Proceedings of the 29th North American Conference on Chinese Linguistics (NACCL-29). 2017. Volume 1. Edited by Lan Zhang. University of Memphis, Memphis, TN. Pages 99-113.

Taiwan Mandarin Tone Sandhi
Variation of the Intonational Phrasing in Fast Speech

Yi Jen Chen
National Chengchi University, Taiwan

This paper discusses the domain of Taiwan Mandarin tone sandhi under the framework of Nespor & Vogel's (1986) Prosodic Hierarchy based on a Taiwan Mandarin spoken corpus of 324 sentence tokens consisting of 6-24 third tone syllables. The analysis first shows that under fast speech (i.e. above 200 beats per minute) both cyclical and simultaneous tone sandhi may apply within a long utterance. Second, the phonological and the intonational phrasing as defined by Nespor & Vogel needs a revision to account for the various surface tone patterns of the sentence tokens. Third, the revised prosodic hierarchy incorporates explanation of the classifier cliticization without making stipulations.

0. Introduction

Mandarin Chinese is a tone language consisting of four lexical tones and a neutral tone, and Taiwan Mandarin (TM) is one of the Mandarin dialects spoken in Taiwan. In terms of its lexical tones, the first tone (marked as 1 in the following derivations) is a high level tone; the second (2) a mid-rising tone; the third (3) a low tone; and the fourth (4) a high-falling tone. When a sequence of third tones occurs in an utterance, the non-final third tones may undergo a phonological rule and surface as second tones. This rule is referred to as third tone sandhi (3TS). The domain within which 3TS applies influences the number of derived second tones and is argued to be the prosodic domains such as the foot, the phonological phrase or the intonational phrase by previous studies (Shih 1986; Hung 1987; Cheng 1987; Hsiao 1991; Lin 2002). In other words, when the prosodic domains of a single sentence vary, different tone patterns can be observed, which is reflected in the Taiwan Mandarin corpus. Therefore, to describe and predict the various surface tone patterns, it is necessary to define the prosodic domain within which 3TS applies.

However, different studies assume different prosodic domains. Shih (1986), Hung (1987) and Hsiao (1991) assume that 3TS applies cyclically from the foot to the intonational phrase, while Cheng (1987) adopts the prosodic hierarchy comprising the word, clitic group, phonological phrase and the intonational phrase level and argues that 3TS applies cyclically from the word to the intonational phrase. Despite that the 3TS domains they propose differ from each other, these proposals are unanimous in assuming that the highest domain within which 3TS applies is the intonational phrase. When the

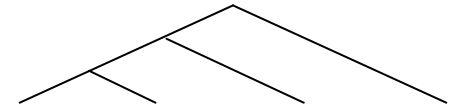
speech rate is fast, simultaneous 3TS optionally applies to the whole intonational phrase. In these approaches, the domain formations under the intonational phrase (e.g. foot/phonological phrase formation etc.) are clearly defined and their variations restricted. Yet, the construction and reconstruction of the intonational phrase are not as clearly defined.

Accordingly, based on a Taiwan Mandarin corpus of 324 sentence tokens containing 6-24 third tone syllables, the following sections of the paper first discuss the observation of the surface tone patterns, then review the problems incurred by the above-mentioned approaches to 3TS in sentences and suggest a revision of the prosodic hierarchy approach to account for intonational phrase variations. The cliticization of classifiers is also explained.

1. Literature review

Previous studies have shown that 3TS domain is not determined by the syntactic structure, but derived from it. For instance, in (1), if 3TS applies within syntactic constituents, the natural tone pattern 2323, cannot surface.

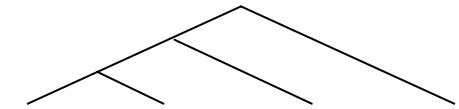
(1) Which kind of wine is good?



na3	*zhong3*	*jiu3*	*hao3*
'which	kind	wine	good'
(2	3)	3	3
(2	2)	3)	3
(2	2)	2)	3)
*(2	3)	(2	3)

To solve this problem, previous studies have regarded the domain of 3TS as determined by the prosodic categories (i.e. the foot, PhP and IP), and the boundaries of the prosodic categories are derived from syntax. One example of the prosodic approach is Shih's (1986) *Foot Formation Rules* (FFR), who states that immediate constituents (*na3 zhong3* 'which kind') form into disyllabic feet first, and then the unpaired syllables (*jiu3 hao3* 'wine', 'good') form into binary feet if they branch to the same direction. It derives the tone pattern 2323. Also, under a fast speech rate, 3TS optionally applies to the whole sentence, which is the IP domain (*na3 zhong3 jiu3 hao3* 'which kind of wine is good?'), producing a 2223 tone pattern. The derivations are shown in (2).

(2) Which kind of wine is good?

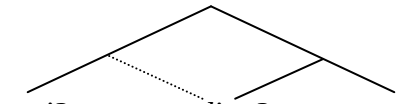


na3	*zhong3*	*jiu3*	*hao3*
'which	kind	wine	good'
(2	3)	3	3
(2	3)	(2	3)
(2	2	2	3)

The prosodic approaches have succeeded in proving and explicating the mismatches between 3TS domain and the syntactic boundaries, though they disagree in the formation of the prosodic boundaries or the prosodic categories 3TS applies to. Shih (1986) and Hsiao (1991) both see the foot as the domain of 3TS, but their theories differ to some extent. Hsiao (1991) proposes a set of *Beat Assignment Rules* and *Foot Formation Rules*. Unlike Shih (1986), in his theory foot formation forms beats, rather than syllables into feet, and that every syllable has to be assigned a beat before foot formation operates. In addition, the beat assignment of function words only takes place after the beat assignment and foot formation of lexical words. He also separates the foot from the PhP and the IP into different hierarchies. The foot belongs to the metrical hierarchy. The PhP and the IP belong to the prosodic hierarchy. These two hierarchies operate simultaneously and interact with each other in the grammar, and the domain of foot formation is restricted by the domain of the PhP and by the IP. In other words, foot formation operates cyclically on the PhP and the IP level.

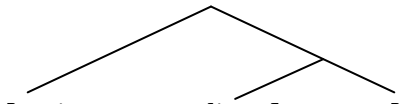
However, both Shih (1986) and Hsiao (1991) need to stipulate the cliticization of classifiers to derive correct surface tone patterns for phrases with classifiers. The classifier *dian3* 'some' in (3a) is within an immediate constituent with *shui3* 'water', so Shih (1986)'s FFR require that *dian3* and *shui3* form a foot first. If this were the case, a wrong surface tone pattern 323 would surface. To repair this problem, she stipulates that the classifier cliticizes to the verb on its left to form a word regardless of the syntactic structure, in order to form *mai3* and *dian3* into a foot and derive a 223 pattern, this stipulation weakens the FFR. The cliticization is shown by the dotted line in (3a). Hsiao (1991) also draws on cliticization, which is of the phonological phrase. The boundaries of the PhP and the IP are indicated by the square brackets in (3b). The classifier *dian3* cliticizes to *mai3* to form a PhP, and foot formation can only operate within the PhP on the PhP level. Hence, *mai3* undergoes 3TS on the PhP level, and *dian3* undergoes 3TS on the IP level. The correct tone pattern 223 is thus derived.

(3) Comparison of Shih's (1986) and Hsiao's (1991) models
a. Cliticization in Shih (1986)



mai3	*dian3*	*shui3*
'buy	some	water'
(2	3)	3
((2	2)	3)

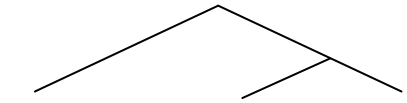
b. Cliticization in Hsiao (1991)



[*mai3*	*dian3*]	[*shui3*]—PhP	
[		]—IP	
'buy	some	water'	
x		x	
	x		
(x	x)	(x)—	Foot formation on the PhP level
(2	3)	(3)	
(x	x	x)—	Foot formation on the IP level
(2	2	3)	

Cheng (1987) follows Hayes's (1986) *Prosodic Hierarchy* and maintains that 3TS operates cyclically within the word, the clitic group, the PhP and the IP level, rather than within the foot. Yet, her model still warrants a stipulation that, unlike other function words, classifiers always cliticize to their left. A demonstration of her model is shown in (4).

(4) Cliticization in Cheng (1987)



mai3	*dian3*	*shui3*	
'buy	some	water'	
(3)	(3)	(3)—	word level
(2	3)	(3)—	clitic group level
(2	2	3)—	PhP level
(2	2	3)—	IP level

Another problem in Cheng's (1987) model is the construction rule of the phonological phrase, which is formed by obligatorily adjoining all the materials on the non-recursive side to their head. The PhP domain varies only in whether the complement on the recursive side is also adjoined. This definition may make the PhP too wide if the non-recursive side of the head contains many clitic groups. Hsiao (1991) also gave a definition of the PhP boundaries, which is marked on the left edge of every branching maximal projection, but this definition deprives the PhP domain of its variability, because this means the head obligatorily adjoins the complement on the recursive side as well. Therefore, the following analysis adopts Nespor & Vogel's (1986) definition that a PhP obligatorily contains all the clitic groups on the nonrecursive side of a head X within its maximal projection XP and optionally contains the first complement on the recursive side. In this way, the length of the PhP domain is restricted while retaining its flexibility to adjoin the complement.

Although previous theories are consistent in the assumption that 3TS does not apply across IP boundaries, their definition of the IP domain remains vague. Since in Cheng (1987) and Hsiao (1991) the IP belongs to the prosodic hierarchy, an IP must be composed of one or more PhPs according to the Selkirk's (1984) *Strict Layer Hypothesis* that requires a prosodic constituent on one level (e.g. an IP) to immediately dominate only constituents of the next lower level (e.g. a PhP). Hsiao (1991) further adds that the IP should also conform to the Selkirk's (1984) *Sense Unit Condition*, which commands that the constituents within an IP must bear either a head-modifier or a head-argument relation. Nevertheless, to provide an explicit definition of the IP, the definitions of the lower prosodic categories (i.e. the word, the clitic group, the PhP) must also be given.

Via observing the tone patterns in the corpus, the following analysis examines whether Nespor & Vogel's (1986) prosodic hierarchy, as in (5), is able to account for the variability of Taiwan Mandarin 3TS domain formation. In their model, the foot and the syllable are below the phonological word; however, this study assumes that the metrical categories like the foot and the syllable should belong to a different hierarchy (see Hsiao 1991). Thus, the prosodic hierarchy in (5) does not include these two metrical categories.

(5) The prosodic hierarchy

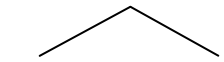
Utterance
|
Intonational Phrase (IP)
|
Phonological Phrase (PhP)
|
Clitic group (CG)
|
Phonological word (PhWd)

2. Data collection and corpus organization

As pointed out by Shih (1986), Cheng (1987) and Hsiao (1991) among others, 3TS applies either cyclically or simultaneously to the IP level under a fast speech rate, so the corpus data were recorded under the speech rate of 200 beats/word per minute with an attempt to see how 3TS applies to the IP and where the IP boundaries sit in the utterances of 6 to 24 syllables. The informants are two female and two male Taiwan Mandarin native speakers aging from 25 to 27. Each of them was asked to read 81 sentences naturally with at least the speed of 200 beats per minute, while a metronome was present for their reference. All in all, the corpus contains 324 sentence tokens of 49 types of syntactic structures.

In Nespor & Vogel (1986), the PhWd is defined as coextensive with the terminal node of the syntactic tree. The following analysis incorporates Cheng's (1987) definition of the word and treats any lexical item or "anything that is derived through a morphological process" as a PhWd (p.40), so compounds are also seen as individual PhWds. To investigate whether 3TS applies on the PhWd level or skips applying on this level, 624 binary branching XPs were examined. Each of the XPs either contains a head and its modifier (*xiao3 lao3 shu3* 'little mouse') or a verb head and its object (*mai3 bao3 yang3 pin3* 'buy skin-care products') as shown in (6). If the final syllable of the modifier (*xiao3* 'little') or the final syllable of the verb head (*mai3* 'buy') undergoes 3TS, it can be assumed that 3TS operates on the PhWd level. If they do not undergo 3TS, 3TS skips this level.

