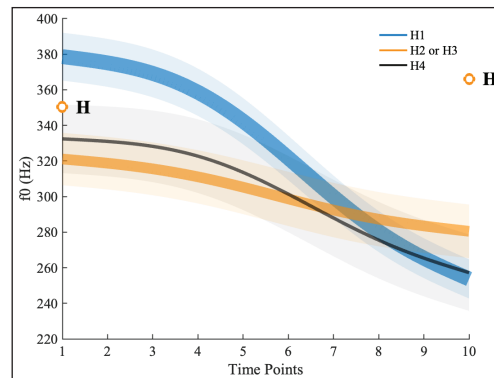
Figure 5: Classification plot of Speaker 06, HØ.H tokens.

The resulting aggregated histogram plot is shown in **Figure 6**. It is important to note that these histograms reflect the distribution of *falling/non-level* trajectories only (as level trajectories were removed from the classification analysis).[27] From the overall appearance, the HØ.L context was not much different from HØ.H: there was both inter- and intra-speaker variation, as the histograms for most speakers peaked at one or both ends but not in the middle. However, upon closer observation, we may see potential shortcomings in the classification process. Consider Speaker 12 in **Figure 7a**, who was representative in showing three types of HØ.L trajectories: higher falls, lower falls and levels. First, averaged H1/Default L and H2/Interpolation trajectories are clearly separated, and the HØ.L tokens were rarely identified as H4/Weakly-articulated (thin black line): the Naive Bayes classifier trained on Speaker 12 was capable of categorizing the individual tokens with high confidence. However, on closer examination, the classification results were incongruent with my prediction in **Table 1**: the contrast between an inserted L tone and linear interpolation in the HØ.L context was predicted to be a difference between steeper/earlier falls (similar to HL.L) vs. smoother/later falls (H-.L). **Figure 7a**, on the other hand, shows that when averaged, the *slopes* of H1/Default L and H2/Interpolation were almost identical, while the starting and ending f0s were largely different — it is more accurate to characterize the H1 vs. H2 contrast as higher vs. lower falls here.

There are two possible explanations for these findings. First, recall that the training input for the HØ.L context contains phonological /HL/μμ tones (followed by /L/μμ) and simulated H-.L trajectories. /HL/μμ as a lexical tone has the highest starting pitch (hence the '51' transcription on the Chao tone letter scale). Using /HL/μμ as a stand-in for the Default L hypothesis would inevitably lead to HØ.L tokens with higher starting pitch being classified as H1. In addition, we may observe from **Figure 7a** that the right context L tone (represented as the orange circle) had

---

[27] One may ask whether it is more beneficial to train Naive Bayes models containing *three* classes in this case: Default L (H1), Interpolation (H2) and Spread Tone (H3). Although this is possible, we would in turn be unable to neatly represent the Weakly-articulated hypothesis (H4) as 'neither Default L nor Interpolation' on the histograms (see Shaw & Kawahara (2018)).

a low average f0: around 150Hz for Speaker 12. HØ.L tokens that ended with a relatively low f0 were in fact more similar to an Interpolated hypothesis in this case. In short, it is more accurate to say that the classifier models in the HØ.L context were trained to distinguish between */HL/ μμ-like* falls and linear interpolations, which may be a questionable practice in itself. To that end,
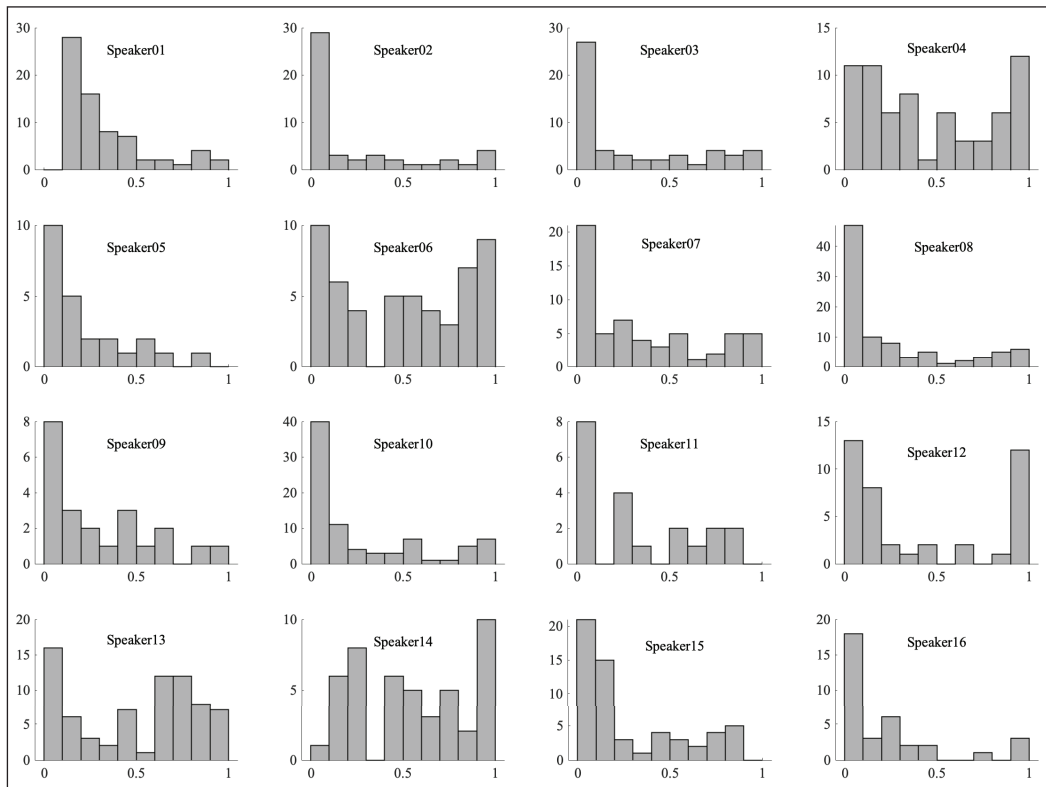

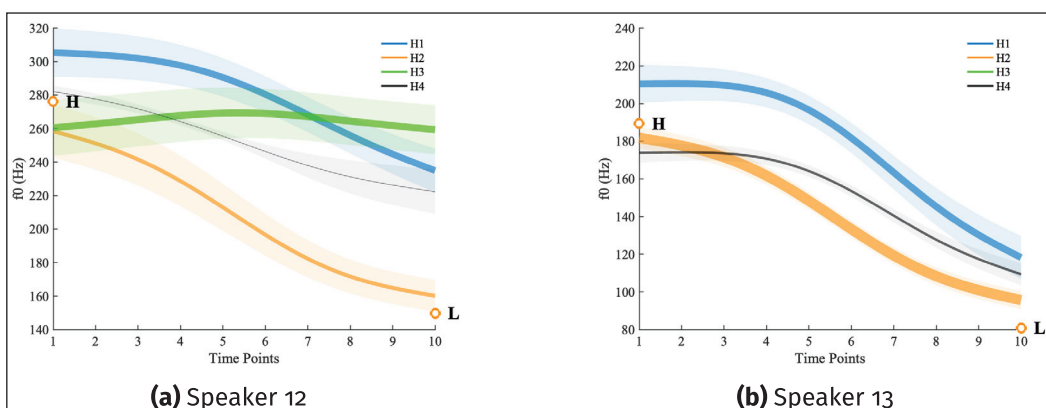**Figure 6:** Histograms of HØ.L tokens (after removing level tokens). 0 = Default L; 1 = Interpolation or Spread Tone.

