Variable Pitch Realization of Unparsed Moras in Suzhou Chinese: Evaluation Through F0 Trajectory Simulation and Classification

Yuhong Zhu
The Ohio State University

1 Introduction

This study investigates the phonetic realization of metrically unparsed, phonologically toneless moras in Suzhou Chinese (Northern Wu). Counter to previous tonal analyses that either assume a 'default tone' realization of toneless Tone-Bearing Units (TBUs) or argue for a linear phonetic interpolation through tonelessness, I demonstrate that toneless moras in Suzhou show *both* default L tone insertion and linear interpolation, with considerable cross-speaker variation. The results are in line with a phonological model with optional/probabilistic instead of categorical processes (Coetzee & Pater, 2011; Coetzee & Kawahara, 2013), and are comparable to the Mandarin neutral tone data in Zhang et al. (2019). Furthermore, the toneless data of four speakers point to a possible split conditioned by biological sex, as female speakers more often choose the interpolation interpretation and male speakers overwhelmingly prefer default L insertion.

Tonelessness, the status of a TBU without an associated phonological tone, has made its appearance in numerous tonal and intonational accounts (Pierrehumbert & Beckman, 1988; Chen, 2000; Yip, 2002). Despite its popularity as an analytical tool, phonologists and phoneticians have not reached a consensus on its precise phonetic realization. A toneless TBU cannot be simply without any pitch/f0, as long as it is not completely devoiced. As Yip (2002:11) aptly puts it: 'They are, nonetheless, pronounced with some particular pitch'.

Two major proposals are available when it comes to interpreting tonelessness: toneless TBUs are sometimes assigned a predetermined (often L) default tone as 'the last act of phonology' Yip (2002:64). Default tones are often situated in the discussion of tonal underspecification, offering a toneless TBU the least marked value when phonological processes such as tonal spreading are unavailable (see, for instance, the analyses of Margi and Yoruba in Pulleyblank 1986). The choice of default tone/pitch height is language-specific; it can be an exclusive surface/phonetic tone (e.g. a /H/ vs. /L/ language with default M), or may be phonetically identical to an underlying tone (e.g. a /H/ vs. /L/ language with default L). In the latter case, the only indication for an inserted default tone is its phonological inactiveness — inserted tones do not participate in any phonological processes.

The alternative explanation stems from early works on intonation (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988) and argues for the possibility of a string of TBUs to be sparsely toned even on the surface. F0 value of the toneless TBUs is then determined by an interpolation rule that refers to surrounding phonological tones. Compared to the default tone approach, a major difference the interpolation analysis predicts is that toneless TBUs should assume contexually variable pitch, such that the (interpolated) F0 value of a toneless TBU between two Hs should be quantitatively distinct from that of a toneless TBU between a H and a L.

In the current study, I present a way of quantitatively assessing tonelessness in Suzhou Chinese, adopting the methodology of Shaw & Gafos (2015); Shaw & Kawahara (2018) and Zhang et al. (2019). The simulation & classification paradigm of Shaw& Gafos and Shaw & Kawahara not only generates robust token-by-token interpretation of tonelessness, but also provides by-speaker and by-context aggregates that warrant further sociolinguistic investigation. The paper is structured as follows: §2 summarizes phonological arguments for tonelessness in Suzhou Chinese, as well as previous work on quantitatively assessing tonelessness and more generally, 'targetlessness' (Shaw & Kawahara, 2018). In §3 and §4 I present the methodology and results for

[*] I am grateful for the invaluable input from Jason Shaw during the initial stage of this project. I have also benefited from helpful comments and suggestions from the Phonies discussion group at The Ohio State University. The usual disclaimer applies.

Proceedings of *AMP 2022*

the simulation and classification studies respectively. §5 discusses possible phonological and sociolinguistic factors that condition different interpretations of tonelessness. §6 concludes the paper.

2 Background

2.1 *Tonelessness in Suzhou Chinese* Suzhou Chinese or Suzhou Wu is a dialect of Northern Wu Chinese spoken in Suzhou City, Jiangsu Province, China. There are over ten million native speakers, but the amount of speakers and their proficiency are diminishing due to the pressure of language standardization policies (Yu, 2010). Below in Table 1 I show underlying representation of the seven lexical tones in Suzhou, along with their traditional name among Chinese linguists and example words.

Table 1: Lexical tones of Suzhou Chinese

Representation	Traditional name	Example	Gloss
/H/μμ	*yinping*	[ti:^{H}]	'low'
/LH/μμ	*yangping*	[di:^{LH}]	'carry'
/LHL/μμ	*yangshang/yangqu*	[ti:^{LHL}]	'ground'
/HL/μμ	*yinshang*	[ti:^{HL}]	'bottom'
/HLH/μμ	*yinqu*	[ti:^{HLH}]	'emperor'
/H/μ	*yinru*	[tɪ^{H}ʔ]	'trickle'
/LH/μ	*yangru*	[dɪ^{LH}ʔ]	'flute'

The last two lexical tones in Table 1 are traditionally referred to as *ru* or 'checked' tones, and synchronically surface as monomoraic/short vowel syllables with (non-moraic) glottal stop codas (Zhu 2020 for phonetic data). I will henceforth refer to them as *light* tones.

One central piece of argument for tonelessness in Suzhou comes from disyllabic tone sandhi patterns with initial light tones, or *light-initial* sandhi (Zhu, 2021a,b, forthcoming). Light-heavy disyllables in Suzhou contain a third/final mora that is invariably low-pitched in isolation. The low pitch is attributed to a phrase-final L% boundary tone, instead of restrictions on non-initial (non-foot-head) H tones. (de Lacy, 2002). A concise example pair demonstrating this point is given below:

(1) Toneless final/third mora in Suzhou light-heavy disyllables. Heavy-heavy disyllable with identical tonal material shown for comparison.