(6) Binary branching XPs



xiao3 *lao3 shu3* 'little mouse'
mai3 *bao3 yang3* pin3 'buy skin-care products'

The next higher prosodic category is the CG, which includes one non-clitic (lexical) word plus, if there are any, one or more clitic (function) words. It should be noted that the present analysis assumes Hung's (1987) and Lin's (2006) hypothesis that prepositions are lexical, not functional, in Mandarin because they are developed from the grammaticalization of verbs, so they still preserve some verbal characteristics such as taking aspectual markers and functioning as the main predicate of a sentence. This study supports their hypothesis in arguing that the phonological phrasing of prepositions also behaves like verbs, so prepositions are not considered as clitic words (see Section 4).

The total number of clitic tokens is 20 (12 pronouns and 8 classifiers). If the clitic word (*dian3*) cliticizes to the left, its preceding non-clitic word undergoes 3TS, as in (7). If the clitic word (*wo3*) cliticizes to the right, the clitic word itself undergoes 3TS, as in (8).

(7) Left cliticization

gei3	*dian3*	*shui3*
'give	some	water'
(3)	(3)	(3)—PhWd
(2	3)	(3)—CG

(8) Right cliticization

wo3	*xiang3*	*xi3*	*leng3*	*shui3*	*zao3*
'I	want	take	cold	water	shower'
(3)	(3)	(3)	(2	2	3)—PhWd
(2	3)	(3)	(2	2	3)—CG

On the PhP level, a CG which is a head X in the syntactic structure obligatorily adjoins all the CG(s) on its nonrecursive side within XP and optionally includes the first complement on its non-branching recursive side. Here, the recursive side refers to the complement side, and the non-recursive side refers to the opposite side (Giorgi and Longobardi, 1991). If the recursive side branches, optional inclusion of the complement is not permitted. Optional phonological phrasing is the most seen in that of the verb heads because they occur with modifiers and complements much more frequently than the noun heads. Therefore, the analysis focuses on 584 verb head tokens and classifies them into 5 types, as in (9). The first type denotes the verb heads containing only the recursive sides. The second type of verb heads has only the nonrecursive sides. The third type of verb heads has both the recursive and the nonrecursive sides. To examine whether Nespor & Vogel's (1986) definition of PhP construction and reconstruction is capable of describing the phonological phrasing of the verb heads, three targets of observation are made, as shown in (10).

(9) Classification of verb heads

	[±branching] Complement	Number of tokens
With a recursive side	-	252
	+	140
With a nonrecursive side		76
With both sides	-	92
	+	24
Number of V heads		584

(10) Targets of observation

a. Whether the head obligatorily includes all the CG(s) on its nonrecursive side within the maximal projection

b. Whether the head obligatorily includes, optionally includes, or obligatorily not includes the CG(s) on the recursive side

c. Whether the recursive side of the adjoined CG(s) branches or not

A verb head X is regarded as obligatorily includes all the CG(s) on its nonrecursive side within XP when the final syllable (*jin3*) on the nonrecursive side undergoes 3TS as in (11). If it undergoes 3TS the verb head forms a PhP with the CG(s) on its nonrecursive side and vice versa.

(11) Verb head forming a PhP with the CG on its nonrecursive side

gan3 jin3		*zhao3*
'quickly		find'
(2	3)	(3)—PhWd
(2	3)	(3)—CG
(2	2	3)—PhP

If the verb head forms a PhP with the CG(s) on the recursive side, the final syllable (*yao3*) of the verb head undergoes 3TS, as in (12). Note that the phonological phrasing of the NP also operates on this level, so the derivation in (12) shows the phonological phrasing of the NP before the verb head forms a PhP with the NP. As for (10c), the branchingness of the recursive side will also be observed when the inclusion occurs.

(12) Verb head forming a PhP with the CG on its recursive side

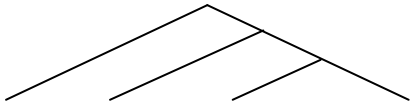
yao3	*xiao3*	*lao3 shu3*
'bite	small	mouse'
(3)	(3)	(2 3)—PhWd
(3)	(3)	(2 3)—CG
(3)	(3	2 3)—PhP (NP)
(2	3	2 3)—PhP (VP)

Here the branchingness refers to that of the prosodic structure, not of the syntactic structure, as in (13) and (14). If the prosodic structure of the complement (*xiao3 lao3 shu3*) does not branch, the recursive side is considered non-branching. If the prosodic structure of the complement (*xiang3 sheng3 can1 qian2*) branches, the recursive side is considered branching. As will be shown in the following analysis, the variation of the IP domain and the surface tone pattern is strongly related with the flexibility of phonological phrasing.

(13) The prosodic structure of the non-branching complement on the PhP level

yao3 *xiao3* *lao3 shu3*
'bite small mouse'
(3) (3 2 3)—PhP (NP)

(14) The prosodic structure of the branching complement on the PhP level



Lao3 Li3 *xiang3* *sheng3* *can1 qian2*
'Old Li want save food cost' (Old Li wants to save food cost.)
(2 3) (3) (3) (1 2)—PhWd
(2 3) (3) (3) (1 2)—CG
(2 3) (3) (3) (1 2)—PhP
(2 2 3) (3 1 2)—IP (1)
(2 3) (2 3 1 2)—IP (2)

Although the IP domain in Nespor & Vogel's (1986) definition merely requires that it is formed by one or more PhPs, the present analysis adopts Selkirk's (1984) and Hsiao's (1991) view that the constituents within an IP should bear either a head-modifier or a head-argument relation. The argument here designates the subject or the object of the verb head. With a unified definition of the lower prosodic domains (i.e. PhWd, CG, PhP), when the derived surface tone pattern on the PhP level, like the one on the PhP level in (14), is not consistent with the collected surface tone pattern in the corpus, like the ones on the IP level in (14), it is assumed that the domain of the IP is at play. The construction and reconstruction rule of the IP can thus be inferred.

3. Cyclic and simultaneous TS within an utterance

Since previous studies only discuss the derivation of 3TS patterns in short sentences, the present study collected sentences of 6-24 syllables in order to see if the length of an utterance has any influence on the derivation. The observation of the collected data finds that both cyclic and simultaneous 3TS can apply to a single utterance as in (15). Statistics also suggests that sentence length does have influence on whether an utterance has both cyclic and simultaneous application in different IPs. This phenomenon is observed respectively in sentences of more than 7 syllables. The number of sentence tokens undergoing both cyclic and simultaneous application is shown in (16). There is a noticeable gap between 14- to 23-syllabled sentence tokens because the corpus lacks sentences of 14 to 17 syllables and of 19 syllables, but the influence of length is still obvious. Sentences that are shorter than 7 syllables do not show such a phenomenon. In

fact, this finding corresponds to Shih's (1986) assumption that while it sounds natural for a two-foot phrase to form an IP, it sounds relatively unnatural when an IP consists of more than three feet (i.e. more than 6 syllables). Therefore, sentences of more than 6 syllables should naturally be reconstructed into at least 2 IPs. Since 3TS does not apply across the IP boundaries, it is reasonable that cyclic 3TS applies to one IP and simultaneous 3TS applies to the other in long sentences.

(15) Cyclic and simultaneous 3TS in an utterance

a. '…so Miss Jiang dares to buy skin-care products from Little Female Ghost'

suo3 yi3 jiang3 xiao3 jie3 gan3 zhao3 xiao3 nu3 gui3 mai3 bao3 yang3 pin3

'so Jiang Miss dare find Little Female Ghost buy skin-care products'

(2 3) (3) (2 3) (3) (3) (3) (2 3) 3 3 3 3 —PhWd
(2 3) (3) (2 3) (3) (3) (3) (2 3) 3 3 3 3 —CG
(2 3) (3 2 3) (2 3) (3 2 3) 3 3 3 3 —PhP
(2 2 3 2 3 2 2 3 2 3) (2 2 2 3)—IP

b. '…apply to the face and to the toes'

wang3 lian3 mo3 ye3 wang3 jiao3 zhi3 mo3

'to face apply and to toes apply'

(3) (3) (3) 3 3 3 3 3 —PhWd
(3) (3) (3) 3 3 3 3 3 —CG
(3) (2 3) 3 3 3 3 3 —PhP
(3 2 3) (2 2 2 2 3)—IP

(16) Number of sentence tokens undergoing both cyclic and simultaneous 3TS

No. of syllables	7	8	9	10	11	12	13	23	22	23
No. of tokens	1	5	6	6	9	7	1	4	2	4

4. A revision of Nespor & Vogel's (1986) prosodic hierarchy

The following analysis examines whether Nespor & Vogel's (1986) prosodic hierarchy can account for the cliticization of function words as well as IP formation in the corpus. Due to limited space, the analysis focuses mainly on the definition of the PhP and the IP, because the definitions of the PhWd and the CG do not differ significantly from the original ones. First, a PhWd is defined as any lexical item or any morphologically derived word, and this does not contradict with the original definition that requires a PhWd to be coextensive with the terminal node of a syntactic tree. On the next higher level, a CG is formed by a non-clitic word (i.e. a lexical word) plus one or more clitic words if there are any function words. This definition also remains the same. However, it is important to note that this study claims prepositions to be lexical words in Mandarin because they are argued to retain some characteristics of verbs, which is the prosodic phrasing in the derivation of 3TS.

As for the PhP formation, Nespor & Vogel (1986) contends that a head X obligatorily forms a PhP with all the CG(s) on the nonrecursive side within XP and

optionally includes the first complement on the nonbranching recursive side. The corpus data correspond to the reconstruction rule in that the head does optionally include the nonbranching complement, as in (17). The category "either" means that whether the head forms a PhP with the complement or not has no influence on the surface tone patterns. When the complement branches, the verb head does not form a PhP with the complement, as shown in (18). When the nonrecursive side branches, the verb heads obligatorily include all the CGs on the nonrecursive side, as shown in (19).

(17) Verb heads phrasing PhPs with nonbranching complements

Inclusion	Non-inclusion	Either	No. of verb heads
3.5% (12)	18% (62)	78.5% (270)	100% (344)

(18) Verb heads not phrasing PhPs with branching complements

Inclusion	Non-inclusion	Either	No. of verb heads
0	63% (103)	37% (61)	100% (164)

(19) Verb heads phrasing with the nonbranching nonrecursive side

Inclusion	Non-inclusion	Either	No. of verb heads
67% (35)	0	33% (17)	100% (52)

What Nespor & Vogel's (1986) definition fails to explain is how the verb heads form PhPs with the branching nonrecursive side. It is observed that even when the nonrecursive side branches, the verb head still forms a PhP with it, but the inclusion is restricted to only one CG on the nonrecursive side, as in (20). The domain formation of such verb heads is shown in (21).

(20) Verb heads phrasing with the branching nonrecursive side

Inclusion of 1CG	Non-inclusion	Either	No. of verb heads
39% (55)	30% (42)	31% (43)	% (140)

(21) Verb heads forming a PhP with one CG on the branching nonrecursive side
'Old Li asks my cousin to go north'

Lao3	*Li3*	*qing3*	*biao3*	*jie3*	*wang3*	*bei3*	*zou3*	
'Old	Li	ask	cousin		toward	north	go'	
(2	3)	(3)	(2	3)	(3)	(3)	(3)	—PhWd
(2	3)	(3)	(2	3)	(3)	(3)	(3)	—CG
(2	3)	(3)	(2	3)	(3)	(**2**	**3**)	—PhP
(2	2	3	2	2	3	2	3)	—IP

In (21) since prepositions are seen as lexical words, they form clitic groups by themselves on the CG level. On the PhP level, when the preposition head (*wang3*) does not form a PhP with the complement (*bei3*), the nonrecursive side of the verb (*zou3*)

branches. The verb head must include the CG *bei3* to derive the correct surface tone pattern. Accordingly, the branchingness of the nonrecursive side should be considered in Nespor & Vogel's (1986) definition of the construction rule of PhP. A revision is given in (22).