(a) Speaker 12   **(b)** Speaker 13

Figure 7: Classification plots for two speakers, HØ.L tokens.

one may argue that /HL/μμ fall represents the realization of a lexical tone, which is inherently different from a *sandhi* fall resulting from light-initial sandhi (e.g., [Hμ.**HμLμ**]). While I remain agnostic on this issue, there is an alternative that may circumvent the problem of distinguishing lexical vs. sandhi tones: since there is ample data for the sandhi 'L' tone in Suzhou, one may attempt to create simulated datasets of HL tones as part of the classification input. This, however, would yield classifiers that distinguish two *simulated* classes, which may also be undesirable as a method.

Notwithstanding the somewhat surprising Naive Bayes classification results, the clear separation of largely parallel higher vs. lower falling pitch trajectories provides additional counter-evidence to the 'Weakly-articulated' hypothesis: a core assumption of Y. Chen & Xu (2006: 51) is that tonelessness always has a stable target, and variation in f0 is no more than differing degrees of reaching the intended target, which is caused by the highly variable phonetic duration of toneless TBUs. Two sets of higher vs. lower falls, following this claim, should in turn average to two trajectories *converging* towards the intended toneless target (since there is no principled reason to believe that higher vs. lower falls differ systematically in *duration*). This is, however, not the case, as the two averaged trajectories in **Figure 7a** remain almost parallel with considerable pitch height difference.

Aside from the two types of falling trajectories identified by the classifier, the remaining majority of Speaker 12's tokens were classified as *levels* (green line). Levels in this particular context could only be attributed to a Tone Spreading hypothesis, where the toneless mora realized with the same pitch as its left context. I can conclude from the HØ.L data that H3 — the Spread Tone hypothesis — was also one of the possible realizations of toneless moras in Suzhou Chinese.

Similar to the HØ.H context, there is also considerable cross-speaker variation in realizing a HØ.L trajectory. For example, Speaker 13 had *none* of their HØ.L tokens classified as level (green), and instead only had falling trajectories of varying degrees. This can be seen in **Figure 7b**. Again, there was a clear separation between averaged trajectories representing H1/Default L and H2/Interpolation, with non-overlapping confidence intervals. Similar to Speaker 12, few tokens were identified as weakly-articulated.

4.3 LØ.H — toneless mora with preceding L and following H

Lastly, I present the results of the LØ.H context, where a toneless mora was preceded by a tautosyllabic L, and followed by a H tone in the next syllable. For this context, two hypotheses might correspond to the same surface realization: both H1/Default L and H3/Spread Tone may lead to a low level heavy syllable (/LμØμ/ → [Lμμ]). H2/ Interpolation and Weakly-articulated/ H4, on the other hand, could be identified from different degrees of rises.

The aggregated classification histogram in **Figure 8**, when compared to **Figure 3**, shows a slightly different pattern. Recall that for the HØ.H context, there were speakers with a majority of high level tokens (Interpolation or Spread Tone) and those with a majority of falls (Default L) — one may consider them as direct opposites on a 'level vs. contour realization' continuum. For the LØ.H context, however, there seemed to be no speaker with an overwhelming majority of contour tokens (in this case, rises): speakers either realized most of their tokens as levels (H1 or H3), or had a bimodal pattern with considerable amounts of both levels (Default L or Spread Tone) and rises (Interpolation). I give an example of each condition below, with Speaker 04 standing for a 'majority low levels' realization and Speaker 09 for a 'levels and rises' (i.e. bimodal) realization.
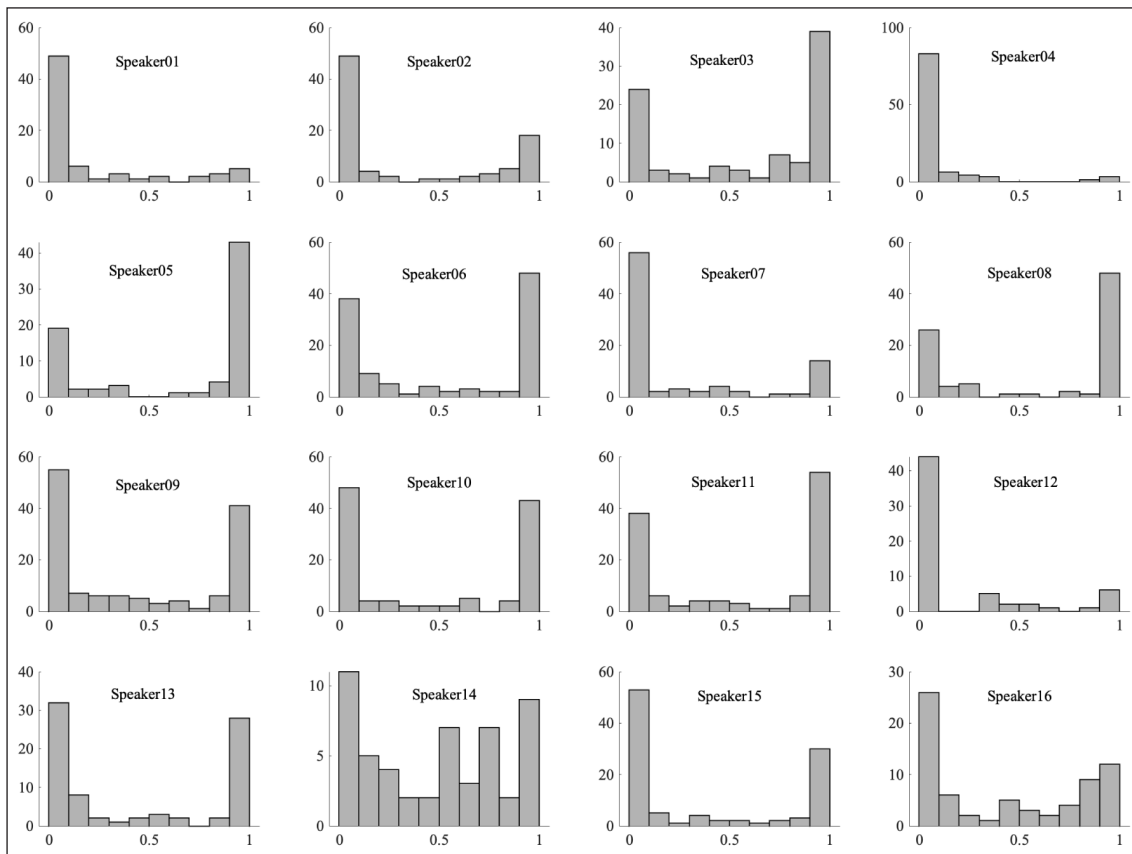
Figure 8: Histograms of LØ.H tokens. 0 = Default L or Spread Tone; 1 = Interpolation.