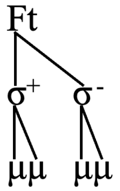
a. /LH/μ+ /HH/μμ → [Lμ.HμØμ]. Example: [ba^{L}.ho:^{HØ}] 'white flower'
b. /LH/μμ+ /HH/μμ → [Lμμ.Hμμ]. Example: [mɛ:^{L}.ho:^{H}] 'plum flower'

In (1) I show a pair of Suzhou disyllables with identical underlying tones but different weight profiles (light-heavy vs. heavy-heavy). Crucially, the final/third mora in (1a) is invariably low-pitched in isolation, while the entire second syllable (third/fourth moras) of (1b) is associated to a H tone by disyllabic sandhi. The L% boundary tone in the latter case does not override the H phonological tone. In short, the third mora of a light-heavy disyllable cannot be associated to a phonological tone, while there is no restriction on third/fourth moras with the heavy-heavy counterpart.

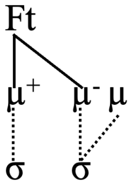
An unparsed TBU is often unable to license a phonological tone: studies on tone-foot interaction such as de Lacy (2002) and Breteler (2018); Breteler & Kager (2022) have proposed licensing constraints that penalize the linking between an unfooted/unparsed TBU (syllable/mora) and an autosegmental tone. The same generalization holds for Suzhou: while there is sufficient evidence from tone sandhi to believe that footing in Suzhou is trochaic (Duanmu 1995; Shi & Jiang 2013; Zhu forthcoming), a syllabic trochee leads to a cross-linguistically dispreferred parse of the light-heavy disyllables — a light foot head with a heavy foot dependent (Head-Dependent Asymmetry; Dresher & van der Hulst 1998; Iosad 2013). Instead, Suzhou non-exhaustively parses a light-heavy disyllable with a bimoraic trochee built directly on moras (Kager, 1993; Kager & Martínez-Paricio, 2018; Breteler, 2018). The last/third mora is not parsed into the bimoraic foot, and therefore unable to host a phonological tone. Footing in Suzhou is essentially quantity-sensitive, with syllabic trochees as the 'default' parse and moraic trochees for light-heavy disyllables. This is shown in (2):

(2) Quantity-sensitive footing in Suzhou Chinese. Ft = Foot; Superscript +/- = head/dependent

a. Heavy-heavy (syllabic trochee)



b. Light-heavy (moraic trochee)[1]



Both the phonetic data of light-initial sandhi and the phonological argument for quantity-sensitive footing point to the existence of toneless moras in Suzhou. There is, however, one remaining issue with the phonetic realization of the toneless mora: with regard to the two possible interpretations in §1, the light-heavy disyllable data do not differentiate between a default L tone interpretation and a interpolation (to L%) interpretation. That is, the invariable low pitch may result from either an inserted (and phonologically non-active) L tone at the phonology-phonetics interface, or an interpretation to a boundary L%. To tease apart these two possibilities, one would need to quantitatively assess the f0 trajectories of toneless moras when surrounded by different phonological tones (e.g. HØH, HØL). I summarize previous work on evaluating 'targetless'/toneless phonological units in the next subsection.

2.2 *Quantitatively evaluating tonelessness* The methodology of the current study is based on the Japanese devoiced vowel study of Shaw & Kawahara (2018). Similar to the alternative interpretations of tonelessness, there also have been debates on whether the devoiced vowels in Japanese are fully-articulated vowels without voicing (H1), or simply 'targetless' vowels without any gestural specification (H2). In addition to the hypotheses above, Shaw and Kawahara have also identified two other possibilities — as phonological processes may be optional (Coetzee & Pater, 2011), both targeted and targetless vowels may appear over a large amount of speech data (H3); the devoiced vowels might also be *neither* targeted nor targetless, but instead have a 'reduced' articulatory gesture in between a full vowel and a targetless one (H4).

To tease apart the four hypotheses, Shaw and Kawahara proposed two computational studies that offer valuable insight on the status of Japanese devoiced vowels. They collected two types of articulography data, words with fully-articulated vowels (e.g. /ɸuzoku/) and ones with the devoiced vowels in question (e.g. /ɸu̥ zoku/). In a stochastic simulation study, the authors sampled articulatory trajectory data using (i). a straight interpolation line connecting the preceding and following vowels (representing the surrounding gestural targets); (ii). naturally-occurring variability observed in the full vowel and devoiced vowel data. The results were simulated 'interpolation' trajectories that resembled naturalistic data in their variability. The authors then carried out pairwise comparisons using MANOVA to determine whether the devoiced vowel trajectories were significantly different from (i). fully-articulated vowels; (ii). simulated interpolations. While most pairwise comparisons confirmed that devoiced vowels were distinct from both full vowels and interpolations, several speakers' devoiced vowels were not significantly different from full vowels (possibility of H1) or interpolations (possibility of H2).

In order to acquire robust, token-by-token estimates of the devoiced vowel trajectories, Shaw and Kawahara also conducted a classification study. They trained a Naive Bayes classification model using the full vowel vs. simulated interpolation data, then fitted the devoiced vowel data to the classifier. The classifier gave probability estimates of each devoiced vowel token, which could then be aggregated by speaker to determine each speaker's strategy of realizing the devoiced vowels. The results indicated that speakers indeed differed in their ways of articulating Japanese devoiced vowels, with some favoring fully-articulated vowels no different from their voiced counterparts (H1), and others mostly showing interpolated articulation, with no intended vowel target in the devoiced vowel position (H2). Crucially, there was within-speaker variation of devoiced vowel trajectories to some level: several speakers had both types of tokens according to the classifier (H3).