(22) Revision of the PhP construction rule

a. A head X obligatorily forms a PhP with the CG(s) on the nonrecursive side within its maximal projection (XP).

a-1. When the prosodic structure of the nonrecursive side branches, the head forms a PhP with only one CG on the nonrecursive side within XP.

a-2. When the prosodic structure of the nonrecursive side does not branch, the head forms a PhP with all the CGs on the nonrecursive side within XP.

b. A head X optionally forms a PhP with the first complement on the non-branching recursive side.

Having defined the formation of the PhP, we can now move on to the definition of the IP. The original construction rule for the IP only requires it to be composed of one or more PhPs. However, if this were the case, unnatural IPs would be produced, as in (23). As Nespor & Vogel (1986) observes, there is a tendency to avoid short IPs in natural speech. Therefore, the formation of IP should be revised as well.

(23) Unnatural IP formation

'Old Li asks my cousin to go north'

Lao3	*Li3*	*qing3*	*biao3*	*jie3*	*wang3*	*bei3*	*zou3*	
'Old	Li	ask	cousin		toward	north	go'	
(2	3)	(3)	(2	3)	(3)	(3)	(3)	—PhWd
(2	3)	(3)	(2	3)	(3)	(3)	(3)	—CG
(2	3)	(3)	(2	3)	(3)	(**2**	**3**)	—PhP
(2	3)	(3)	(2	3)	(3)	(**2**	**3**)	—IP

Selkirk (1984) and Hsiao (1991) suggest that IP formation should conform to the *Sense Unit Condition*, which demands either a head-modifier or a head-argument relation between the constituents in an IP. Despite that this condition does eliminate the possibility for an unnatural IP with the inner constituents bearing no semantic relations at all, the IPs in (23) do not violate the condition because individual words such as *qing3* and *wang3* can be sense units by themselves. This brings us to consider the length restrictions of IP formation. Shih (1986) and Hsiao (1991) both make sure that a 3TS domain is longer than at least two syllables by proposing foot formation rules. However, the foot formation approach is incompatible with the cliticization of classifiers without adding stipulations. Furthermore, Hsiao (1991) points out that the metrical and the prosodic hierarchy should be separated, which means that the foot and the syllable in Nespor & Vogel's (1986) prosodic hierarchy should belong to the metrical hierarchy.

This brings about the possibility that the length of an IP is restricted by the requirement of foot formation which avoids monosyllabic feet, and this is the work of the metrical hierarchy, not the prosodic hierarchy. Therefore, the construction rule of the IP does not need to include the length restriction of the metrical hierarchy. A revision of the original definition is given in (24).

(24) A revision of the IP construction rule
a. An IP constitutes one or more PhPs.
b. The constituents within an IP must satisfy the Sense Unit Condition

5. The cliticization of classifiers

The other issue this paper tackles is the cliticization of classifiers, that is, whether it can be accounted for by the prosodic hierarchy model without making further stipulations. The corpus contains two kinds of clitics: classifiers and pronouns. The cliticization of these two pronouns behave differently from each other in that classifiers cliticize only to the left, while pronouns cliticize to both directions, as seen in (25). Since one of the pronouns undergoes cyclic 3TS, it is considered to skip the phrasing on this level. The CG formations of classifiers and pronouns are shown in (26).

(25) Direction of cliticization

	Direction			Total
	Leftward	Rightward	Simultaneous TS	
Pronouns	1	10	1	12
Classifier	8	0	0	8
Total				20

(26) Cliticization of classifiers and pronouns
a. '…give **some** cold water'
gei3 dian3 leng3 shui3
'give some cold water'
(3) (3) (2 3)—PhWd
(2 **3**) (2 3)—CG
(2 3) (2 3)—PhP
(2 3 2 3)—IP

b. 'Miss Li says **you** only apply top skin-care products'

Li3	*xiao3*	*jie3*	*jiang3*	*ni3*	*zhi3*	*mo3*	*ding3*	*ji2*	*bao3*	*yang3*	*pin3*	
'Li	Miss		say	you	only	apply	top		skin-care		products'	
(3)	(2	3)	(3)	(3)	(3)	(3)	(3	2)	(2	2	3)	—PhWd
(3)	(2	3)	(3)	(**2**	3)	(3)	(3	2)	(2	2	3)	—CG
(3	2	3)	(3)	(2	2	3)	(3	2	2	2	3)	—PhP
(3	2	2	3)	(2	2	2	3	2	2	2	3)	—IP

Like this study, Nespor & Vogel (1986) observes two kinds of clitics and proposes the directional and the non-directional ones. The directional clitics they found are in Greek possessives. They also state that the direction of cliticization must be seen as an inherent property of them, and such inherent property does not exist in non-directional clitics. In this model, the cliticization of Mandarin classifiers is explained because they always cliticize to the left. The cliticization of pronouns, on the other hand, is non-directional, so they can either cliticize to the left or the right clitic word. Without making stipulations in the model, the difference in the direction of cliticization is explicated.

6. Conclusion

This paper investigates how 3TS applies in sentences of different lengths, and finds that in utterances of more than 7 syllables both cyclic and simultaneous application of 3TS are observed. After the examination of Nespor & Vogel's (1986) prosodic hierarchy, revisions are made regarding the PhP and IP construction rule. It is assumed that the branchingness of the nonrecursive side also plays an important role in the formation of PhPs. For the IP construction, the requirement of the *Sense Unit Condition* is added, and the tendency to avoid monosyllabic IPs is believed to be constrained by factors outside the prosodic hierarchy (i.e. foot formation).

REFERENCES

CHENG, LISA LAI SHEN. 1987. Derived domains and Mandarin third tone sandhi. *Chicago Linguistic Society* 23.2:16-29.

HSIAO, YUCHAU E. 1991. *Syntax, Rhythm, and Tone: A Triangular Relationship*. Taipei: Crane Publishing Co., Ltd.

HUNG, TONY. 1987. *Syntactic and semantic aspects of Chinese tone sandhi*. San Diego: UCSD dissertation.

LIN, HUI-SHAN. 2002. Constraint reranking and Mandarin tone sandhi. *Chinese Phonology* 11.191-213.

LIN, T.H.J. 2006. Comparative grammar of Chinese and English: An elementary introduction to the grammars of Chinese and English from a comparative

perspective. Online:
http://ling.nthu.edu.tw/faculty/thlin/pdf/comparative_grammar_en_ch.pdf

NESPOR, MARINA and IRENE VOGEL. 1986. *Prosodic Phonology*. Dordrercht: Foris Publication.

SELKIRK, ELISABETH. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge: The MIT Press.

SHIH, CHILIN. 1986. *The prosodic domain of tone sandhi in Chinese*. San Diego: UCSD dissertation.

Proceedings of the 29th North American Conference on Chinese Linguistics (NACCL-29). 2017. Volume 1.
Edited by Lan Zhang. University of Memphis, Memphis, TN. Pages 114-124.

Intonational Phrasing in Zhuolan Raoping

Yuchau E. Hsiao (蕭宇超)
National Chengchi University, Taipei

This paper constructs a corpus of tone sandhi of Zhuolan Raoping, and discusses intonational phrasing in this Hakka dialect. While examining the strict layer hypothesis, this research observes that the intonational phrase (ι) is not sensitive to the formation of sense unit. The corpus also indicates that the ι-length is under restrictions. I have proposed a constraint ranking under the standard prosodic theory, and shown that the match theory is not well-motivated and may result in incorrect predictions.

0. Introduction

This paper discusses the intonational phrasing in Zhuolan Raoping, a Hakka dialect spoken in central Taiwan. I will first address the theoretical background and the tonal basics of this dialect. I will then explain the corpus I built in this research, and propose an analysis. Finally I will comment on an alternative analysis, and then offer a conclusion.

1. Standard prosodic theory

The partial hierarchy in (1) is posited by the Standard Prosodic Theory (Nespor and Vogel 1986, Hayes 1989, Hsiao 1991, among others). An intonational phrase consists of one or more phonological phrases, and a phonological phrase consists of one or more phonological words.

(1) Prosodic Hierarchy

ι	Intonational Phrase
φ	Phonological Phrase
ω	Phonological Word

The Strict Layer Hypothesis in (2) is a constraint for the prosodic hierarchy (Nespor and Vogel 1986, Hayes 1989, Hsiao 1991, among others). A prosodic constituent at a given level of the hierarchy is composed of constituent(s) at the next lower level of the hierarchy. There is no recursion, inversion, nor skipping of prosodic level.

(2) Strict Layer Hypothesis
There is a hierarchy of prosodic constituent types such that, in a prosodic tree, any constituent at a given level of the hierarchy consists exclusively of constituents at the next lower level of the hierarchy.

Selkirk (1986) posits four end-based parameters to construct a phonological phrase, as in (3); namely, a phonological phrase boundary is marked at the right or left edge of an XP or X^{head}.

(3) Phonological Phrasing (Selkirk 1986)
φ = XP], [XP, X^{head}], [X^{head}

In Hsiao (1991, 1995), as in (4) and (5), I observed that a phonological phrase boundary in Taiwanese is marked at the right edge of a non-adjunct and non-clitic XP (cf. also Chen 1987, 2000).

(4) Phonological Phrasing (Hsiao 1991)
φ = {Right, XP^{-a}} where –a = non-adjunct

(5) Phonological Phrasing (Hsiao 1995)
φ = {Right, $XP^{-a \wedge -c}$} where –a = non-adjunct; –c = non-clitic

There are also some proposals regarding intonational phrasing. Selkirk (1984) suggests that the immediate constituents of an intonational phrase may be grouped into a sense unit, as in (6). She proposes a Sense Unit Condition, which considers that two constituents may form a sense unit if they show a modifier-head relation or an argument-head relation, as in (7).

(6) Intonational Phrasing (Selkirk 1984)
The immediate constituents of an intonational phrase must together form a sense unit.

(7) Sense Unit Condition: (Selkirk 1984)
Two constituents C_i, C_j form a sense unit if (a) or (b) is true of the semantic interpretation of the sentence:
a. C_i modifies C_j (a head)
b. C_i is an argument of C_j (a head)

In (8), Nespor and Vogel (1986) follow the Strict Layer Hypothesis and indicate that phonological phrases are grouped into an intonational phrase, and in (9), they suggest that there is a preference of constructing intonational phrases of average length.

(8) Intonational Phrasing (Nespor and Vogel 1986)
Join into an n-ary branching ι all φs included in a string delimited by the definition of the domain of ι.

(9) ι-Length (Nespor and Vogel 1986)
There is a tendency to establish ι's of a more or less uniform, average length.

2. Tonal basics

There are six base tones in Zhuolan Raoping, including three high-register tones and three low-register tones, as in (10). In addition, there is a derived mid tone that occurs only in the surface, as in (11). The rule in (12) shows that high-falling, high-level and low-falling map to mid before any low-register tone, but they map to low before any high-register tone. The low-level, 11, does not undergo tone sandhi.