As shown in **Figure 9a**, the vast majority of LØ.H tokens were identified as 'level', standing for either a Default L insertion hypothesis or a Tone Spreading hypothesis. There was, in fact, one token ([zəᴸ.sɛᴸᴼ.tɪːᴴᴸ] 十三點, 'idiot', an idiom) that was classified as Interpolation/H2, also evident from its visible rise. Morphologically, the phrase would have invoked two tonal domains((sə.sɛː)

(tɪː), literally 'thirteen' and 'o'clock'), but the highly idiomatic usage (as a common insult term) has potentially made the expression behave like one tonal unit. I leave the topic of morphosyntax and its influence on toneless realization for future research. The remaining LØ.H tokens, on the other hand, were all identified as 'low levels', and were averaged to a clearly falling trajectory. For this speaker, the right context H exerts little to no influence on the realization of Ø, as it often continues to fall throughout the bimoraic syllable (typical for L tones in Chinese languages; see, for instance, Yip 2002: 22).
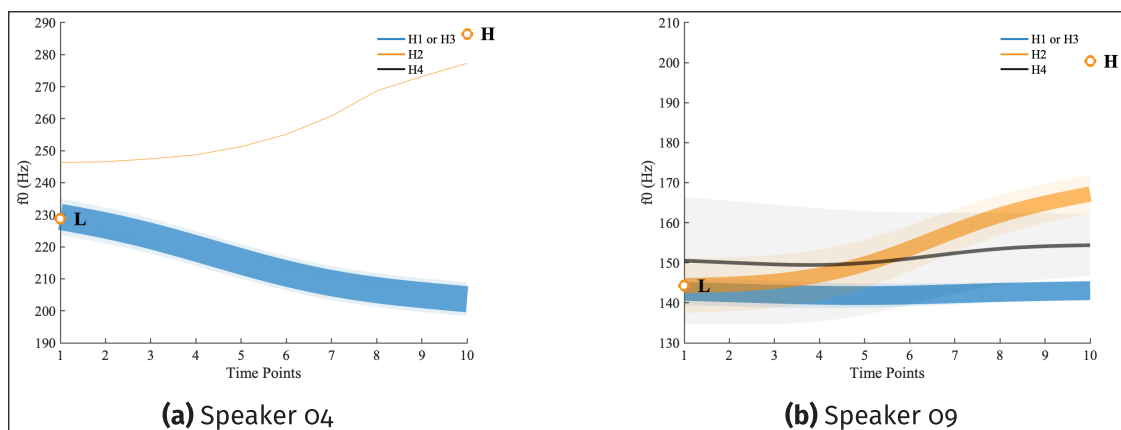
Figure 9: Classification plots for two speakers, LØ.H tokens. The H2 (orange) line in 9a does not have a confidence interval as there was only one H2/Interpolation token.

In contrast, Speaker 09 had an almost even mix of low level tones and rising tones. The rises resembled an interpolation from the left context L to the right context H (shown as orange circles in **Figure 9b**), indicative of an Interpolation hypothesis (H2).[28] This type of 'bimodal' realization of tonelessness again provides counter-evidence to the proposal that toneless moras in Suzhou, being the 'weak' elements in running speech, are realized with one stable target (Y. Chen & Xu 2006).

As an interim summary, systematic and categorical variation was prevalent in how Suzhou speakers realized their toneless moras. I found variation both within (in the form of multiple identified classes for the same speaker) and across speakers (in the form of multiple distribution patterns under the same tonal context). In addition, the three tonal contexts investigated in the current study also demonstrated variation that was sensitive to context. The HØ.H context witnessed three types of speakers: ones with a majority of high falls, ones with a majority of high levels, and ones with both. HØ.L, due to its unique tonal makeup, was both challenging

---

[28] While the averaged trajectories all started at roughly the same low f0, the averaged endpoint of the Interpolation/ H2 class was significantly lower in pitch than the right tonal context. Interestingly, this resembles the 'asymptotic approximation' model proposed by Y. Chen & Xu (2006: 67): the right context H as the articulatory target had a relatively high f0, and there was not enough duration for the interpolation to complete.

for the Naive Bayes model,[29] but also informative in providing direct evidence for the Spread Tone hypothesis (H3): none of the other three hypotheses (Default L, Interpolation, Weakly-articulated) was able to account for high *level* tokens in the HØ.L context. Realization in the LØ.H differed from HØ.H in that although speakers with many low levels and those with both low levels and low rises were found, there was a lack of participants who had a majority of low rises (representing H2/Interpolation) only. Translated into posterior probability histograms (**Figure 8**), there were speakers with a peak at 0 or two peaks, but none with a single peak at 1.

5 Discussion

5.1 Identifying systematic variation from noisy phonetic data

This study approaches variation in phonetic data using the simulation and classification methods in Shaw & Kawahara (2018). By training classification models containing a 'targeted' hypothesis and a 'null' hypothesis (H1/Default L vs. H2/Interpolation in the current study), the current analysis sheds light on several crucial aspects regarding the realization of tonelessness in Suzhou Chinese.

First, taken holistically, there was no single hypothesis capable of comprehensively capturing the realization data — variation was present both within and across speakers. Instead of a one-to-one relationship (e.g., /Øμ/ → [Lμ], in SPE terms), it is more appropriate to adopt a phonological model allowing for variable realization (e.g., Coetzee & Pater 2011 and Coetzee & Kawahara 2013) and argue that toneless realization in Suzhou is one-to-many.

Secondly, the training input for the Naive Bayesian models was set up to evaluate not only H1/Default L and H2/Interpolation, but also a 'Weakly-articulated' or 'reduced realization' hypothesis (H4; Shaw & Kawahara 2018, Kawahara et al. 2022). Similar to what Shaw, Kawahara and colleagues found, there was insufficient evidence from my data to argue that Ø in Suzhou has a stable and *reduced* pitch target. I return to this point in more detail in §5.2.

Lastly, the possibility of a Spread Tone hypothesis (H3) is what distinguishes the current study from similar toneless research such as M. Zhang et al. (2019): in the Mandarin tonelessness study by M. Zhang and colleagues, the toneless tokens were always evaluated against one 'fully-toned' hypothesis (Tone 3 or Tone 4 depending on the stimuli). In contrast, I have established *two* possible full tone targets for toneless moras in Suzhou in (6): Default L (H1) or Spread Tone (H3). Crucially, The HØ.L analysis was split into two classification tasks — the first distinguishing levels from contours (based solely on slope), while the second further diving contours into Default L (H1) and Interpolation (H2). I demonstrate that in addition to the commonly-accepted

---

[29] While the grand mean loss rate of all classification models was 8.7%, the mean classification loss of the HØ.L context was 16.8%. That is, 16.8% of the HØ.L *training* data were incorrectly classified by the trained models, implying that the models performed worse in terms of accuracy when compared to other contexts.

Default L analysis (M. Chen 2000; Yip 2002; see also §2.1), it was also possible for a Ø mora to receive its tonal specification from its tautosyllabic neighbor tone: /TμØμ/ → [Tμμ]. Similar results are reported in Liu et al. (2021), where two types of Neutral Tones in Taiwanese Southern Min realize as a stable L target and 'tone spreading/extension' respectively. This tone spreading process, along with linear pitch interpolation, has been argued by Pierrehumbert (1980) and Pierrehumbert & Beckman (1988) as one of the ways underspecified TBUs may realize in non-tonal languages (e.g., English and Japanese). Here, I show that a tonal language like Suzhou Chinese may also choose tone spreading as its toneless realization strategy. An anonymous reviewer has insightfully pointed out that this observation echoes what Hyman (2009) argues for in his 'Property-driven Typology' approach — as numerous accounts have shown that tonal languages can exhibit stress-like properties (e.g., metrical structure, interaction with intonation), and the same is true for typical stress languages with tonal operations, a typology that aims to sort languages into 'stress', 'pitch accent', 'tone' prototypes is a somewhat fruitless endeavor. A typological continuum from 'stress language' to 'tone language' does not in itself provide much insight, since one can rarely place any language on the scale with certainty.