The simulation & classification paradigm has since been used to address the issue of tonelessness as well. In Zhang et al. (2019), the authors investigated the phonetic realization of contexually-toneless and underlyingly-toneless syllables in Mandarin Chinese. Applying similar computational methods, the study found that: (i). contexually-toneless tokens were qualitatively different from underlyingly-toneless ones, as

[1] The dotted lines represent the fact that the prosodic constituents are not in a strictly-layered fashion (cf. Selkirk 1986), but words are nevertheless syllabified.

the former largely consisted of fully-toned articulations while the latter truly toneless interpolations; (ii). there was considerable cross-speaker variation, as some speakers used more interpolation even in the contextually-toneless category.

Bearing in mind these two computational studies on targetless vowels and toneless syllables, I will report the methods and results of a simulation study and a classification study on Suzhou toneless moras. Similar to Shaw & Kawahara (2018), there are four hypotheses regarding the realization of toneless moras in Suzhou. They are:

(3) Competing hypotheses regarding toneless moras in Suzhou

a. The toneless mora is assigned a default L tone on the surface. It does not demonstrate pitch variability beyond the degree of that of a phonological /L/. (H1)

b. The toneless mora assumes interpolated and *variable* pitch in different tonal contexts: for instance, high level between two H, rising between L and H. (H2)

c. The toneless mora is *optionally* L-toned, and *optionally* assumes interpolation. It demonstrates interpolated pitch in some cases, and low pitch in others. (H3)

d. The toneless mora realizes as a 'reduced' L tone. It is relatively low-pitched, yet distinct from phonological /L/. (H4)

3 Simulation study

3.1 *Methods* The simulation study compares three sets of f0 trajectory data: toneless moras Ø, phonological /L/ tones, and simulated interpolations. The toneless Ø and /L/ trajectories come from elicitation data in my fieldwork, while the interpolation data are stochastically sampled (see below for detailed methods). In order to tease apart the hypotheses in (3), the toneless mora trajectories need to be surrounded by different phonological tones (e.g. HØH, HØL, see §2.1). Trisyllabic and quadrisyllabic noun phrases with initial light-heavy disyllables provide such a context: for instance, a light-heavy-heavy-heavy noun phrase [ba^{L}.ho:^{HØ}.koŋ^{H}.ɹø^{H}] 'white flower park' corresponds to the tonal context HØH (Ø surrounded by two H tones). Each toneless condition is then paired with a corresponding L-toned condition (e.g. HØH vs. HLH), elicited from heavy-initial trisyllabic and quadrisyllabic phrases. There are in total three tonal contexts: HØH vs. HLH, HØL vs. HLL and LØH vs. LLH (LØL does not differentiate between default L tone and interpolation hypotheses, and is thus excluded). Below in (4) I give a pair of example words for each tonal contexts:

(4) Tonal contexts with examples. Crucial comparisons are in italics.

a. HØH vs. HLH
[ba^{L}.*ho:*^{HØ}.*sæ:*^{H}.t^{h}ã:^{H}] 'white shrimp soup' vs. [sɛ:^{H}.*tɕiø*^{HL}.*poʔ*^{H}] 'three-nine-eight'

b. HØL vs. HLL
[ɦo^{L}.*sã:*^{HØ}.*wɛ:*^{L}.tsã:^{H}] 'student organization leader' vs. [sæ:^{H}.*tsɨ:*^{HL}.*ts^{h}*ɛ:^{L}] 'cooked kelp'

c. LØH vs. LLH
[p^{j}ɛ^{H}.*tɕ^{h}in*^{LØ}.*tsø:*^{HL}.tsoŋ^{L}] 'badly beaten (idiom)' vs. [sɨ:^{HL}.*jã:*^{L}.*k^{w}a:*^{H}.tɕ^{h}i:^{H}] 'lazy and unwilling (idiom)'

The crucial pitch comparisons take place in the second syllable and the initial mora of the third syllable. For ease of segmentation, pitch trajectories of the second syllable were submitted for further analysis (containing the preceding tonal context + /L/ or Ø mora), while trajectories of the third syllable (containing the following tonal context) were omitted[2].

The elicitation word list consists of 33 HØH words, 20 HØL words and 36 LØH words for the toneless set. For the /L/ tone set, there are 23 LLH words and 33 /HL/-toned words followed by either a H or a L[3]. Each word is repeated three times in a psudeo-randomnized list. Four native speakers of Suzhou Chinese (two males, two females; age range 31-58, with an average of 47) were recruited to read each word in a carrier

[2] Note that the third syllable still provides a basis for one of the interpolation target tones. See below for discussion.

[3] Due to the tone sandhi patterns of Suzhou, [HL.H] words are rather rare, only appearing in certain verb-object phrases. As the /HL/ lexical tone has the same falling pitch realization regardless of its following tone, I have collapsed words with [HL.L] and [HL.H] sequences.

phrase [kã:^{HL} _____ bə^{H}.ŋəu^{LH}.t^{h}ɪn^{H}] (‘say _____ for me’). Due to travel restrictions, the elicitation took place remotely as each speaker was instructed to use their smartphone to record in a relatively quiet room.

The elicitation recordings were annotated using Praat (Boersma & Weenink, 2022) and relevant pitch trajectories extracted using the ProsodyPro script, with each trajectory containing ten time-normalized f0 data points (Xu, 2013). Tracking errors (doubling/halving) were hand-corrected. In order to compress the data for further statistical analysis (also to acquire linguistically-meaningful coefficient measurements; see discussion in Shaw & Kawahara 2018), I transformed the ten-point trajectory data using Discrete Cosine Transform (DCT), keeping the first three DCT coefficients of each pitch trajectory as they accounted for more than 99% of the variation in the data (by Pearson coefficient *r*).