(10) Tone inventory
High-registered: 55, 53 and 5̲
Low-registered: 11, 31 and 2̲ (where checked tones are underlined)

(11) Derived tone
High-registered: 33

(12) Tone sandhi

$$\left\{\begin{matrix}53\\55\\31\end{matrix}\right\} \longrightarrow 33 \,/\, \_\_\_\_ \left\{\begin{matrix}11\\31\\\underline{2}\end{matrix}\right\}$$

$$\longrightarrow 11 \,/\, \_\_\_\_ \left\{\begin{matrix}53\\55\\\underline{5}\end{matrix}\right\}$$

3. The corpus

In this research, I build a corpus of Zhuolan Raoping, with the help of two male informants, both aged 64. The corpus contains 428 pentasyllabic line tokens, 112 hexasyllabic line tokens, and 571 longer line tokens. There are 4,414 syllable tokens in total. The tone occurrences in the data are coded with different structures, as in (13). (13a) indicates an SU-final, a non-adjunct XP-final and line final position. (13b) indicates an SU-final and a non-adjunct XP-final but not line final position. (13c) indicates an SU-final and adjunct XP-final but not a line final position. (13d) indicates SU-final or non-final.

(13) Structural coding:
- a. $)^{SU}\ ]^{XP\text{-}a}\ ]^{LF}$: SU-final, non-adjunct XP-final and line final position.
- b. $)^{SU}\ ]^{XP\text{-}a}$: SU-final and non-adjunct XP-final but not line final position.
- c. $)^{SU}\ ]^{XP+a}$: SU-final and adjunct XP-final but not line final position.
- d. Y:
  - d.1 $)^{SU}$: SU-final and non-XP-final but not line final position.
  - d.2 Z: non-SU-final, non-XP-final and not line final position.

An example is given in (14), where there are two possible ways to parse the sense units. For this line, we obtained two actual renderings, as in (14a) and (14b). (15a) and (15b) show how the tone occurrences in (14a) and (14b) are counted respectively. (16) is the sum of (15a) and (15b).

(14) $[lo\ moi]^{NP}\ [da\ \ [ten\ fa]^{NP}\ ]^{VP}$

	老 妹	打	電 話	
	sister	call	phone	'My sister called.'
	31 31	31	55 55	Input
	( )	(	( ))	SU pattern 1
	(( )	)	( )	SU pattern 2
a.	33 31 /P	11	33 55 /P	Actual rendering 1
b.	33 31	11	33 55 /P	Actual rendering 2
	33 31	11 /P	33 55 /P	Unacceptable

(15) Instances of calculating

a. = (14a)

	Words	B	S	Total	Percentage	/P
$)^{SU}\ ]^{XP\text{-}a}\ ]^{LF}$	*fa* 話	1	0	1	20%	1
$)^{SU}\ ]^{XP\text{-}a}$	*moi* 妹	1	0	1	20%	1
$)^{SU}\ ]^{XP+a}$						
Y	*lo* 老, *da* 打, *ten* 電	0	3	3	60%	0
Total		2	3	5	100%	2
Percentage		40%	60%	100%		

b. = (14b)

	Words	B	S	Total	Percentage	/P
$)^{SU}\ ]^{XP\text{-}a}\ ]^{LF}$	*fa* 話	1	0	1	20%	1
$)^{SU}\ ]^{XP\text{-}a}$	*moi* 妹	1	0	1	20%	0
$)^{SU}\ ]^{XP+a}$						
Y	*lo* 老, *da* 打, *ten* 電	0	3	3	60%	0
Total		2	3	5	100%	1
Percentage		40%	60%	100%		

(16) Sum of (15a) + (15b)

	Words	B	S	Total	Percentage	/P
)^{SU}]$^{XP\text{-}a}$]^{LF}	*fa* 話	2	0	2	20%	2
)^{SU}]$^{XP\text{-}a}$	*moi* 妹	2	0	2	20%	1
)^{SU}]$^{XP+a}$						
Y	*lo* 老, *da* 打, *ten* 電	0	6	6	60%	0
Total		4	6	10	100%	3
Percentage		40%	60%	100%		

Another example is given in (17), where there are also two possible ways to parse the sense units. For this line, we obtained three actual renderings, as in (17a-c). (18a-c) show how the tone occurrences in (17a-c) are counted respectively. (19) is the sum of (18a-c).

(17) [*lo moi*]^{NP} [*teu ha*]^{AdvbP} [*gong* [*oi shid pon*]^{VP}]^{VP}

	老 妹	頭 下	講	愛 食 飯
	Sister	just	say	like eat meal ‘My sister just said she wanted to eat.’
	31 31	53 55	31	31 <u>5</u> 55 Input
	( )	( )	(	(( ))) SU pattern 1
	( )	(( )	)	(( )) SU pattern 2
a.	33 31 /P	33 55 /P	33	33 <u>2</u> 55 /P Actual rendering 1
b.	33 31	33 55 /P	33	33 <u>2</u> 55 /P Actual rendering 2
c.	33 31	33 55	33	33 <u>2</u> 55 /P Actual rendering 3
d.	33 31 /P	33 55	33	33 <u>2</u> 55 /P Marginal
e.	33 31	33 55	33 /P	33 <u>2</u> 55 /P Unacceptable

(18) Instances of calculating

a. = (17a)

	Words	B	S	Total	Percentage	/P
)^{SU}]$^{XP\text{-}a}$]^{LF}	*pon* 飯	1	0	1	12.5%	1
)^{SU}]$^{XP\text{-}a}$	*moi* 妹	1	0	1	12.5%	1
)^{SU}]$^{XP+a}$	*ha* 下	1	0	1	12.5%	1
Y	*lo* 老, *teu* 頭, *gong* 講, *oi* 愛, *shid* 食	0	5	5	62.5%	0
Total		3	5	8	100.0%	3
Percentage		37.5%	62.5%	100%		

b. = (17b)

	Words	B	S	Total	Percentage	/P
)^{SU}]$^{XP\text{-}a}$]^{LF}	*pon* 飯	1	0	1	12.5%	0
)^{SU}]$^{XP\text{-}a}$	*moi* 妹	1	0	1	12.5%	1
)^{SU}]$^{XP+a}$	*ha* 下	1	0	1	12.5%	1
Y	*lo* 老, *teu* 頭, *gong* 講, *oi* 愛, *shid* 食	0	5	5	62.5%	0
Total		3	5	8	100.0%	2
Percentage		37.5%	62.5%	100%		

c. = (17c)

	Words	B	S	Total	Percentage	/P
)^{SU}]$^{XP\text{-}a}$]^{LF}	*pon* 飯	1	0	1	12.5%	1
)^{SU}]$^{XP\text{-}a}$	*moi* 妹	1	0	1	12.5%	0
)^{SU}]$^{XP+a}$	*ha* 下	1	0	1	12.5%	0
Y	*lo* 老, *teu* 頭, *gong* 講, *oi* 愛, *shid* 食	0	5	5	62.5%	0
Total		3	5	8	100.0%	1
Percentage		37.5%	62.5%	100%		

(19) Sum of (17a) + (17b) + (17c)

	Words	B	S	Total	Percentage	/P
)^{SU}]$^{XP\text{-}a}$]^{LF}	*pon* 飯	3	0	3	12.5%	3
)^{SU}]$^{XP\text{-}a}$	*moi* 妹	3	0	3	12.5%	1
)^{SU}]$^{XP+a}$	*ha* 下	3	0	3	12.5%	2
Y	*lo* 老, *teu* 頭, *gong* 講, *oi* 愛, *shid* 食	0	15	15	62.5%	0
Total		9	15	24	100.0%	6
Percentage		37.5%	62.5%	100%		

(20) shows the total number of tones occurring in difference structures. The parentheses indicate the numbers of the low-level tone.

(20) Statistics including 11

	B	S	Total	Percentage	/P
$)^{SU}\ ]^{XP\text{-}a}\ ]^{LF}$	648 (165)	66 (0)	714 (165)	16.18%	3
$)^{SU}\ ]^{XP\text{-}a}$	744 (68)	131 (3)	875 (71)	19.82%	1
$)^{SU}\ ]^{XP+a}$	73 (8)	52 (3)	125 (11)	2.83%	2
Y	751 (556)	1923 (55)	2674 (611)	60.57%	0
Total	2305 (797)	2109 (61)	4414 (858)	100.0%	6
Percentage	52.22%	47.78%	100%		

The low-level tone, 11, is irrelevant to tone sandhi, and thus the occurrences of this base tone are deducted, as in (21a-d), which show accurate statistic numbers and percentages of tone sandhi.

(21) Statistics excluding 11

a. SU-final, non-adjunct XP-final and line final

	B	S	Total
$)^{SU}\ ]^{XP\text{-}a}\ ]^{LF}$	483	66	549
Percentage	87.98%	12.02%	100%

b. SU-final and non-adjunct XP-final but not line final

	B	S	Total
$)^{SU}\ ]^{XP\text{-}a}$	676	128	804
Percentage	84.08%	15.92%	100%

c. SU-final and non-adjunct XP-final but not line final

	B	S	Total
$)^{SU}\ ]^{XP+a}$	65	49	114
Percentage	57.02%	42.98%	100%

d. SU-final or non-final

	B	S	Total
Y	195	1868	2063
Percentage	9.45%	90.55%	100%

4. Proposed analysis

In this section, I propose an analysis using standard prosodic theory and general alignment. But before that, several patterns are observed from the corpus. First, line-final tones mostly retain their base forms, found in 87.98% of the data. Second, non-adjunct XP-final tones mostly retain their base forms, found in 84.08% of the data. Third, half of the adjunct XP-final tones retain their base forms, found in 57.02% of the data, while half

of them surface with their sandhi forms, found in 49.98% of the data. Fourth, the formation of sense unit does not affect tone sandhi; a SU-final tone surfaces with its sandhi form. Fifth, a pause may optionally occur after an XP. Sixth, a pause cannot be preceded by a sandhi tone, but base tone is not necessarily followed by a pause. Finally, the length difference between the established ι's within an utterance is no more than three syllables. Based on the corpus, I propose a set of constraints in (22), and a partial constraint ranking is proposed in (23).

(22) Constraints

a. ALIGN-R(φ, XP^{-a}): assign one violation mark for every phonological phrase, φ, whose right edge does not coincide with the right edge of a non-adjunct XP.
b. ALIGN-R(φ, XP^{+a}): assign one violation mark for every phonological phrase, φ, whose right edge does not coincide with the right edge of an adjunct XP.
c. ALIGN-R(ι, φ): assign one violation mark for every intonational phrase, ι, whose right edge does not coincide with the right edge of a phonological phrase, φ.
d. ALIGN-R(ι, B): assign one violation mark for every intonational phrase, ι, whose right edge does not coincide with the right edge of a base tone, B.
e. *S/P: assign one violation mark for every pause, P, which is preceded by a sandhi tone, S.
f. ι-RHYTHM: assign one violation mark for every pair of intonational phrases, ι's, who lengths differ in four syllables or more.

(23) Constraint ranking

*S/P, ALIGN-R(ι, B), ALIGN-R(ι, φ) >> ι-RHYTHM

Tableaux (24) and (25) show how this analysis works.

(24) [lo moi]^{NP} [*da* [*ten fa*]^{NP}]^{VP}

		*S/P	ALIGN-R (ι, B)	ALIGN-R (ι, φ)	ι-RHYTHM
☞a.	((*lo moi*)$_{φ}$)$_{ι\ /P}$ ((*da ten fa*)$_{φ}$)$_{ι\ /P}$ S B S S B				
☞b.	((*lo moi*)$_{φ}$ (*da ten fa*)$_{φ}$)$_{ι\ /P}$ S B S S B				
c.	((*lo moi*)$_{φ}$ *da*)$_{ι\ /P}$ ((*ten fa*)$_{φ}$)$_{ι\ /P}$ S B S S B	*(!)	*(!)	*(!)	