In addition to presenting an empirical case of variable toneless realization, an understudied phenomenon in Suzhou Chinese, the current study also contributes to the methodological discussion of using computational models to analyze noisy phonetic data. The robustness of Naive Bayes classification models manifests primarily in three ways. Firstly, they allow relatively fast and reliable classification of a large amount of phonetic data (here, 16 speakers X 486 tokens per speaker = 7776 individual tokens), where manual classification by fieldworker(s) would be either logistically difficult or untenable. More importantly, these models reduce errors due to subjectivity, especially when multiple fieldworkers involved in a study have different levels of exposure to tone languages. Apart from few classification errors that could be easily identified (visually or by listening to the tokens), the current data show that classification models only performed worse when the task at hand was equally challenging for a human transcriber: namely, sorting falling pitch tokens into steeper/higher falls vs. smoother/lower falls in the HØ.L context (§4.2). Another caveat is that classification models only yield interpretable results when trained on input data representative of the hypotheses. During the classification of HØ.L tokens, I used HL.L (containing lexical /HL/μμ tones) as a proxy for the L-insertion hypothesis, which might not be comparable to the HØ.L experimental tokens (a product of light-initial disyllabic tone sandhi) to begin with. Lastly, the current method utilizes stochastically sampled data assuming an Interpolation hypothesis. That is, it is capable of demonstrating what the f0 trajectories for interpolated tokens *would* look like, while also incorporating natural-like pitch perturbations. This becomes particularly important when raw data fitting the hypothesis (i.e., real phrases with 'apparent' pitch interpolation) are difficult to obtain.

In addition, I want to reiterate the benefit of evaluating pitch tokens individually rather than analyzing grand mean contours, a point also explicitly discussed by Kawahara et al. (2022: 115–116). A grand mean contour that averages all tokens from all speakers under a certain tonal context would conflate two possible outcomes: one where most Ø tokens are weakly-articulated Mid pitch (H4), and one where there is an equal distribution of Default L (H1) and Interpolation (H2) Ø tokens. **Figure 10** plots the distribution of classification results in Suzhou by showing two aggregate posterior probability histograms, one for HØ.H and another LØ.H.[30] An immediate observation is that the classification results, when pooled across all speakers, showed a clear bimodal pattern in both HØ.H and LØ.H contexts (regardless of the bias towards one end of the probability scale in both cases). It is reasonable to conclude that both HØ.H and LØ.H contained two *categorical* realizations — Default L (H1) and Interpolation (H2). Such a double realization pattern, however, might be overlooked if the analyst approaches the noisy phonetic data in the form of averaged trajectories. I explore this caveat further with regard to the Weakly-articulated hypothesis (H4) in the following subsection.
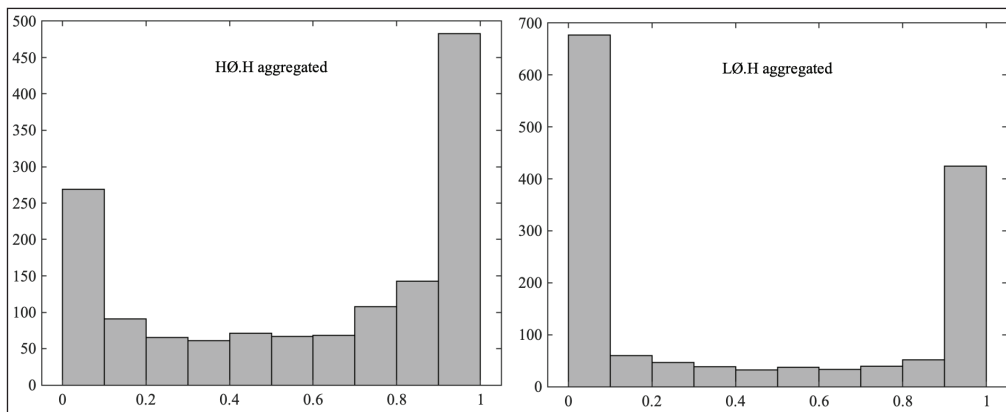
Figure 10: Posterior probability histograms for the HØ.H and LØ.H contexts, aggregated over all speakers.

5.2 Exploring the 'Weakly-articulated' hypothesis of toneless realization

This subsection further explores the stable target proposal by Y. Chen and colleagues. One of the main arguments of the 'Weakly-articulated' hypothesis is that a sequence of neutral tones in Mandarin, regardless of their left and right tonal contexts, would always converge on a pitch value characteristic of a 'mid' tone. Averaged over four speakers and twelve repetitions, trisyllabic neutral tone sequences in Mandarin would end around 150Hz in all tonal contexts, which was in the middle of a typical H (250Hz) and a L tone (120Hz) in their data (see Y. Chen & Xu 2006, Fig. 3). Interestingly, one would have reached a similar conclusion if plotting the Suzhou data as averaged trajectories. Consider the following line plot containing the averaged HØ.H, HØ.L and LØ.H trajectories of all speakers.

---

[30] The HØ.L context contained *three* categorical realizations (H1, H2 and H3) and is excluded.

Interpolation and H3/Spread Tone — identifiable by both Naive Bayes classification models and visual/manual inspection by the fieldworkers. Take the HØ.H histogram in **Figure 10** for example: the majority of HØ.H tokens belonged to either the 0–0.1 bin or the 0.9–1 bin on the posterior probability scale, indicating that Ø realized with either a clear low pitch or a clear high pitch, but rarely in between (i.e., around 0.5 on the probability scale). Following the proposal of Y. Chen & Xu (2006), I show a hypothetical distribution of weakly-articulated Ø in the HØ.H context, and contrast it with a 'bimodal' speaker in my study.
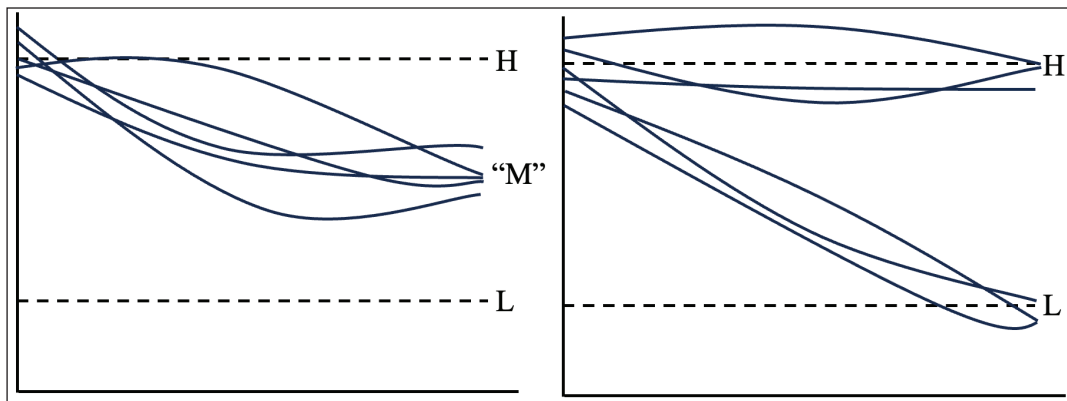

Figure 12: Hypothetical distributions of weakly-articulated realization (left) vs. bimodal realization (right) under the HØ.H context.