The simulated/interpolation data set was based on the toneless Ø data. First, pitch targets for the preceding/following phonological tones were acquired by averaging the maximum pitch for all H tones and minimum for all L tones: for instance, for each HØL trajectory, the maximum pitch in the second syllables and minimum in the third syllables were extracted. Pitch targets for the HØL tonal context were then calculated from averaging these maximum/minimum values. After acquiring the averaged pitch targets for each tonal context, I fitted a straight line connecting the targets, representing the interpolation between preceding/following tones. The resulting straight lines were also transformed into three DCT coefficients. To simulate pitch variation observed in naturalistic speech for this set, I fitted standard deviations of the toneless Ø data to the straight interpolation lines, sampling interpolated pitch trajectories assuming normal distributions[4]. In short, the resulting simulated pitch trajectories were sampled combining straight-line interpolations connecting the preceding/following tones, and naturally-occurring variance from the raw speech data. In anticipation for the following classification study, I simulated the same amount of interpolation trajectories to that of Ø trajectories. Similar to the toneless and /L/ tone data sets, I generated the first three DCT coefficients to represent the simulated/interpolation data set. Below in Table 2 I list the summary statistics for one speaker and one tonal context as an example. Note that there were in total 4 speakers x 3 tonal contexts = 12 sets of Ø vs. /L/ vs. simulated data.

Table 2: Summary statistics for Speaker01, HØH vs. HL vs. simulated H-H condition

		1st DCT	**2nd DCT**	**3rd DCT**
HØH	Mean	528.56	22.82	0.18
	SD	42.32	22.03	3.66
HL	Mean	567.71	48.68	-2.08
	SD	49.06	24.04	4.04
Simulated H-H	Mean	577.42	-8.37	0
	SD	42.32	22.03	3.66

Standard deviation values for the ‘Interpolation H-H’ row is identical to that of ‘HØH’, representing the fact that the same level of variability is applied to the simulated interpolation data. Once DCT coefficients of the interpolation data set were sampled, they were transformed back to ten-point f0 values using inverse DCT (Shaw & Kawahara, 2018). As a result, we arrive at three sets of f0 trajectory data, two elicited from natural speech and one stochastically simulated. In addition to assessing the three data sets per speaker per tonal context by visualizing the f0 trajectories, I also conducted MANOVA tests, making pairwise comparisons between toneless Ø vs. /L/ tone, toneless Ø vs. interpolation and /L/ tone vs. interpolation groups. The MANOVA tests would indicate for each speaker and each tonal context, whether any of the two groups above were significantly different from each other. Since there are 4 speakers x 3 tonal contexts x 3 pairwise comparisons, the alpha value was Bonferroni-corrected to 0.05/36 = 0.000138. Coming back to the competing hypotheses in (3), each hypothesis corresponds to a set of predictions for each tonal context. They are:

(5) Predictions of the competing hypotheses

H1: Ø realizes as a low pitch target in all three tonal contexts; HØ trajectories are similar to HL, LØ similar to LL

[4] I chose normal distributions as the raw data DCT coefficients did not significantly differ from normality, confirmed by Shapiro-Wilk tests.

H2: Ø realizes as high level pitch in HØH, high falling pitch in HØL, and low rising pitch in LØH; Ø trajectories should be similar to simulated interpolation trajectories
H3: Ø are optionally low-pitched (H1) and optionally interpolation (H2)
H4: Ø realizes as a 'reduced' low pitch target; HØ trajectories are significantly different from HL, and LØ different from LL

3.2 *Results* I start by presenting MANOVA results of the pairwise comparisons (visualization of the pitch trajectories will be presented along with classification results; see §4.2). In order to ensure replicable results that are not dependent on a particular simulated data set, I report average MANOVA statistics over 10 simulations.

Table 3: MANOVA results for the three tonal conditions: HØH, HØL and LØH

	Speaker01			**Speaker02**			**Speaker03**			**Speaker04**		
	df	λ	p	df	λ	p	df	λ	p	df	λ	p
HØH vs. HL	193	0.67	*	132	0.40	*	144	0.44	*	155	0.75	*
HØH vs. Simulated H-H	189	0.54	*	167	0.76	*	169	0.88	**n.s.**	203	0.51	*
HL vs. Simulated H-H	193	0.37	*	132	0.33	*	144	0.41	*	155	0.46	*
HØL vs. HL	174	0.90	**n.s.**	132	0.81	*	122	0.80	*	125	0.73	*
HØL vs. Simulated H-L	151	0.59	*	98	0.42	*	125	0.38	*	143	0.72	*
HL vs. Simulated H-L	174	0.72	*	132	0.53	*	122	0.30	*	125	0.40	*
LØH vs. LL	126	0.65	*	120	0.59	*	114	0.75	*	134	0.65	*
LØH vs. Simulated L-H	143	0.28	*	169	0.12	*	177	0.46	*	199	0.13	*
LL vs. Simulated L-H	126	0.18	*	120	0.09	*	114	0.37	*	134	0.16	*

The degrees of freedom varied across each pairwise comparison mainly due to the fact that different numbers of word tokens were excluded from each speaker recording (due to background noise, poor recording quality, excessive creaky voice, etc.). Shown in Table 3, each significant pairwise comparison indicates that the two sets of DCT values for said speaker are significantly different. Similar to Shaw & Kawahara (2018), we can reject H1 if the Ø vs. /L/ pairwise comparison is significant, and reject H2 if the Ø vs. Simulated comparison is significant. If both comparisons are significant for a speaker-tonal context combination, we are left with H3 or H4 as the two possibilities: toneless moras categorically take the form of *either* /L/ or simulated interpolation, or are distinct from *both* /L/ and interpolation.