(25) [*lo moi*]^{NP} [*teu ha*]^{AdvbP} [*gong* [*oi shid pon*]^{VP}]^{VP}

		*S/P	ALIGN-R(ι, B)	ALIGN-R(ι, φ)	ι-RHYTHM
☞a.	((*lo moi*)$_{φ}$)$_{ι\ /P}$ ((*teu ha*)$_{φ}$)$_{ι\ /P}$ ((*gong oi shid pon*)$_{φ}$)$_{ι\ /P}$ S B S B S S S B				
☞b.	((*lo moi*)$_{φ}$ (*teu ha*)$_{φ}$)$_{ι\ /P}$ ((*gong oi shid pon*)$_{φ}$)$_{ι\ /P}$ S B S B S S S B				
☞c.	((*lo moi*)$_{φ}$ (*teu ha*)$_{φ}$ (*gong oi shid pon*)$_{φ}$)$_{ι\ /P}$ S B S B S S S B				
d.	((*lo moi*)$_{φ}$)$_{ι\ /P}$ ((*teu ha*)$_{φ}$ (*gong oi shid pon*)$_{φ}$)$_{ι\ /P}$ S B S B S S S B				*!
e.	((*lo moi*)$_{φ}$ (*teu ha*)$_{φ}$ *gong*)$_{ι\ /P}$ ((oi shid pon)$_{φ}$)$_{ι\ /P}$ S B S S S S B	*(!)	*(!)		

In (24), candidate (c) can be ruled out by any of the higher-ranked constraints, and then candidates (a) and (b) are both selected. In (25), candidate (e) can be ruled out either by star sandhi tone pause or align base tone; candidate (d) is ruled out by I-rhythm, and then candidates (a), (b) and (c) emerge.

5. Alternative analysis

At this moment, we can consider an alternative analysis using Selkirk's Match theory. This theory proposes that there is a tendency for prosodic constituents to mirror syntactic constituents. In this sense, the syntax-prosody match is no longer a matter of alignment but operates on correspondence. Selkirk's match constraints are given in (26), which require a match between phonological phrase and XP, and between intonational phrase and clause. A constraint ranking can be posited as (27).

(26) Constraints

a. MATCH(φ, XP): assign one violation mark for every phonological phrase, φ, that is not matched by a corresponding XP.
b. MATCH(XP, φ): assign one violation mark for every XP that is not matched by a corresponding phonological phrase, .
c. MATCH(ι, CL): assign one violation mark for every intonational phrase, ι, that is not matched by a corresponding clause.
d. MATCH(CL, ι): assign one violation mark for every clause that is not matched by a corresponding intonational phrase, ι.

(27) Constraint ranking
*S/P >> MATCH(CL, ι) >> MATCH(ι, CL) >> ι-RHYTHM

Tableaux (28) and (29) show that under this analysis, an illegal output can be wrongly selected, while the real optimal output can be wrongly ruled out. The symbol 👎 indicates a wrongly selected output, while the symbol *☞ indicates the real optimal output that is wrongly ruled out.

(28) [*lo moi*]^{NP} [*da* [*ten fa*]^{NP}]^{VP}

		*S/P	MATCH (CL, ι)	MATCH (ι, CL)	ι-RHYTHM
*☞a.	((*lo moi*)$_{φ}$)$_{ι\ /P}$ ((*da ten fa*)$_{φ}$)$_{ι\ /P}$ S B S S B		*!	**	
☞b.	((*lo moi*)$_{φ}$ (*da ten fa*)$_{φ}$)$_{ι\ /P}$ S B S S B				
c.	((*lo moi*)$_{φ}$ *da*)$_{ι\ /P}$ ((*ten fa*)$_{φ}$)$_{ι\ /P}$ S B S S B	*!	*	*	

(29) [*lo moi*]^{NP} [*teu ha*]^{AdvbP} [*gong* [*oi* [*shid pon*]^{NP}]^{VP}]^{VP}

		*S/P	MATCH (CL, ι)	MATCH (ι, CL)	ι-RHYTHM
*☞a.	((*lo moi*)$_{φ}$)$_{ι\ /P}$ ((*teu ha*)$_{φ}$)$_{ι\ /P}$ ((*gong* (*oi* (*shid pon*)$_{φ}$)$_{φ}$)$_{φ}$)$_{ι\ /P}$ S B S B S S S B		*!	**	
*☞b.	((*lo moi*)$_{φ}$ (*teu ha*)$_{φ}$)$_{ι\ /P}$ ((*gong* (*oi* (*shid pon*)$_{φ}$)$_{φ}$)$_{φ}$)$_{ι\ /P}$ S B S B S S S B		*!	*	
☞c.	((*lo moi*)$_{φ}$ (*teu ha*)$_{φ}$ (*gong* (*oi* (*shid pon*)$_{φ}$)$_{φ}$)$_{φ}$)$_{ι\ /P}$ S B S B S S S B				
d.	((*lo moi*)$_{φ}$)$_{ι\ /P}$ ((*teu ha*)$_{φ}$ (*gong* (*oi* (*shid pon*)$_{φ}$)$_{φ}$)$_{φ}$)$_{ι\ /P}$ S B S B S S S B		*!	**	*
e.	((*lo moi*)$_{φ}$ (*teu ha*)$_{φ}$ *gong*)$_{ι\ /P}$ ((*oi* (*shid pon*)$_{φ}$)$_{φ}$)$_{ι\ /P}$ S B S S S S B	*!	*		

6. Conclusion

An intonational phrase consists of one or more phonological phrases, and thus ends in a base tone. Intonational phrasing is not sensitive to the formation of sense unit. A pause cannot occur after a sandhi tone. The corpus shows that the length difference between the established ι's within an utterance is no more than three syllables. I have proposed a constraint ranking under the standard prosodic theory, and show that the match theory is not well-motivated, and may result in incorrect predictions.

ACKNOWLEDGEMENT

This paper presents partial findings of the research project, MOST# 105-2410-H-004-181-MY2. I would like to thank Ministry of Science and Technology for the sponsorship.

REFERENCES

CHEN, MATTHEW Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4:109-150.

CHEN, MATTHEW Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.

HAYES, BRUCE. 1989. The prosodic hierarchy in meter. *Phonetics and Phonology* 1:201-260.

HSIAO, YUCHAU E. 1991. *Syntax, Rhythm and Tone: A Triangular Relationship.* Doctoral Dissertation. University of California, San Diego. Taipei: Crane Publishing Co., Ltd..

HSIAO, YUCHAU E. 1995. *Southern Min Tone Sandhi and Theories of Prosodic Phonology*. Student Book Co., Ltd., Taipei.

NESPOR, MARINA and IRENE VOGEL. 1986. *Prosodic Phonology*. Dordrecht: Foris Publications.

SELKIRK, ELISABETH. 1984. *Phonology and Syntax: the Relation between Sound and Structure*. MIT Press.

SELKIRK, ELISABETH. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3:371-405.

Proceedings of the 29th North American Conference on Chinese Linguistics (NACCL-29). 2017. Volume 1.
Edited by Lan Zhang. University of Memphis, Memphis, TN. Pages 125-140.

Analysis of Tone-Melody Relationship Problems in *Huju*

Lu Lu
College of William and Mary

This study provides a quantitative analysis of comparing the relationship between music melody and linguistic tone through different prosodic domains. The result suggests that there is a high correspondence (70%) between music melody and language tone in *Huju*, which differs from musicians' opinion that the word tones have a limited restriction on the music melody of *Huju*. Besides, the fewer tones a language has, there tends to be fewer possibilities to link tone and music melody. Most of the mismatch conditions happen when the linguistic property is sacrificed to satisfy the music property in *Huju*. The study tries to fill the vacancy and provides a preliminary model of studying the tone-melody relationship in tonal languages.

0. Introduction

Language and music are the two qualities of which we human beings are born with. There are at least three research perspectives from which a possible relationship between language and music has been studied. The most general studies have focused on how the language and music are intimated related. The two domains might share a common origin with respect to the development of human brain. Although music and speech have different specialized representations, such as interval in music and speech categories in language, these elements share basic sound processing mechanisms. Another group of scientists is committed to studying the differences between music and language. They argue that music lacks the semantic accuracy and systematic grammar of the language, or assume that music can be usefully analyzed using linguistic models. The third approach examines the overlap, interface or the relations between the two elements, which is also the research interest of this paper. It should be noted that the analyses of such studies are based on the concept of the shared origin of music and language.

For tonal languages like Chinese, to investigate the relationships between language tones and music melodies is always the focus. As the first linguist who proposed five-scale tonal representation system, Chao (1956) compares the pitch level in tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition. He claims that different styles of singing or chanting forms in Chinese matched with language tones. The tone should be considered as one essential part of singing, and it is better for composers to involve language tones into their consideration when composing songs. Therefore, to analyze the music-language relationships in tonal

languages, the following questions need to be answered: 1) Does music completely follow speech-melody in tonal languages? 2) Is there a clash between melodic patterns based on speech and purely musical melodic patterns? 3) If there is a clash, what are the reasons behind it? In other words, what causes such clash, but more importantly, under what circumstances will that clash happen?

Different research methods are used to quantify the correspondence, such as native speaker observation, note shape/tone shape comparison, F_0(pitch) comparison and contour comparison. Previous researches focused a lot on the relationship between two adjacent syllables or the adjacent words with its related musical notes. The core idea is based on the fact that "the expression of tone in speech is always in relations to what came before and what comes after it" (Wee 2007). Nevertheless, previous studies in Chinese languages mainly focus on Cantonese songs/operas and Mandarin pop songs. The findings indicate that the fewer tones a language has, there tends to be fewer possibilities to link tone and music melody. Chinese scholars have only studied Mandarin (4 tones) and Cantonese (7 tones) music, while no researchers have ever done studies exploring the tone-melody relationship with the Shanghai dialect (5 tones), which has a moderate tonal system somewhere between Mandarin and Cantonese.

Nespor and Vogel (1986, 2007) claim that it is the prosodic constituency rather than the syntactic structure that accounts for the ability of listeners to understand the ambiguous sentence. However, the prosodic constitutes have never been involved in the research in this area among Chinese languages. Therefore, this paper aims to answer the above questions through studying the linguistic tone and music melody mapping problems in *Huju* through prosodic domains.

1. Background

1.1 Studies on Cantonese and Mandarin

Among all the existing tone-melody relationship studies of Chinese language and music, studies on Cantonese Opera and pop songs are relatively thorough. Those studies suggest that the degree of conformity between language tone and music melody varies with languages and genres.

Chao (1956) first posits a hierarchy of song types in Mandarin: singsong, chanting, recitative, tonal composition, and atonal composition. Singsong have the most correspondence with tone while the atonal compositions, such as contemporary songs, have the least correspondence between tone and melody. Wong and Dieh (2002)'s perception experiment on four Cantonese songs finds that the songwriters abandon the ratio scale of F_0 (fundamental frequency) difference while native Cantonese–speaking listeners still apply an ordinal F_0 scale to understand the lyrics when listening to the music. The correspondence between music and lexical melodies is near 91.81%. Schellenberg (2009) finds that Cantonese and Mandarin singers employ different approaches in the manifestation of tone in singing, but these strategies match those used by composers in the two languages. For example, Mandarin listener cannot identify

individual sung words out of context, but the listeners in Cantonese can use pitch and contour to recognize it. A match happens when the transition from one syllable to the next moves in the same direction as the transition from the note on which the first syllable is set to the note on which the second syllable is set, while a mismatch happens when the transition goes in the opposite direction. Besides, Wee (2007) proposes that headship in music and linguistics is the prominent features when preserving tonal integrity in music.

1.2 Tones in *Huju*

Huju, also known as Shanghai Opera, is a Shanghai dialect based regional opera, which is popular in the Yangtze River delta. There are three major aria singing types in *Huju*, including *changqiang* 长腔, *huang qiang* 簧腔 and *xiaodiao* 小调. According to You (2006), there are 28 initials and 41 vowels in Shanghai Opera. The only differences found between tones in *Huju* and tones in Shanghai dialect are the *yangqu* and *yinru* tones. But the differences are too minor that it can be disregard. Therefore, I adopt the same tonal system of Shanghai dialect to analyze the *Huju* in this paper (see Table 1).