Shown on the left is a stable realization of HØ as 'HM' independent of the right context H, which is representative of H4 (Y. Chen & Xu 2006). The variation observed in the trajectory data does not correspond to phonologically distinctive categories, and is better treated as naturalistic 'noise' inherent to speech production. The right schematic in **Figure 12**, on the other hand, contains two clear categories: high falling pitch corresponding to a Default L insertion hypothesis (H1), high levels corresponding to either Linear Interpolation (H2) or Tone Spreading (H3). I term this pattern 'bimodal' realization due to the peaks at both ends in the corresponding Naive Bayes classification histogram (**Figure 3**). As discussed in §5.1, a critical drawback of analyzing noisy pitch data as means (e.g., ANOVA) is that it is difficult to tease apart the two scenarios in **Figure 12** solely from mean values. Both scenarios discussed above would have a fall to mid trajectory when averaged, given that there were comparable amounts of HL fall vs. high level tokens in the bimodal realization (see **Figure 10**, HØ.H aggregated). A machine classification analysis, on the other hand, can simultaneously assess both the realization of individual tokens and the overall distribution — a weakly-articulated realization would peak around 0.5 posterior probability (at least for the HØ.H context), representing the observation that many tokens could neither be classified as high levels nor as HL falls (also see Shaw & Kawahara 2018; Kawahara et al. 2022). I did not find an overall peak around 0.5 probability in the Suzhou data, rejecting H4.

Before ending the discussion, I do want to reiterate a crucial difference in elicitation items between Y. Chen & Xu (2006) and the current study, and to a lesser degree, the Suzhou data of Ling (2014). The target syllables in Chen and Xu's design were functional words of varying length. Likewise in Ling (2014), lexical words were all followed by a genitive particle [gə?] 葛. In contrast, the current study has examined tonelessness arising from tone sandhi in *lexical words*. As such, the current findings are not necessarily contradictory to those of Y. Chen & Xu (2006), in that there may as well be two types of tonelessness in Suzhou: functional words that consistently converge towards a low target, and lexical sandhi Ø that demonstrates variability. This possibility has been explicitly discussed by M. Zhang et al. (2019), who differentiate between 'lexically-toneless' syllables and 'contextually-reduced' ones in Mandarin.

5.3 Noisy/variable phonological grammars

This study has demonstrated that the mapping between toneless moras in Suzhou and their surface exponents can be one to many. Some discussion on how a phonological grammar can generate multiple surface forms is due. Here, I briefly describe one framework compatible with one-to-many realizations: variants of Noisy Harmonic Grammar.

Both Coetzee & Pater (2011) and Coetzee & Kawahara (2013) have extensively discussed phonological models that can generate variable outputs from the same input. Coetzee & Pater (2011) introduce one implementation of variable Optimality Theory grammars, Noisy Harmonic Grammar (Noisy HG; Smolensky & Legendre 2006), where constraints are assigned numerical values and the candidate with the highest aggregated constraint scores (called *Harmony*) is selected as the output. As its name suggests, variation in Noisy HG can result from the 'noise' that changes the numerical weights of some constraints every time the grammar is used. In addition to Noisy HG, it is also possible to transform a Harmonic Grammar into a probability distribution over the possible candidates in an adaptation named Maximum Entropy grammar (Coetzee & Pater 2011: 420–421). While the variable realizations are represented differently in the two types of HG, they are largely similar in the usage of continuous numerical constraint weights and learning mechanisms.[32]

Different versions of HG are able to handle variation sensitive to context (e.g., English variable t/d deletion dependent on dialect and phonological context). Meanwhile, the models discussed thus far seem to be inadequate to account for variation where the surface form distribution is word specific — different lexical items sharing the same phonological context within a single dialect may also demonstrate systematic differences in realization. Coetzee & Kawahara (2013) augment the noisy HG model by incorporating *word frequency* to capture the

---

[32] See Flemming (2021) for a detailed discussion of the differences between Noisy Harmonic Grammar and Maximum Entropy Grammar regarding the evaluation of constraint violations.

empirical observation that more frequent words, all else being equal, undergo reductive change to a greater extent. They propose to include a frequency-dependent scaling factor that changes the weight of faithfulness constraints, such that the OT grammar permits faithfulness violations for more frequent words, leading to more reduced tokens. This model brings together aspects of traditional generative grammars (e.g., OT) and recent developments in usage-based/exemplar models. Under such framework, it would also be viable to incorporate speaker-specific constraint scaling to account for different types of speakers in Suzhou (e.g. those who use a mix of toneless realizations vs. ones that overwhelmingly produce one single surface form).

5.4 Potential sources of variation: speaker age, sex and usage frequency

After observing the phonetic variation in the data, a natural follow-up question is what exactly conditions the variation. This section explores several factors that may have affected the variable realization of toneless moras in Suzhou.

One reviewer asked whether between-speaker variation could be largely attributed to speaker age and sex (Eckert 1989; 2017), since I have included males and females ranging from 30 to 70 at the time of the fieldwork. Although an in-depth look at the potential of an ongoing sound change (conditioned by speaker age and sex) is beyond the scope of the current paper, I present some preliminary data below regarding the issue (see also Zhu 2023b: 9–10 for similar discussions).

For ease of visualization and analysis, I choose to use the average posterior probability given by the Naive Bayes models to represent each speaker's realization pattern. For example, Speaker 02's HØ.H tokens, when submitted to a Naive Bayes classifier, yielded an average posterior probability of 0.76 (where a probability of 0 stands for Default L/H1, and 1 for Interpolation/ H2 or Spreading/H3; cf. **Figure 4a**). For the sake of simplicity, I refrain from discussing Spread Tone tokens and the HØ.L tonal context: Spread Tone tokens are conflated with other hypotheses in the HØ.H and LØ.H contexts, and the HØ.L context differentiates three types of realization but does not allow for a straightforward interpretation of the posterior probability (see also fn. 27). Below I show the scatter plots of averaged posterior probabilities by speaker age, sex and tonal contexts (HØ.H and LØ.H).

From visually inspecting **Figure 13**, one may observe that: (i). The HØ.H context overall corresponds to higher posterior probabilities; (ii). Females tend to have higher probabilities. That is, the HØ.H context included more Interpolation/H2 tokens (also discussed in §4.3), and females were more likely to use interpolation as their toneless realization. To tease apart the effect of age, sex and tonal context, I fit a mixed-effects regression model using the lme4 package in R (Baayen et al. 2008; Bates et al. 2014). I included Age (numerical), Sex (categorical) and Tonal Context (categorical) and their two-way interactions as fixed effects, along with by-speaker random intercepts. The model estimates are reported in **Table 3**.
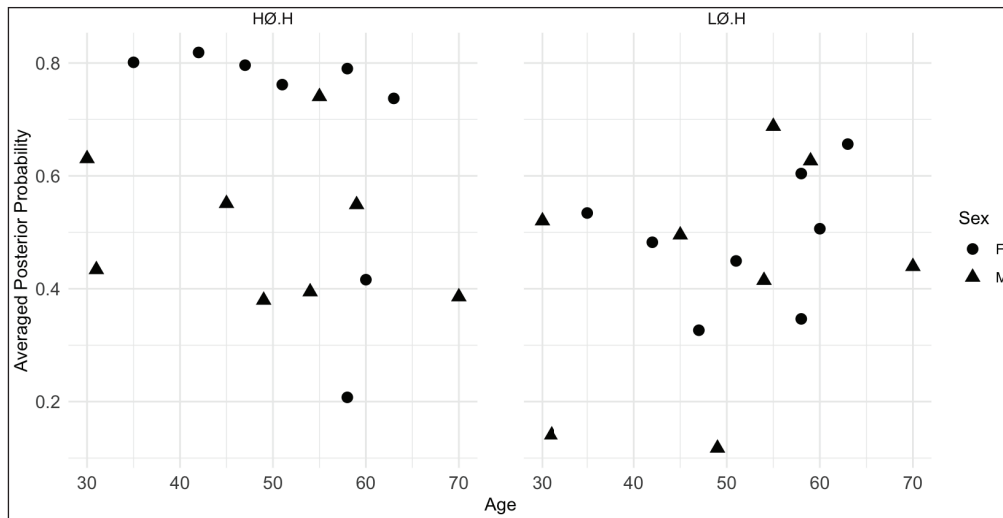
Figure 13: Averaged posterior probability values (y-axis) plotted against speaker age (x-axis). Different shapes stand for speaker sex and panels stand for tonal context.