The two crucial pairwise comparisons were significant for most speaker-tonal context combinations, allowing us to reject both H1 and H2 for most cases. Noticeably, there were two non-significant comparisons by the Bonferroni-correted alpha values: for Speaker01, the HØL vs. HL comparison had a p-value of 0.00028, while the HØH vs. Simulated H-H comparison for Speaker03 had a p-value of 0.00047. Although a lack of significance does not immediately translate to similarity between groups (cf. Shaw & Kawahara 2018: 506), it is possible that Speaker01's toneless articulation between H and L is similar to a falling /HL/ tone (default L insertion, H1), while Speaker03's toneless articulation between two H is a high-level pitch resembling Simulated H-H (interpolation, H2).

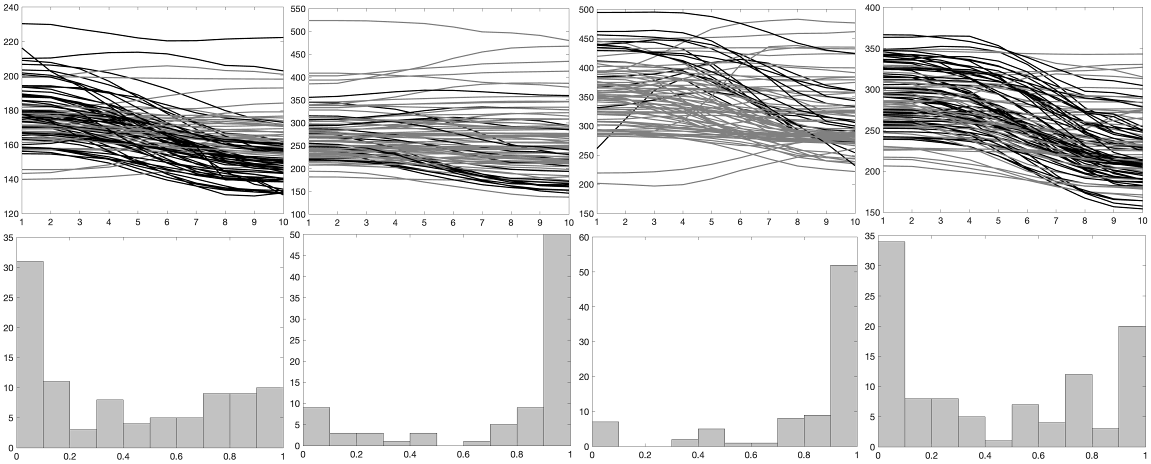
Significance over most of the MANOVA pairwise comparisons indicates that speakers neither uniformly used default /L/ tone insertion nor kept the Ø mora toneless throughout and interpolated in most tonal contexts. In other words, there was no uniform realization of tonelessness that held over different speakers or tonal contexts in Suzhou Chinese. A caveat is that since the test was performed over groups of multiple variables, it is not possible to tease apart H3 (optional) and H4 (reduced) with the MANOVA analysis: we are unable to interpret specific toneless tokens from the MANOVA statistics. I address the individual realization of tonelessness in the classification study below.

4 Classification study

4.1 *Methods* The classification study used the same elicitation and simulation pitch data described in §3.1. The aim of the classification study is to ascertain whether realization of tonelessness in Suzhou is optional/probabilistic between default /L/ insertion and interpolation. In order to acquire token-by-token estimates of the toneless moras, I trained a Naive Bayes classifier with the /L/ tone set and the simulated interpolation set as the input. For instance, to determine the realization of Ø in HØH context for a particular speaker, HL trajectories and simulated H-H ones for that speaker were submitted to the Naive Bayes classifier, with a 70% : 30% train/test split. After training, the model would be able give posterior probability estimates on how likely each Ø token is L-toned (similar to HL) or linear interpolation (similar to H-H). The individual probability estimates were represented as assigned labels ('/L/' vs. 'interpolation', whichever option had higher assigned probability). In addition, overall realization strategy of each speaker-tonal context combination can be seen from histograms of the posterior probabilities[5]. For H1 to be true, the histogram should have a peak in /L/ tone probability; for H2 to be true, the histogram should have a peak in the opposite end (interpolation); if the optional (H3) hypothesis is true, the histogram should be bi-modal at both ends; if the reduced (H4) hypothesis is true, the histogram should center around 0.5 probability, as the realization would resemble neither /L/ nor interpolation, and the classifier would give indeterminate 50%-50% estimates.

4.2 *Results* Below I present f0 plots for the toneless moras with labels assigned by the Naive Bayes classifiers, along with the posterior probability histograms, separated by tonal context. Figure 1 gives the relevant plots for the HØH context. Revisiting the predictions in (5), a default /L/ realization of Ø in HØH would be a falling pitch in the extracted second syllable, while an interpolation realization would connect the two H tones and resemble a high level pitch.

Figure 1: HØH classification plots and posterior probability histograms, Speakers 01-04. Gray lines stand for HØH trajectories classified as interpolation, black for /L/. x-axis for the histogram stands for posterior probability of interpolation articulation determined by the Naive Bayes classifier.