Table 1 Five tones in Shanghai dialect

Tone Types	Pitch Level	Citation Tone	Base Tone
Yinpin	53	HL	HL
Yinqu	34	MH	MH
Yangqu	13	LM	LH
Yinru	5	Hq	MH
Yangru	12	LMq	LH

1.3. Basics of Tone-sandhi rules in Shanghai Dialect

Shanghai dialect is known for word-tone sensitive language; therefore, all the lexical items undergo tone-sandhi change. There are two types of tone-sandhi in Shanghai dialect: 1) Guang-yong shi bian-diao 广用式变调 (Tone-sandhi in broad used form), and 2) Zhai-yong shi bian-diao 窄用式变调 (Tone-sandhi in narrow used form) (Zhang 2017). No matter what kinds of tone-sandhi rules are used, all but the leftmost morpheme lose their underlying base tones. These base tones are associated in a one-to-one fashion, from left to right, across the entire tone-sandhi domain (Sherard 1972; Yip 1980; Zee & Maddieson 1980; Wright 1983; Selkirk & Shen 1990; Duanmu 1992). This paper will adopt the idea that Shanghai dialect has only three base tones and their corresponding relationships to citation are listed below.

The process of tone sandhi change in Shanghai dialect can be concluded as follow,

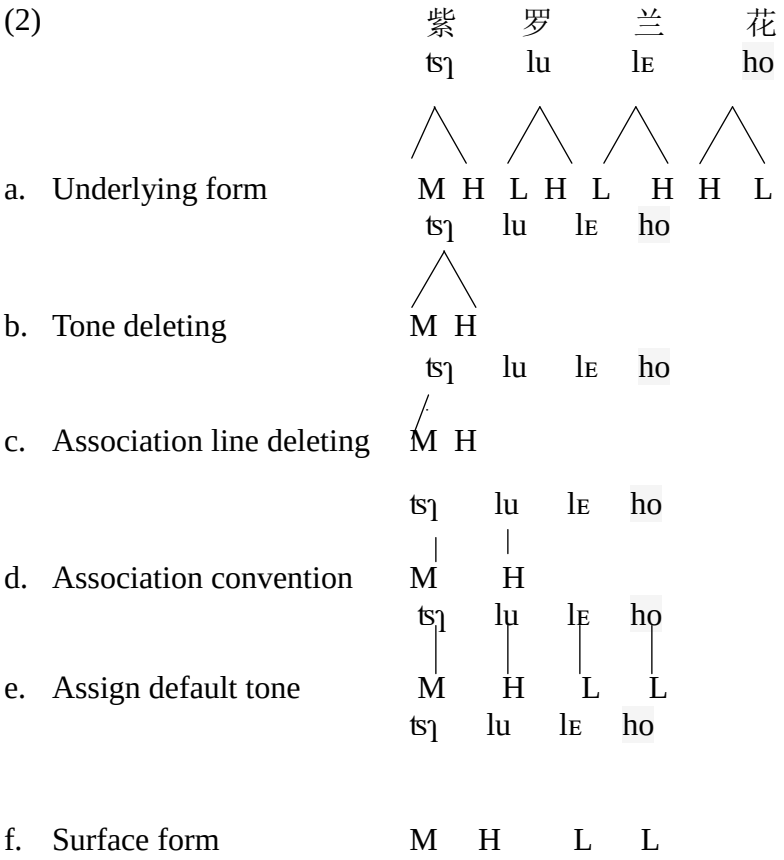
(1) a. Tone deletion: In each TS domain, delete all underlying tones, except an initial syllable

b. Associate convention: associate tones to syllables one to one from left to left

c. Default tone: assign default tone L to the remaining syllables.

An example of TS rule application is provided below in (2),

(2)



2. Prosodic Studies in Shanghai Dialect

2.1 Selkirk and Nespor and Vogel

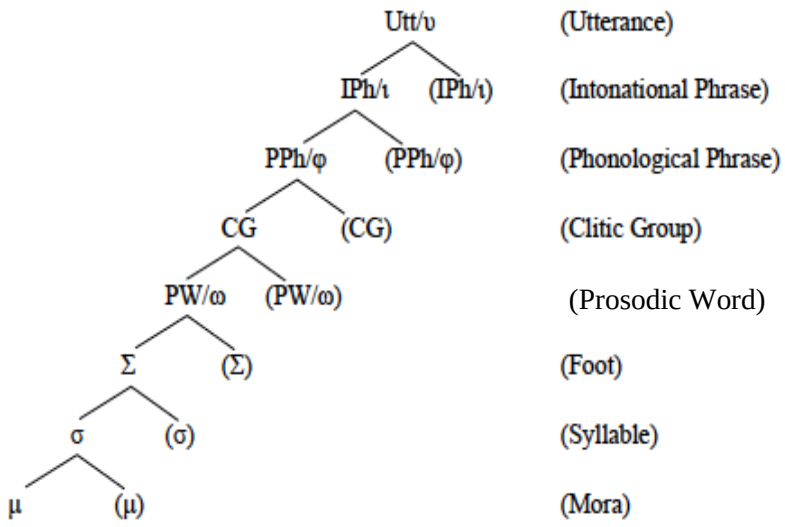
Prosodic phonology is considered as the representative theory of the interface studies between phonology and other linguistic components, such as syntax, semantics, etc. Selkirk (1978) develops the first model of prosodic phonology, which is a six-layer prosodic hierarchy, including Utterance (Utt /υ), Intonational Phrase (IP /ι), Prosodic Phrase (PPh /φ), Prosodic Word (PW /ω), Foot (Ft/ Σ), and Syllable (Syl/ σ). According to Selkirk and Shen (1990), the prosodic structure assigned to a sentence in Shanghai dialect is the minimal structure consistent with the well-formedness constraint and the

mapping rules. For any level of prosodic structure, only the bracketing of the sentence into constituents of one particular X-bar type is relevant and precisely only at the left or right edge of such a constituent. In Shanghai dialect, Selkirk and Shen argues that the left edge of a syntactic word belonged to the categories, noun, verb, or adjective (=a "lexical item") always coincide with the edge of a prosodic word, so the syntax-phonology mapping rule in Shanghai dialect is listed in (3),

（3） Prosodic Word: {Left, Lex^0}, where Lex^0 stands for word belonging to the lexical categories N, V, A

Nespor and Vogel's (1986) book is another authoritative work in prosodic phonology. Nespor and Vogel propose their prosodic constituency hierarchy: the syllable, the foot, the phonological word, the clitic group, the phonological phrase, the intonational phrase, and the phonological utterance.

Figure 1. Prosodic hierarchy (Nespor and Vogel)



2.2 Strict Layer Hypothesis

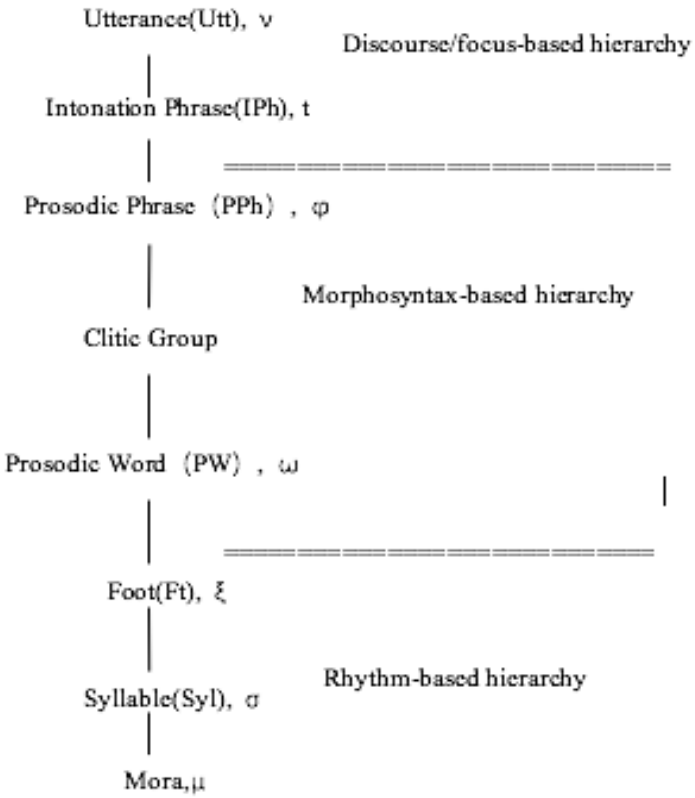
The only well-formedness condition on prosodic constituency is the Strict Layer Hypothesis (SLH) proposed by Selkirk (1984,1986), Nespor and Vogel (1986) and Hayes (1989). It assumes that a prosodic unit of a given level n immediately dominated the unit of the lower level n-1, and is exhaustively contained in a constituent of the immediately higher level n+1. Therefore, the recursive prosodic structure does not occur. For example, a prosodic word can only contain foot, and may not include prosodic word or clitic group or phonological phrase.

Later on, within the OT framework, Selkirk (1996) proposes four general constraints for Strict Layer Hypothesis (Hereafter SLH). Selkirk holds that the constraints of Layeredness and Headness are inviolable and should not be dominated in the constraint ranking universally, while the Exhaustivity and Non-Recursivity are not observed by all languages. SLH is a well-formedness condition on the prosodic constituency, however, it may not be the universal principle for all the languages. For example, the prosodic recursivity is allowed in Mandarin Chinese and Shanghai dialect, (Ladd 1986,1990; Hyman et al. 1987; Odden 1987; and others).

2.3 Zhang (2017)

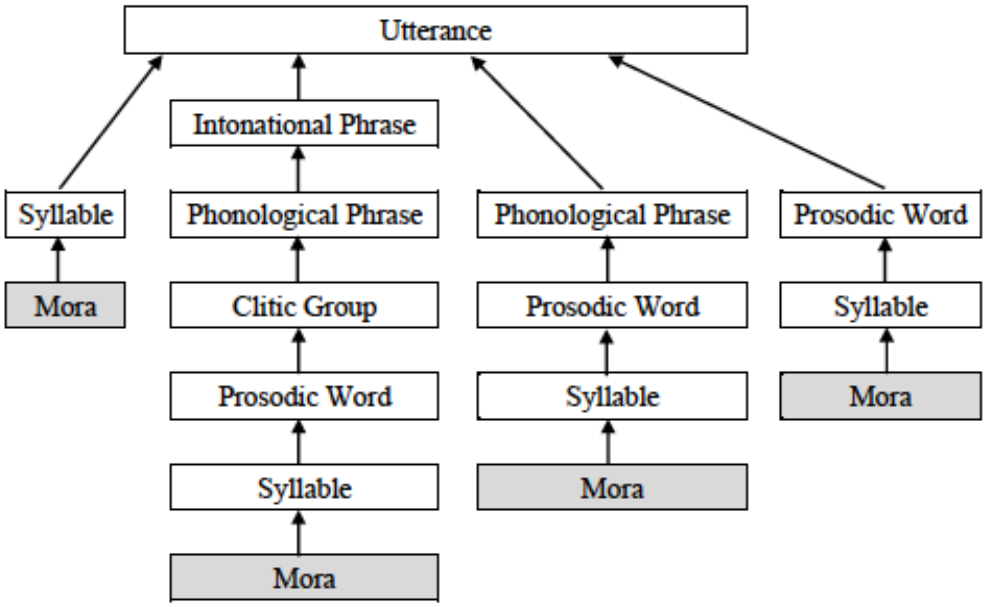
Zhang (2017) proposes a trisected model for prosodic hierarchy.

Figure 2. Prosodic hierarchy (Zhang)



He also agrees that SLH is not applicable in all languages and the prosodic recursion is allowed. A prosodic hierarchy, which entails four possible types of organization of prosodic constituents in Mandarin Chinese is proposed. Level-skipping or the violation of Exhaustivity is allowed.