Fixed effects	Estimate	Standard Error	t value
(Intercept)	1.14	0.34	3.39*
Age	–0.01	0.01	–1.44
Sex-M	–0.48	0.40	–1.20
Tonal Context-LØ.H	–0.68	0.17	–3.90*
Age*Sex-M	0.01	0.01	0.79
Age*Tonal Context-LØ.H	0.01	0.003	3.00*
Sex-M*Tonal Context-LØ.H	0.13	0.07	1.74

Table 3: Model results predicting posterior probability from Age, Sex and Tonal Context.

We may interpret the t-value informally by adopting the absolute value of 2 as a threshold (Baayen et al. 2008: 398). The model estimates show that while Tonal Context did have a significant effect on posterior probability (lower in the LØ.H context), Age and Sex both failed to reach significance. Notably, there was a significant interaction between Age and Tonal Context: the (overall) negative effect of Age on posterior probability was less extreme under the LØ.H context. In other words, between-sex difference in **Figure 13**, left panel (HØ.H) was neutralized to some extent on the right panel (LØ.H). That said, the potential influence of Age and Sex on toneless realization remains inconclusive. Future studies may adopt a more carefully controlled word list, ideally taking into account segmental/tonal contexts and word frequency, and obtain phonetic data from a larger group of speakers in order to determine the role of sociolinguistic factors in toneless realization. This also leads to the discussion of a second potential contributor to variation, word frequency.

Following the tenets of Coetzee & Kawahara (2013), it is possible to characterize the Interpolation (H2) realization as a form of tonal reduction/deletion: while remaining agnostic about the diachrony of toneless moras in Suzhou, it is conceivable that the speakers' synchronic grammar assigns a 'default' f0 value to the final mora in light-heavy disyllables, and optionally deletes/reduces the tonal specification of said mora. As variable reduction has been repeatedly shown to occur in more frequent words (Bybee 2001), we may expect that daily vocabularies in Suzhou (e.g., [ɦoᴸ.sã:ᴴᴼ.kø:ᴴᴸ.buːᴸ] 學生幹部 'student leadership') would realize more often with interpolation (H2) than ones that are formal, low-frequency compounds (e.g., [do.ᴸ.zɛ:ᴴᴼ.sənᴴ.mɪnᴴ] 獨裁聲明 'proclamation of dictatorship', both were target words in the current study). The following figure shows the averaged trajectories of these two phrases over all speakers and repetitions.

A preliminary conclusion from **Figure 14** is that the [HØ] syllable in a high-frequency phrase is overall more 'flat' in pitch compared to that of a low-frequency phrase, even when both phrases contain the same HØ.H tonal context and have identical morphosyntactic structure (i.e. disyllable-disyllable noun compounds). Translated into the competing hypotheses, one can argue that high-frequency phrases more often correspond to Interpolation/H2 while low-frequency ones realize with more Default L insertion/H1, which is in line with Coetzee and Kawahara's (2013) account of reductive variation and change. I am, however, unable to fully explore this observation, as there is no publicly-available spoken corpus of Suzhou Chinese to determine the relative frequency of all the tokens in this study. In order to continue the investigation of potential frequency effects, the next step is to establish a database representative of daily speech in Suzhou.
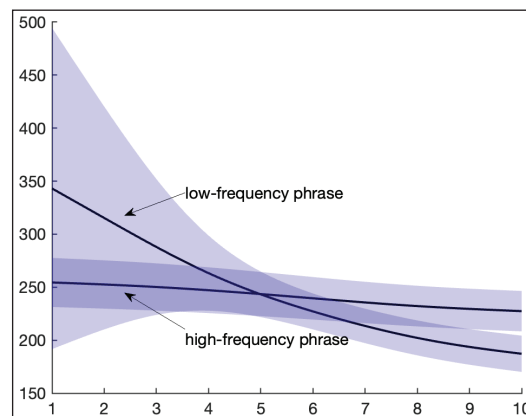

Figure 14: Smoothing spline trajectories of the [HØ] syllable for all tokens of a high-frequency phrase 'student leadership', and those of a low-frequency phrase 'proclamation of dictatorship' (both are HØ.H words).

6 Conclusion

This paper challenges the traditional belief in Chinese tonology literature that toneless syllables or moras in Chinese languages have a single, predetermined surface realization, be it a 'Default L' tone or an intermediate 'reduced' tone. Through original fieldwork and computational analysis

following Shaw & Kawahara (2018), I show that Suzhou Chinese realizes its toneless mora in multiple categorically distinct forms. Also noteworthy is the observation that while a toneless mora can realize without a tonal target throughout the production process (cf. 'targetless' in the Articulatory Phonological tradition), its pitch value is not randomly assigned (following some normal distribution), but generally fits a linear interpolation between the surrounding phonological tones (cf. Pierrehumbert 1980). The findings of this study contribute to our knowledge of both tonelessness and variable phonological processes.

I discuss several potential limitations of the current study and future directions. One central argument of the 'asymptotic approach' pitch model by Xu and colleagues (Xu 1999; Xu & Wang 2001) is that the realization of the stable mid toneless target is *duration dependent*: a toneless TBU is often also reduced in duration, and would not have enough time to fully realize the mid target. Consequently, toneless syllables or moras with relatively short duration may appear as if they were level tones. It is possible for example, that the bimodal realization in the HØ.H data was largely due to variation in the toneless mora duration. A more complete analysis would therefore benefit from statistical models testing the correlation between duration and pitch trajectory types. I refrain from discussing duration in this study due to the potential complication with speaker-specific speaking rate: for a speaker who produced more level-like tokens, it is unclear whether this was because they were speaking relatively fast (in relation to their own speaking rate, or to all Suzhou speakers in general), or they were preferring a certain toneless realization in particular.

In addition, the Toneless data set of the current study did not contain any LØ.L tokens (toneless mora surrounded by two L tones). The tonal context cannot discriminate between H1, H2 and H3 (all three hypotheses predict a low level Ø realization), but can instead act as a direct test against H4/Weakly-articulated: the LØ syllable would have to realize as a rise to mid pitch, regardless of the right context L, if the weakly-articulated hypothesis is true. I leave this as a follow-up investigation future studies may explore.

Lastly, despite the relatively high classification confidence of the Naive Bayes models, it remains unknown whether native speakers of Suzhou can perceive different toneless realizations (e.g., Default L vs. Interpolation). An informal probe during the fieldwork interviews suggests that Suzhou speakers were sensitive to phonologically contrastive tonal categories (e.g., /H/ μμ vs. /HL/μμ), as they would comment that a word was pronounced 'with the wrong tone' if a /HL/ tone morpheme was produced with a high level pitch. However, there was considerably less correction when, for instance, a HØ.H token was produced with either Default L Insertion or Interpolation. As such, we may predict that words with variably pitched toneless moras would lead to less confusion or identification error in a word identification task, when compared to words that are fully toned yet realize with similarly variable pitch. In other words, listeners might be more 'permissive' when perceiving toneless moras in Suzhou, precisely because tonally underspecified moras have considerable pitch variation in production as well.