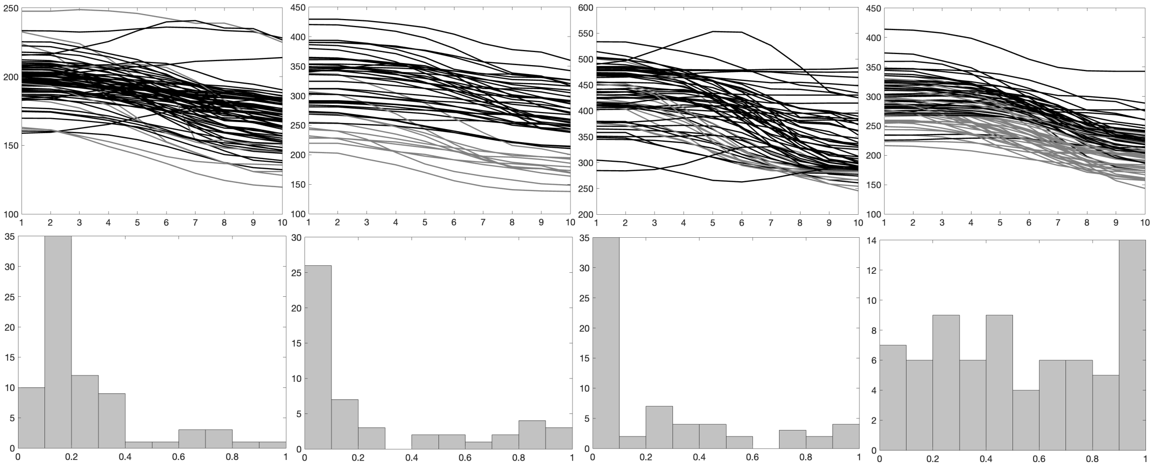


In Figure 1, each classification plot of the HØH data is paired with a histogram of posterior probability of interpolation. That is, higher concentration of tokens near 0 probability represents more tokens classified as having a default /L/ tone, and more black lines in the classification plot. Conversely, peaks at 1 probability indicates that more tokens are classified as interpolation, corresponding to more gray lines in the classification plot. As shown in both the trajectory plots and respective histograms, all four speakers included a mix of both /L/ tone and interpolation in their realizations of HØH, while speaker-specific preference was also easily observable: Speakers 01 and 04 favored /L/ insertion overall, indicated by a majority of falling pitch in the [HØ] syllable (black lines); Speakers 02 and 03, on the other hand, realized the [HØ] syllable mostly as high level pitch or even slightly raising pitch (gray lines), indicating an interpolation between tonal targets. Although the phonetic realization of tonelessness was never clear-cut like a categorical phonological processes, we could still conclude that each speaker had a 'preferred' realization strategy, be it /L/ tone or interpolation. The classification results in a sense also corroborated with the MANOVA statistics in §3.2: the non-significant comparison between HØH and Simulated H-H for Speaker03 corresponds to the classification plot and histogram, as Speaker03 indeed had the least amount of tokens classified as /L/-toned.

[5] This is possible due to there being only two classification labels: lower posterior probability for one label entails higher probability for the other.

Below in Figure 2 are plots related to the HØL tonal context. The trajectory difference between a /L/ realization an interpolation realization will be less obvious here: if a toneless mora is inserted with a /L/ tone, the trajectory will be an early fall (equivalent to HLL); on the other hand, interpolation through Ø towards the following L tone would have a less steep 'late' fall, possibly with higher average pitch in the extracted second syllable.

Figure 2: HØL classification plots and posterior probability histograms, Speakers 01-04.

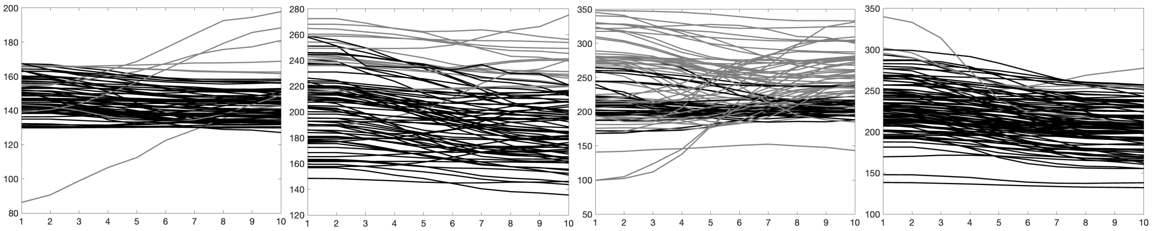


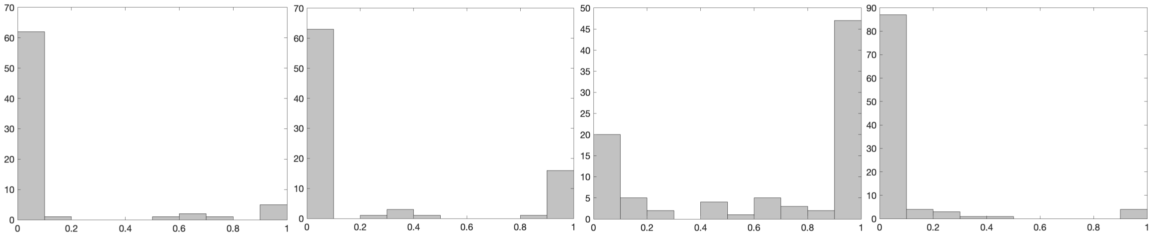
Results for the HØL context show a somewhat different picture. Speakers 01-03 all had the majority of their toneless tokens classified as /L/-toned, while only a small percentage of words (typically with lower average pitch) were classified as interpolation. Speaker 04, however, had heavy overlaps between the /L/-toned (black) and interpolation (gray) trajectories, also confirmed by the somewhat even distribution of posterior probabilities in the accompanying histogram. Judging from the classification results, we may conclude that with regard to the HØL context, three speakers used /L/ insertion as the main realization strategy while one used 'reduced' /L/ tokens that were in between canonical /L/ tones and straight interpolations.