Figure 3. Prosodic hierarchy in Mandarin Chinese



As for Shanghai dialect, according to Zhang, regardless of its internal structure, it must undergo tone sandhi. But the case will become more complicated when they involve function words. In contrast to the lexical items, not all the function words will form a TS domain with the lexical item to its left. As for SLH, Zhang proposes that prosodic recursivity is prohibited between the units of different hierarchies (language universal), but optionally in the units of the same hierarchy (language specific), therefore, the violation of constraints of Exhaustivity, Nonrecursivity, and Layeredness may all be allowed in the Shanghai dialect.

In conclusion, the account of the phonological system of the Shanghai dialect developed in this paper will be based on the framework of the aforementioned prosodic phonology theory. In the following analysis, PW+PW can be dominated by another PW', CG+CG can be dominated another CG'; However, PPh+PPh cannot be dominated by PPh', instead it must be dominated by its higher level IP, because IP and PPh belong to different hierarchies. Therefore, the following two layers are acceptable in my analysis,

(4)

3. Analysis on Shanghai Opera

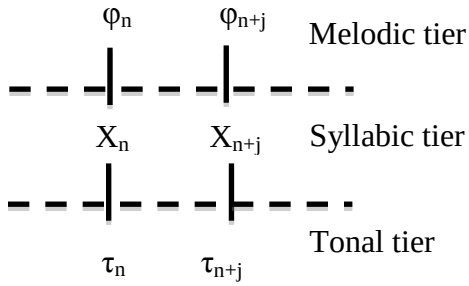
3.1 Data Collection

The main data/music scores used in this paper come from the book 沪剧唱段 108 首 [108 collections of Huju pieces], published by Shanghai Music Press. Currently, there is no official corpus of Chinese dialectical opera for academic study. This book is considered as the most authoritative music collection of *Huju*. All the transcriptions are based on the music track provided by the book. Ten selected pieces represent the three aria singing types in Shanghai Opera.

3.2 Overview

Each music piece is analyzed through three tiers: melodic tier, syllable tier and tonal tier.

Figure 4. Three Tiers



The syllable X_n represents the syllables in texts, and syllables X_nX_{n+j} forms a prosodic domain.

τ_n in tonal tier represents the linguistic tonal feature associated with syllables X; τ_{n+j} represents the tonal feature associated with syllables $X_{n+j.}$

φ_n in melodic tier, which is associated with the syllables X_n, represents the musical note pitch at the edge; φ_{n+j} represents the musical note pitch at the edge associated with the syllables $X_{n+j.}$

Therefore, the following metric is proposed,

(5) Within a prosodic unit, the syllable $X_n X_{n+j}$[1]

Music melody and language tones have an optimal match

When $\tau_n<\tau_{n+j}$, $\varphi_n <\varphi_{n+j}$, or

$\tau_n>\tau_{n+j}$, $\varphi_n >\varphi_{n+j}$, or

$\tau_n=\tau_{n+j}$, $\varphi_n= \varphi_{n+j}$,

[1] The quantity of J depends on the number of syllables within a single prosodic unit.

Music melody and language tones have a non-optimal match,

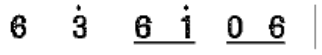
When $\tau_n<\tau_{n+j,}$, $\varphi_n= \varphi_{n+j}$ or
$\tau_n>\tau_{n+j}$, $\varphi_n=\varphi_{n+j,}$, or

Music melody and language tones mismatch,

When $\tau_n<\tau_{n+j,}$ $\varphi_n>\varphi_{n+j}$ or
$\tau_n>\tau_{n+j}$, $\varphi_n<\varphi_{n+j,}$, or
$\tau_n=\tau_{n+j}$, $\varphi_n>\varphi_{n+j}$, or
$\tau_n=\tau_{n+j}$, $\varphi_n<\varphi_{n+j}$,

Briefly speaking, to get matched between music and language tier, if τ_n is lower than τ_{n+j}, the φ_n cannot be higher than the φ_{n+j}; if τ_n is higher than τ_{n+j}, φ_n cannot be lower than φ_{n+j}.

3.3 Sample Analysis

In this section, a sample analysis of the piece 为你打开一扇窗 'Open a window for you' will be provided. Appendix I is a numbered musical notation score[2]. The numbered musical notation is widely used in music publication in China. As for octave transcription, I used subscript b, $_b$, to represent the low key, and the superscript #, $^{\#}$, to represent the high key for convenience. For example, 

, I transcribed as 6 $3^{\#}$ 6 $1^{\#}$ in the music tier. As for the transcription of the syllabic and tonal tiers, I used the following line as an example,

[2] Number musical notation score is known as *jianpu* 简谱. The number 1 to 7 represents the scale degrees. The number always correspond to the diatonic major scale. For example, in the key of C, their relationship with the notes and the solfege is as follows

Note: C D E F G A B
Solfege: do re mi fa sol la si
Notation: 1 2 3 4 5 6 7

In addition, dots above the or below a musical note raise or lower it to other octaves. The number of dots equals to the number of octaves. For example, $\dot{1}$ is an octave higher than 1, $\dot{7}$ is an octave lower than 7

Since it is a song with a time signature of 4/4, the line ‘Open a window for you’ is a full musical phrase, which includes four music bars.

Within a Prosodic Word, 打开 tã k^hE ‘open’

(6)	tã	k^hE
	打	开
Gloss	beat	open
BT	MH	HL
ok	M =	H

Therefore, the full analysis of this PW is listed as follows,

(7) Melodic tier φ_n =6 $\varphi_{n+1=}$ 65 $\varphi_n > \varphi_{n+1}$

Syllabic tier [打 开] $_{pw}$

X X+1 j=1

Tonal tier τ_n =M τ_{n+1} =H $\tau_n < \tau_{n+1}$

Result Mismatch

Example 8 shows that within the prosodic word tã k^hE ‘open’, the music melody presents a descending trend since the musical note of tã is 6 and the ending edge note of k^hE is 5. As for the linguistic tonal contour change, the word undergoes a TS change, therefore it has an ascending trend in the tonal tier (M=H). According to metric (16), the melody and tonal tier mismatch with each other within the prosodic word unit.

Analogously, the analysis of the phonological phrase, 为你打开 ɦuE ŋi tã k^hE ‘open for you’ is listed as follows,

(8) Melodic Tier φ_n =6 $\varphi_{n+3=}$ 65 $\varphi_n > \varphi_{n+1}$

Syllabic Tier [[为]$_{pw}$ 你] $_{cg}$ [打 开]$_{pw}$]$_{pph}$

X X+3 j=3

Tonal Tier τ_n =L τ_{n+3} =L $\tau_n = \tau_{n+1}$

Result Non-optimal Match

Within the Clitic Group 为你 ɦuᴇ ŋ̊i ‘for you’, both the music and tonal tier show an ascending change, therefore, the tone and melody have an optimal match. Although *ɦuᴇ ŋ̊i* together as clitic group has an optimal match, the *ŋ̊i* as a clitic pronoun, has a mismatch between tone and melody. Within the phonological phrase, there is a descending trend in melodic tier since only the edge musical notes are considered, while the tonal tier doesn’t show any change. Therefore, tone and melody has a non-optimal match.

The summary of the whole sentence is listed as follows

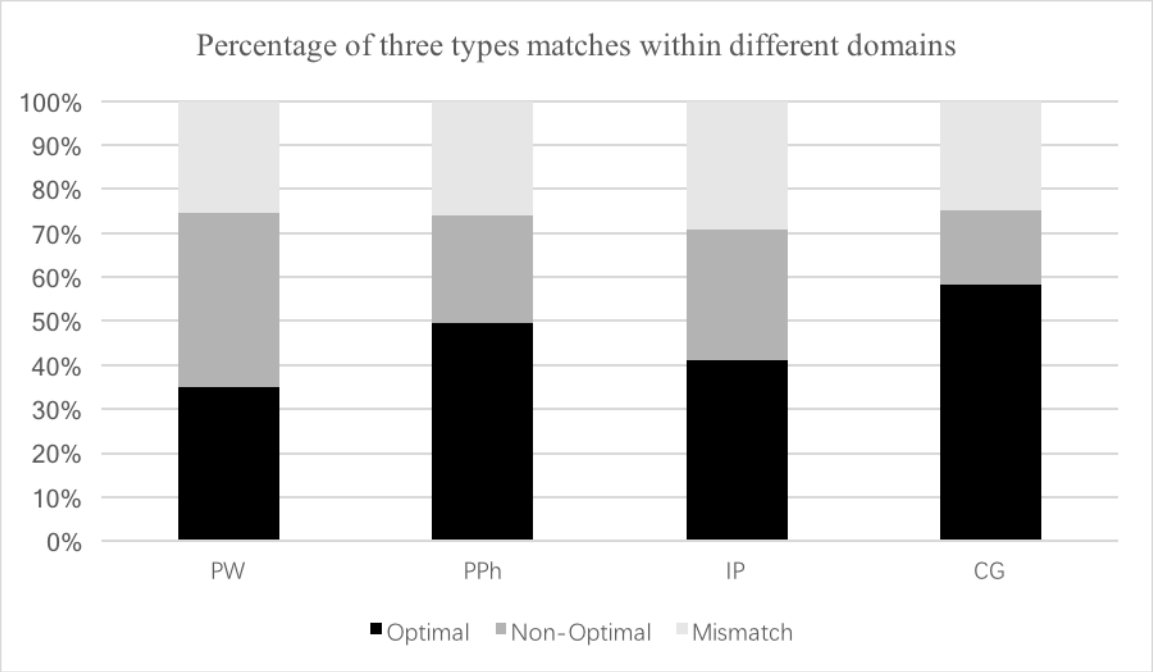
(9)	ɦuᴇ ŋ̊i tã kʰᴇ iɪ sø tsʰɑ̃	Match
Lyrics	为 你 打 开 一 扇 窗	
	[[[[为]你]$_{cg}$ [打开]$_{pw}$]$_{pph}$ [[[一]扇]$_{cg}$[窗]$_{pw}$]$_{pph}$]$_{ip}$	
BT	LH MH MH HL MH MH HL	
Melodic tier	6 3^h1^h 6 65 1^h 1^h2^h 3^h1^h	
TS within PW	LH # MH # M= H# MH #MH# HL	ok/ * / */ok/OK/OK[3]
TS within CG	L = H M=H	OK/ OK
TS within PPh	L=H=L=L # M=H=L	ok/ok
TS within IP	L = L	*

3.4 Results

There are in total 165 lines in ten selected pieces of *Huju*. According to the aforementioned analysis metric, 74.79% of the pairs of syllables confirm to metric (Figure 4) within PW, among which 35.18% are optimal match and 39.61% are non-optimal match. 74.84% confirms with CG, among which 57.86% are perfect match, and 16.98% are non-optimal match; 73.89% confirms within PPh, among which among which 49.49% are perfect match, and 24.40% are non-opposing match. 70.91% confirms within IP, among which 41.21% are perfect match and 29.09% are non-optimal match. Figure 14 is the conclusion chart of the data. Such findings suggest that the music pitch depicts with language tones in a high percentage in *Huju* within different prosodic domains. The CG group has the highest percentage of optimal match, followed by PPh, IP and PW. A summary of the percentage of optimal match, non-optimal match and mismatch are shown in Figure 5

Figure 5. Percentage of three types matches within different prosodic domains

[3] Method of Transcription: OK=optimal match; ok=non-optimal match; *=mismatch



A one-way within subjects group ANOVA with a Greenhouse-Geisser correction was conducted to compare the match situation of music melodys and language tones withinPW, CG, PPh and IP. There was a significant difference on the tone-melody parallel situation between different prosodic domains, $F(2, 162) = 60.417$, $P<0.05$. Post hoc tests using the Bonferroni correction reveals that the tone-melody match condition within PW are significant different from those within CG, PPh and IP, $p<0.05$, which indicates that for those tone-melody match within PW are not necessarily match within CG, PPh and IP. Similar findings were found that tone-melody match situation within CG are significantly different from PPH and IP, $p<0.05$, and the match condition within PPh are significantly different from IP, $p<0.05$. See Table 2 for the summary of the results.