Supplementary Files

The supplementary materials for this article including the full elicitation word list, all classification plots, the raw f0 data and all code can be found here: https://osf.io/nawk5/.

Acknowledgements

Ethics and consent

This research study was approved by the ethics committee of The Ohio State University. The identity of all participants is anonymized and I do not share any identifiable personal information of them. All participants have read and agreed to an informed consent form.

Competing interests

The author has no competing interests to declare.

References

Baayen, R. H. & Davidson, D. J. & Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. DOI: https://doi.org/10.1016/j.jml.2007.12.005

Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2014. Fitting linear mixed-effects models using lme4. DOI: https://doi.org/10.18637/jss.v067.i01

Bermúdez-Otero, Ricardo & Trousdale, Graeme. 2012. Cycles and continua: On unidirectionality and gradualness in language change. In *The oxford handbook of the history of english*, 691–720. Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199922765.013.0059

Boersma, Paul & Weenink, David. 2022. Praat: doing phonetics by computer. http://www.praat.org/.

Breteler, Jeroen. 2018. *A foot-based typology of tonal reassociation: perspectives from synchrony and learnability*: University of Amsterdam dissertation.

Breteler, Jeroen & Kager, René. 2022. Layered feet and syllable-integrity violations: The case of Copperbelt Bemba bounded tone spread. *Natural Language & Linguistic Theory* 40(3). 703–740. DOI: https://doi.org/10.1007/s11049-021-09514-1

Browman, Catherine P. & Goldstein, Louis. 1992. "Targetless" schwa: an articulatory analysis. In Docherty, Gerald J. & Ladd, D. Robert (eds.), *Papers in laboratory phonology ii: Gesture, segment, prosody*, 26–67. Cambridge University Press. https://www.cambridge.org/core/product/identifier/CBO9780511519918A011/type/book_part

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge University Press. https://www.cambridge.org/core/product/identifier/9780511612886/type/book

Chao, Yuenren. 1968. *A grammar of spoken Chinese*. University of California Press. DOI: https://doi.org/10.1017/CBO9780511486364

Chen, Matthew Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.

Chen, Yiya. 2008. Revisiting the Phonetics and Phonology of Shanghai Tone Sandhi. In *Proceedings of the fourth conference on speech prosody*, 253–256. DOI: https://doi.org/10.21437/SpeechProsody.2008-55

Chen, Yiya & Xu, Yi. 2006. Production of Weak Elements in Speech – Evidence from F0 Patterns of Neutral Tone in Standard Chinese. *Phonetica* 63(1). 47–75. DOI: https://doi.org/10.1159/000091406

Coetzee, Andries W. & Kawahara, Shigeto. 2013. Frequency biases in phonological variation. *Natural Language & Linguistic Theory* 31(1). 47–89. DOI: https://doi.org/10.1007/s11049-012-9179-z

Coetzee, Andries W. & Pater, Joe. 2011. The place of variation in phonological theory. In Goldsmith, John A. & Riggle, Jason & Yu, Alan C. (eds.), *The handbook of phonological theory*, 401–434. Blackwell Publishing Ltd. DOI: https://doi.org/10.1002/9781444343069.ch13

de Lacy, Paul. 2002. The interaction of tone and stress in Optimality Theory. *Phonology* 19(2002). 1–32. DOI: https://doi.org/10.1017/S0952675702004220

Dresher, Elan B. & van der Hulst, Harry. 1998. Head-Dependent asymmetries in phonology: Complexity and visibility. *Phonology* 15(3). 317–352. DOI: https://doi.org/10.1017/S0952675799003644

Duanmu, San. 1995. Metrical and tonal phonology of compounds in two Chinese dialects. *Language* 71(2). 225–259. DOI: https://doi.org/10.2307/416163

Duanmu, San. 2007. *The phonology of Standard Chinese*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780199215782.001.0001

Eckert, Penelope. 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1(3). 245–267. DOI: https://doi.org/10.1017/S095439450000017X

Eckert, Penelope. 2017. Age as a Sociolinguistic Variable. In *The handbook of sociolinguistics*, 151–167. DOI: https://doi.org/10.1002/9781405166256.ch9

Flemming, Edward. 2021. Comparing MaxEnt and noisy harmonic grammar. *Glossa: A Journal of General Linguistics* 6(1). DOI: https://doi.org/10.16995/glossa.5775

Goldsmith, John. 1976. *Autosegmental phonology*: MIT dissertation. DOI: https://doi.org/10.1016/B0-08-044854-2/04223-1

Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511616983

Guy, Gregory R. 1991. Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3(1). 1–22. DOI: https://doi.org/10.1017/S0954394500000429

Hayes, Bruce. 1995. *Metrical Stress Theory: Principles and Case Studies*. University of Chicago Press.

Hyman, Larry M. 2009. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences* 31(2–3). 213–238. DOI: https://doi.org/10.1016/j.langsci.2008.12.007

Iosad, Pavel. 2013. Head-dependent asymmetries in Munster Irish prosody. *Nordlyd* 40(1). 66–107. DOI: https://doi.org/10.7557/12.2502

Ito, Junko & Mester, Armin. 2019. Pitch accent and tonal alignment in Kagoshima Japanese. *The Linguistic Review* 36(1). 1–24. DOI: https://doi.org/10.1515/tlr-2018-2005

Kager, René. 1993. Alternatives to the Iambic-Trochaic Law. *NLLT* 11(3). 381–432. DOI: https://doi.org/10.1007/BF00993165

Kager, René & Martínez-Paricio, Violeta. 2018. Mora and syllable accentuation – Typology and representation. In *The study of word stress and accent – theories, methods and data*, 147–186. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/9781316683101.006

Karlin, Robin P. 2018. *Towards an articulatory model of tone: a cross-linguistic investigation*: Cornell University Phd dissertation.

Kawahara, Shigeto & Shaw, Jason A. & Ishihara, Shinichiro. 2022. Assessing the prosodic licensing of wh-in-situ in Japanese. *Natural Language & Linguistic Theory* 40(1). 103–122. DOI: https://doi.org/10.1007/s11049-021-09504-3

Kingston, John. 2007. Segmental influences on F0: Automatic or controlled? In *Tones and tunes*, 171–210. Berlin, Germany: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110207576.2.171

Kingston, John. 2011. Tonogenesis. In *The blackwell companion to phonology*, 2304–2333. Wiley. DOI: https://doi.org/10.1002/9781444335262.wbctp0097

Kirby, James P. 2018. Onset pitch perturbations and the cross-linguistic implementation of voicing: Evidence from tonal and non-tonal languages. *Journal of Phonetics* 71. 326–354. DOI: https://doi.org/10.1016/j.wocn.2018.09.009

Köhnlein, Björn. 2011. *Rule reversal revisited : synchrony and diachrony of tone and prosodic structure in the Franconian dialect of Arzbach*: University of Leiden dissertation. https://openaccess.leidenuniv.nl/handle/1887/17583.

Leben, William R. 1973. *Suprasegmental Phonology*. MIT dissertation.

Lee, Wai-Sum & Zee, Eric. 2008. Prosodic characteristics of the neutral tone in Beijing Mandarin. *Journal of Chinese Linguistics* 36(1), 1–29.

Ling, Feng. 2011. Pitch patterns of prosodic words in Suzhou Chinese. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011)* (August), 1258–1261.

Ling, Feng. 2014. 语流中苏州话连调的声学模式 *[The Acoustic Characteristics of Sandhi Tones in Suzhou Dialect in Connected Speech]*. *Bulletin of Linguistic Studies* 1. 179–188.