Interestingly, the HØL results may have also revealed the limitations of the classification model under this context. We may infer the classification criteria of the Naive Bayes model from the distribution of interpolation (gray) trajectories in the plots — tokens with lower average pitch (ones that fell under the lower half of the graph) were more likely classified as interpolation. Several tokens with almost level pitch were incorrectly classified as /L/-toned, possible due to their high overall pitch. Although it is possible that falls with higher average f0 typically have steeper slopes (the absolute f0 change over a given window is greater for higher-pitched tokens), these inaccurate classifications do show that the computational model might have paid attention to phonetic cues not central to our research questions during a classification task.

Classification plots and histograms for the LØH condition are summarized in Figure 3. Corresponding to the competing hypotheses, if the toneless mora is inserted with a /L/ tone, the entire second syllable will be /L/-toned and similar to the LLH context; if the toneless mora stays without an inserted tone, we would expect an interpolated slight rise towards H in the third syllable.

Figure 3: LØH classification plots and posterior probability histograms, Speakers 01-04.





We again see rather different realization patterns across the four speakers. Speakers 01 and 04, while realizing most of their tokens with a L level pitch, still had several word tokens that clearly assumed a rising form. Speakers 02 and 03 had word tokens at both ends (0 and 1) of the posterior probability scale, but very few items in between. This indicates that the classification models for these two speakers had high confidence that toneless moras either realized as /L/ tone or interpolation (H3, optional realization). There was also considerable difference between these two ‘optional realization’ speakers, with 02 overwhelmingly preferring /L/ tone insertion and 03 preferring interpolation.

So far, I have demonstrated that there was significant variability in how a toneless mora (at the end of light-heavy disyllables) was interpreted in Suzhou Chinese: speakers differed in their preference of inserting a default /L/ tone or keeping the mora toneless throughout and leaving pitch implementation to interpolation. Furthermore, each speaker’s strategy of realizing a toneless mora was not uniform under different tonal contexts either. Speakers 02 and 03, for example, preferred more interpolation realization in contexts HØH and LØH, but nevertheless made frequent use of default /L/ under the HØL context. These findings corroborated with the MANOVA results in §3.2: for most speaker-tonal context combinations, there was no single way of interpreting/realizing tonelessness. Table 4 summarizes each speaker-tonal context combination and the hypothesis that most closely matches with the classification results. In the next section, I offer some preliminary discussion on possible factors conditioning different ways of toneless realization.

Table 4: Summary of different realization strategies of tonelessness

	Speaker01	**Speaker02**	**Speaker03**	**Speaker04**
HØH	H1	H2	H2	H1
HØL	H1	H1	H1	H4
LØH	H1	H3	H3	H1

5 Condition factors to toneless realization

As discussed above, in addition to the tonal contexts these toneless moras appear in, there still might be other factors that would contribute to whether /L/ insertion or interpolation takes place. One such factor is speaker sex: Speakers 02 and 03, the two biologically female speakers, demonstrated some preference towards interpolation in at least two tonal contexts. The two male speakers (01, 04), on the other hand, used /L/ tone insertion in most tonal contexts. In addition, while the preceding tone of the toneless context resulted from specific tone sandhi patterns with little variance (e.g.., the initial H in HØH and HØL came from the same sandhi), the following tone came from all seven lexical tones in Suzhou, partially due to the experiment design. It is also possible that some certain following lexical tones have disproportionately contributed to a specific way of toneless realization.

In order to see how these possible factors interact, I fitted a logistic mixed-effects models with classification label (0 = default /L/; 1 = interpolation) as the dependent variable. The independent variables included: Tonal condition (HØH, HØL and LØH), Biological sex (F/M) and Following tone (/H/, /LH/, /HL/, /HLH/, /LHL/, /H?/, /L?/). I also included by-speaker and by-token random intercepts. The independent variables were sum-contrast-coded, as there was no particular reference level informed by the research question. I report the results in Table 5.

Table 5: Logistic mixed-effects model predicting classification label from Tonal context, Following lexical tone and Sex

	Estimate	**Std Error**	**z-value**	**p-value**	
Intercept	0.4458	0.21064	2.116	0.034311	*
Context-HØH	1.04376	0.12154	8.588	2.00E-16	***
Context-HØL	-0.49073	0.13121	-3.74	0.000184	***
Following-H	-0.29846	0.21244	-1.405	0.160047	
Following-H?	0.11047	0.25616	0.431	0.666291	
Following-HL	0.22898	0.22401	1.022	0.306692	
Following-HLH	0.08535	0.27486	0.311	0.756165	
Following-LH	-0.65697	0.38439	-1.709	0.087425	.
Following-LH?	-0.01837	0.77976	-0.024	0.981201	
Sex-M	-1.42191	0.20705	-6.868	6.53E-12	***

After accounting for the random effects, there was a significant effect of Tonal context: HØH context had more tokens classified as interpolation (indicated by the positive estimate) than both HØL and LØH. None of the following lexical tone effect reached significance, although a following /LH/ might have a trending effect on more /L/ realization. Biological males were overall more likely to use /L/ insertion, indicated by the negative estimate value. The mixed-effects model does confirm that the choice between /L/ insertion and interpolation was affected by factors such as tonal context and speaker sex. Although being preliminary in nature, the regression results do point to an additional advantage for the classification models: that classification labels assigned by the models may consist of robust categorical data, on which further quantitative sociolinguistic analyses can be done. There were yet too few speakers to establish a robust correlation with regard to speaker age, socio-economical class or education level. I leave this to future studies when more speaker data are collected and included.

6 Conclusion

In this study, I have demonstrated that toneless moras arising from non-exhaustive metrical parsing assume variable pitch realization in Suzhou Chinese. The variability lies in both cross-speaker and cross-context patterns, as different speakers might prefer one of the two realization strategies — default /L/ insertion and pitch interpolation — and toneless moras under different tonal contexts showed variable proportions of /L/ vs. interpolation as well. These findings is congruent with an optional/probabilistic model of phonology (Coetzee & Pater, 2011; Coetzee & Kawahara, 2013), where the application of certain phonological processes is not categorical and uniform across the board. Crucially, by applying the stochastic simulation methodology of Shaw & Gafos (2015) and Shaw & Kawahara (2018), I was able to first simulate a set of interpolation pitch trajectories with appropriate variability in natural speech, and use the interpolation data set for further statistical analyses. By subsequently fitting the interpolation and /L/ tone data sets to a Naive Bayes classification model (Shaw & Kawahara, 2018; Zhang et al., 2019), I was able to access token-by-token realization of the toneless Ø data, which was not possible with traditional statistical tests that focused on treatment effects over entire groups (e.g., effect of tonal context over all relevant items). I have also demonstrated that the categorical classification labels, albeit with some degree of inaccuracy, could be used to probe the conditioning factors that have caused the optional phonological realization in the first place. This is particularly useful for sociolinguistic research focusing on ongoing sound changes or synchronic variations, where different surface forms of a symbolic representation coexist, sometimes even within the same speaker.

On a broader note, the computational approach to the Suzhou f0 data also provides strong phonetic support for the formal foot analysis of Suzhou Chinese: the fact that light-heavy disyllables contain pitch variability beyond the level observed in other quantity profiles (e.g., heavy-heavy, light-light) gives more credit to a toneless or 'targetless' interpretation to the last/third mora in question. The phonetic evidence along with phonological arguments of quantity-sensitive footing (Dresher & van der Hulst, 1998; Iosad, 2013) and foot-tone licensing conditions (de Lacy, 2002; Breteler, 2018) contributes to a full account of metrically-guided tonology of Suzhou Chinese.

References

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

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

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

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

Coetzee, Andries W. & Shigeto Kawahara (2013). Frequency biases in phonological variation. *Natural Language & Linguistic Theory* 31:1, 47–89, URL `http://link.springer.com/10.1007/s11049-012-9179-z`.

Coetzee, Andries W. & Joe Pater (2011). The place of variation in phonological theory. Goldsmith, John A., Jason Riggle & Alan C. Yu (eds.), *The Handbook of Phonological Theory*, Blackwell Publishing Ltd., 401–434.

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

Duanmu, San (1995). Metrical and Tonal Phonology of Compounds in Two Chinese Dialects. *Language* 71:2, 225–259.

Iosad, Pavel (2013). Head-dependent asymmetries in Munster Irish prosody. *Nordlyd* 40:1, 66–107.

Kager, René (1993). Alternatives to the Iambic-Trochaic Law. *NLLT* 11:3, 381–432.

Kager, René & Violeta Martínez-Paricio (2018). Mora and syllable accentuation – Typology and representation. *The Study of Word Stress and Accent – Theories, methods and data*, Cambridge: Cambridge University Press, 147–186.

de Lacy, Paul (2002). The interaction of tone and stress in Optimality Theory. *Phonology* 19:2002, 1–32.

Pierrehumbert, Janet B. (1980). *The phonology and phonetics of English intonation*. Phd dissertation, MIT, URL `http://en.scientificcommons.org/736982`.

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

Pulleyblank, Douglas (1986). *Tone in Lexical Phonology*, vol. 4 of *Studies in Natural Language & Linguistic Theory*. Springer Netherlands, Dordrecht, URL `http://link.springer.com/10.1007/978-94-009-4550-0`.

Selkirk, Elisabeth (1986). On derived domains in sentence phonology. *Phonology Yearbook* 3, 371–405, URL `https://www.cambridge.org/core/product/identifier/S0952675700000695/type/journal_article`.

Shaw, Jason A. & Adamantios I. Gafos (2015). Stochastic Time Models of Syllable Structure. *PLOS ONE* 10:5, p. e0124714, URL `https://dx.plos.org/10.1371/journal.pone.0124714`.

Shaw, Jason A. & Shigeto Kawahara (2018). Assessing surface phonological specification through simulation and classification of phonetic trajectories. *Phonology* 35:3, 481–522, URL `https://www.cambridge.org/core/product/identifier/S0952675718000131/type/journal_article`.

Shi, Xinyuan & Ping Jiang (2013). A Prosodic Account of Tone Sandhi in Suzhou Chinese. *Proceedings of the 25th North American Conference on Chinese Linguistics*.

Xu, Yi (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 7–10.

Yip, Moria (2002). *Tone*. Cambridge: Cambridge University Press.

Yu, Weiqi (2010). The popularization of putonghua and the maintenance of Suzhou dialect——A survey on language situation of native students in Suzhou. *Applied Linguistics* 3, 61–69.

Zhang, Muye, Christopher Geissler & Jason Shaw (2019). Gestural Representations of Tone in Mandarin : Evidence From Timing Alternations. *International Congress of Phonetic Sciences ICPhS 2019*, August, 1803–1807.

Zhu, Yuhong (2020). Extending the Autosegmental Input Strictly Local Framework : Metrical Dominance and Floating Tones. *Proceedings of the Society for Computation in Linguistics*, vol. 3, 393–401, URL `https://scholarworks.umass.edu/scil/vol3/iss1/38`.

Zhu, Yuhong (2021a). Moraic Footing in Suzhou Chinese: Evidence from Toneless Moras. *Proceedings of the Annual Meetings on Phonology*, vol. 9, URL `https://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/4862`.

Zhu, Yuhong (2021b). Quantity-sensitive Foot Formation in Suzhou: Evidence from Light-initial Tone Sandhi. *Proceedings of the 32nd North American Conference on Chinese Linguistics*.

Zhu, Yuhong (forthcoming). A Metrical Analysis of Light-initial Tone Sandhi in Suzhou Wu .