Liu, Roger Cheng-yen & Hsieh, Feng-fan & Chang, Yueh-chin. 2021. Targeted and targetless neutral tones in Taiwanese Southern Min. In *Interspeech 2021*, 2631–2635. DOI: https://doi.org/10.21437/Interspeech.2021-434

Luo, Qian & Durvasula, Karthik & Lin, Yen-Hwei. 2016. Inconsistent consonantal effects on F0 in Cantonese and Mandarin. In *5th international symposium on tonal aspects of languages*, 52–55. DOI: https://doi.org/10.21437/TAL.2016-11

Morén, Bruce & Zsiga, Elizabeth. 2006. The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory* 24(1). 113–178. DOI: https://doi.org/10.1007/s11049-004-5454-y

Myers, Scott. 1998. Surface underspecification of tone in Chichewa. *Phonology* 15. 367–391. DOI: https://doi.org/10.1017/S0952675799003620

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. MIT Phd dissertation. DOI: https://doi.org/10.1177/003368828401500113

Pierrehumbert, Janet B. & Beckman, Mary E. 1988. *Japanese Tone Structure*. MIT Press: Cambridge.

Prince, Alan. 1976. *Applying Stress*. Unpublished ms. University of Massachusetts: Amherst.

Prom-on, Santitham & Xu, Yi & Thipakorn, Bundit. 2009. Modeling tone and intonation in Mandarin and English as a process of target approximation. *The Journal of the Acoustical Society of America* 125(1). 405–424. DOI: https://doi.org/10.1121/1.3037222

Qian, Nairong. 1992. 當代吳語研究 *[Contemporary Wu Dialect Studies]*. Shanghai: Shanghai Education Press.

Qian, Nairong & Shi, Rujie. 1983. 苏州方言连读变调讨论之二 *[A Second Discussion on Suzhou Tone Sandhi]*. *Fangyan* (4). 275–296.

Remijsen, Bert. 2013. Tonal alignment is contrastive in falling contours in dinka. *Language* 89(2). 297–327. DOI: https://doi.org/10.1353/lan.2013.0023

Roberts, Brice David. 2020. *An autosegmental-metrical model of Shanghainese tone and intonation*. University of California, Los Angeles Doctoral dissertation.

Rose, Phil. 1990. Acoustics and phonology of complex tone sandhi: an analysis of disyllabic lexical tone sandhi in the Zhenhai variety of Wu Chinese. *Phonetica* 47(1–2). 1–35. DOI: https://doi.org/10.1159/000261850

Shaw, Jason A. & Kawahara, Shigeto. 2018. Assessing surface phonological specification through simulation and classification of phonetic trajectories. *Phonology* 35(3). 481–522. DOI: https://doi.org/10.1017/S0952675718000131

Shi, Menghui. 2020. *Consonant and lexical tone interaction: Evidence from two Chinese dialects*: Leiden University dissertation.

Shi, Xinyuan & Jiang, Ping. 2013. A prosodic account of tone sandhi in Suzhou Chinese. In *Proceedings of the 25th north american conference on chinese linguistics*.

Smolensky, Paul & Legendre, Géraldine. 2006. *The Harmonic Mind*. Cambridge, MA: MIT Press.

Takahashi, Yasunori. 2019. The phonological status of Low tones in Shanghai tone sandhi. *Language and Linguistics* 20(1). 15–45. DOI: https://doi.org/10.1075/lali.00028.tak

Wang, Jialing. 1997. The representation of the neutral tone in Chinese Putonghua. In Wang, Jialing & Smith, Norval (eds.), *Studies in chinese phonology*, 157–184. DE GRUYTER. DOI: https://doi.org/10.1515/9783110822014.157

Wang, Ping. 2011. 苏州方言研究 *[Suzhou Dialect Studies]*. Beijing: Zhonghua Shuju.

Whalen, Douglas H. & Levitt, Andrea G. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23(3). 349–366. DOI: https://doi.org/10.1016/S0095-4470(95)80165-0

Xie, Zili. 1982. 苏州方言两字组的连续变调 *[Tone Sandhi of Bi-characters in Suzhou]*. *Fangyan* 3. 117.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0contours. *Journal of Phonetics* 27(1). 55–105. DOI: https://doi.org/10.1006/jpho.1999.0086

Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3–4). 220–251. DOI: https://doi.org/10.1016/j.specom.2005.02.014

Xu, Yi. 2013. ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In *Proceedings of tools and resources for the analysis of speech prosody (trasp 2013)*, 7–10. Aix-en-Provence, France.

Xu, Yi & Prom-on, Santitham & Liu, Fang. 2022. The PENTA Model: Concepts, Use, and Implications. In Barnes, Jonathan & Shattuck-Hufnagel, Stefanie (eds.), *Prosodic theory and practice*, chap. 11, 377–407. The MIT Press. DOI: https://doi.org/10.7551/mitpress/10413.003.0014

Xu, Yi & Wang, Emily Q. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33(4). 319–337. DOI: https://doi.org/10.1016/S0167-6393(00)00063-7

Ye, Xiangling. 1993. 蘇州方言詞典 *[Suzhou Dialect Dictionary]*. Nanjing: Jiangsu Education Press.

Yip, Moira. 1980. *The Tonal Phonology of Chinese*: Ph.D. Dissertation, MIT. dissertation. DOI: https://doi.org/10.3406/clao.1980.1072

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Yu, Bingqing. 2022. *The IF0 effect in the Hong Kong Cantonese tone system*. MA Thesis, Simon Fraser University dissertation.

Yue-Hashimoto, Anne. 1987. Tone sandhi across Chinese dialects. In *Wang li memorial volumes: English volume*, 445–474. Chinese Language Society of Hong Kong.

Zee, Eric & Maddieson, Ian. 1979. Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. In *Ucla working papers in linguistics* 45. 93–129.

Zhang, Hongming. 2016. *Syntax-Phonology Interface: Argumentation from Tone Sandhi in Chinese Dialects*. New York: Routledge. DOI: https://doi.org/10.4324/9781317389019

Zhang, Jie. 2002. *The Effects of Duration and Sonority on Contour Tone Distribution: A Typological Survey and Formal Analysis*. University of California, Los Angeles dissertation. DOI: https://doi.org/https://doi.org/doi:10.7282/T3RX99XG

Zhang, Muye & Geissler, Christopher & Shaw, Jason. 2019. Gestural Representations of Tone in Mandarin : Evidence From Timing Alternations. In *International congress of phonetic sciences icphs 2019*, 1803–1807.

Zhang, Yixin. 2021. *Neutral Tone in Mandarin: Representation and Interaction with Utterance-level Prosody*. University of Cambridge Phd dissertation.

Zhu, Yuhong. 2023a. A metrical analysis of light-initial tone sandhi in Suzhou Wu. *Natural Language & Linguistic Theory* 41(4). 1629–1678. DOI: https://doi.org/10.1007/s11049-023-09572-7

Zhu, Yuhong. 2023b. Variable Pitch Realization of Unparsed Moras in Suzhou Chinese: Evaluation Through F0 Trajectory Simulation and Classification. In *Proceedings of the 2022 annual meetings on phonology*. DOI: https://doi.org/10.3765/amp.v10i0.5424

Zhu, Yuhong. 2023c. *Tone, Metrical Structure and Intonation in Suzhou Chinese: Data, Theory, Typological Implications*: PhD. Dissertation, The Ohio State University dissertation.