Table 2. Summary of parallel conditions between different prosodic domains

	t	Sig(2-tailed)
PW vs. PPH	6.363	.000
PW vs. IP	11.962	.000
PW vs. CG	9.848	.000
PPh vs. IP	9.296	.000
PPh vs. CG	5.057	.000
IP vs. CG	2.094	.000

The data suggest that there are variations in tone-melody relationship in *Huju*. The music melody does not always correspond to language tone although there is a high

percentage of correspondence. In regardless of the tone-melody condition within the prosodic word or the phonological phrase or the intonation phrase, the musical melody is greatly influenced by the tones of Shanghai dialect. However, for those tone-melody match within PW do not necessarily match within CG, PPh or the IP.

4. Discussion

Nespor and Vogel (1983) suggest that it is not the syntactic constituents, but rather the prosodic constituents that provide the relevant information in the first stage of processing of a given string of speech. However, the analysis results suggest that prosodic constituent is not sensitive in the music speech, at least for the tone-melody relationship. The effect that influences people's understanding of the lyrics in *Huju* is based on the surface form of linguistic tones, not the prosodic constituents. In addition, Zhuang (2013), as a *Huju* musician, provides a comment based on the musician intuition that the word tones have a limited restriction on the music melody in *Huju*. However, the linguistic analysis seems to provide an opposite evidence.

On the one hand, although the findings reveal that prosodic units do not restrict the tone-melody relationship in *Huju*, there is still a high percentage of match concerning the contour shape between music and language. Compared with the 91.81% tone-melody match in Cantonese music, and entirely mismatch condition in Mandarin pop songs, the results confirm with the hypothesis that fewer tones in a language, there are also smaller probabilities of a link between tone and melody. On the other hand, although 70% is considered a relatively high correlation, there is still a 30% clash between music and language in *Huju*. Unfortunately, there is no regulated rules have been found from the mismatch conditions. However, some examples indicate that the mismatch cases are likely to happen when the music form plays a more prominent role than languages form in a music piece. For example, a preference has been found that *changqiang* style has fewer mismatch groups than *xiaodiao* and *huangqiang* within the domain of prosodic word. Compared with *xiaodiao* and *huangqiang*, *changqiang* is one of the arias types that have the most chanting patterns in singing, so it has more linguistic property. In addition, most of the mismatches of clitic group happen on lining word. Lining words are only used in singing, of which music property is more prominent than linguistic property. Therefore, the mismatch conditions happen in *Huju* when the linguistic property is sacrificed to satisfy the music property in a music piece.

5. Conclusions

This paper provides an analysis to investigate the match situation between music melody and linguistic tone within different prosodic domains. The study tries to present a new perspective of adopting the scientific way to reveal the tone-melody relationship in *Huju*. The comparison is based on the linguistic concept—prosodic constituency—including Prosodic Word(PW), Clitic Group(CG), Phonological Phrase(PPh), Intonational Phrase(IP) and Foot(Ft). The findings support the hypothesis that the more

tones in a language, there is a higher probability of a link between language tone and music melody. The results also differ from many musicians' opinions that the word tones have a limited restriction on the music melody of *Huju*. However, the prosodic constituents are not sensitive in the cognition of sung melody, which is different in speech cognition.

When examining the examples of match/mismatch across different prosodic domains, in spite of the fact the optimal or non-optimal match situation may be inclined to present "rules" within different aria singing types, it is hard to conclude a well-formedness rule in general. However, some mismatch cases indicate that the clash between tone and melody usually happen on those words or types of arias that include more musical property than linguistic property. If the linguistic property of a word in music piece plays an upper hand, the tone and melody are inclined to be matched; if the music feature of a word in music piece is more prior than its linguistic property, the tone and melody are tempted to be mismatched with each other.

This study remains limited when it comes to the sample size, types of language and music. Since the prosodic domains do not have a direct influence on the tone-melody relationship, what else elements affect the relationship between music and language? Also, the cognitive understanding of the music and language are different. For example, there won't be a huge divergence among people to decide if a sentence or speech sounds ungrammatical or bad, but people may hold opposite opinions when deciding if the music piece sounds appealing or good. Therefore, there are well-formedness rules in the linguistic area, is such kind of rule still applicable to music? Those questions can be all investigated in the future studies.

REFERENCES

CHAO, YUEN REN. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. *For Roman Jakobson*, 52-59.

CHOW, MINGYING. 2012. *Singing the right tones of the words: the principles and poetics of tone-melody mapping in Cantopop*. Ph.D dissertation. Hongkong University

DUANMU, SAN. 1992. End-Based Theory, Cyclic stress, and tonal domains. *Chicago Linguistics Society*, vol. 28.

NESPOR, MARINA and IRENE VOGEL.1983. Prosodic structure above the word. In Anne Cutler and D. Robert Ladd (eds.) *Prosody: Models and Measurements*, 123-140. Berlin: Springer

NESPOR, MARINA and IRENE VOGEL.1986. *Prosodic phonology*. Dordrecht: Foris.

NESPOR, MARINA and IRENE VOGEL.2007. *Prosodic phonology: with a new foreword*. Vol. 28. Berlin:Walter de Gruyter.

SCHELLENBERG, MURRAY.H. 2009. Singing in a tone language: Shona. In *Selected Proceedings of the 39th Annual Conference on African Linguistics* .137–144.

SCHELLENBERG, MURRAY.H. 2012. Tone realization in sung Mandarin. In *Proceedings of the 3rd International Symposium on Tonal Aspects of Language, Nanjing*.

SELKIRK, ELISABETH O. 1981. On the nature of phonological representation. In J. Anderson, J. Laver and T. Meyers (eds.), *The Cognitive Representation of Speech.* Amsterdam: North Holland.

SELKIRK, ELISABETH O.1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

SELKIRK, ELISABETH O. 2001 "The syntax-phonology interface." The handbook of phonological theory 2: 435-483.

SHEARD, MICHAEL. 1972. *Shanghai Phonology*. Ph.D Dissertation. Cornell University.

ELISABETH O &T. Shen.1990. Prosodic domains in Shanghai Chinese. In S.Inkelas and D. Zec(eds.), *The Phonology-Syntax Connection.* The University of Chicago Press, 313-337

HAYES, BRUCE. 1989. The prosodic hierarchy in meter. In Paul Kiparsky and Gilbert Youmans (eds.), Rhythm and meter, 201-260. Orlando, Florida: Academic Press.

WEE, LEI HUM. 2007. Unraveling the relation between Mandarin tones and musical melody. *Journal of Chinese Linguistics*, *35*(1), 128.

WONG, PATRICK C. & R. L. Diehl, 2002. How can the lyrics of a song in a tone language be understood? *Psychology of Music*, *30*(2), 202–209

WRIGHT, MARTHA.S. 1983. *A metrical approach to tone sandhi in Chinese dialect*. Ph.D. dissertation. University of Massachusetts, Amherst

YIP, MORRIS. 1980.The Tonal Phonology of Chinese. Ph.D. dissertation, MIT

YIP, MORRIS. 1989. Contour tones. *Phonology Yearbook*, vol.6.

YIP, MORRIS. 2002. *Tone: Cambridge Textbooks* in Linguistics. Cambridge University Press.

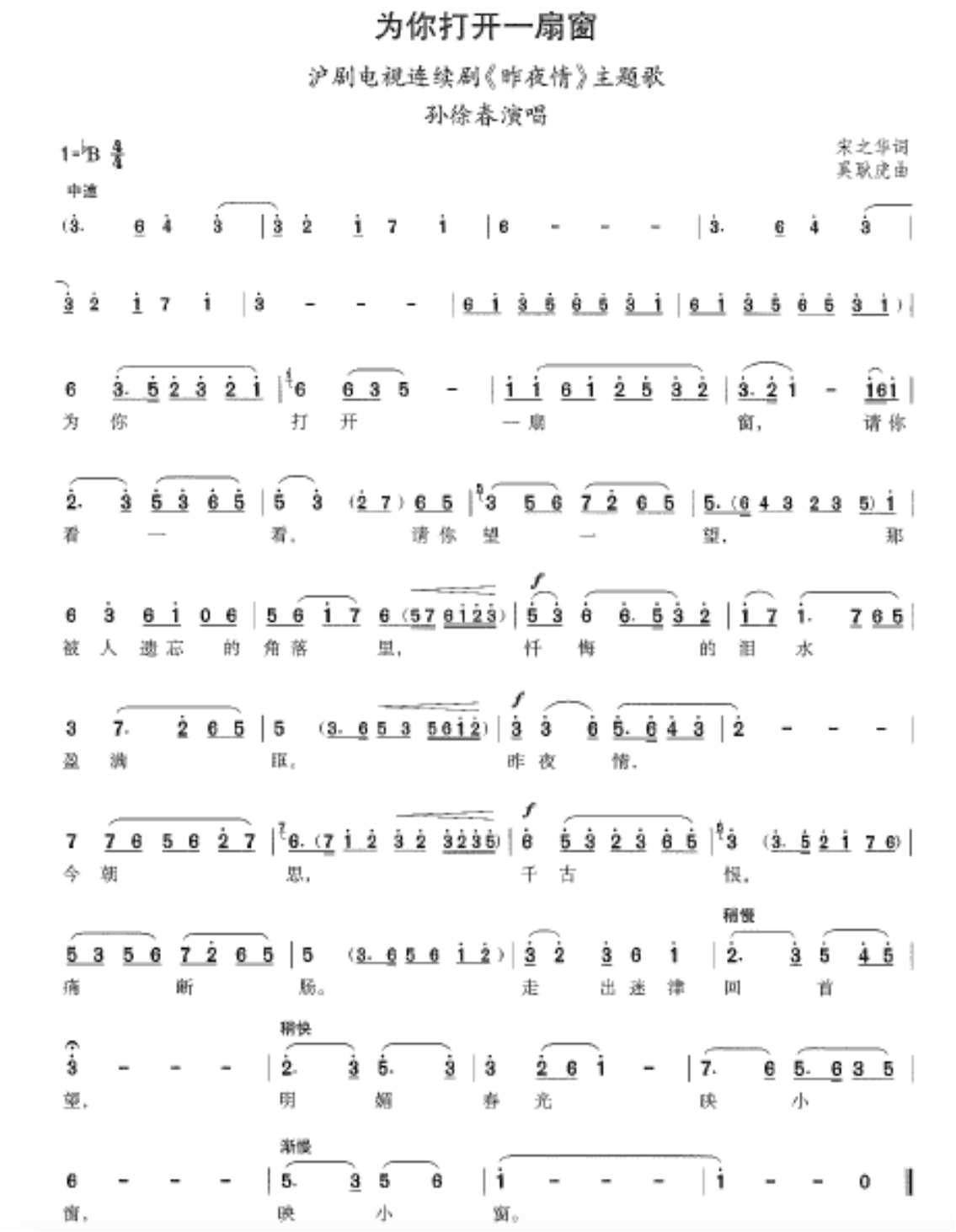
YOU, RUJIE. 2006. *Di fang xi qu yin yun yan jiu*. 地方戏曲音韵研究 [The studies of historical phonology in Chinese local operas]. Beijing: The Commercial Press

ZEE, ERICA and I. MADDIESON. 1980. Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. *Glossa*, vol.16:225-264

ZHANG, HONGMIN2017. *Syntax-Phonology Interface: Argumentation from Tone Sandhi in Chinese Dialects*, Routledge, New York

ZHUANG, YONGPING. 2013. *Huju changqiang shangxi* 沪剧唱腔赏析 [The appreciation of Huju aria]. Shanghai: Shanghai Music Press

APPENDIX I. Music score



Proceedings of the 29th North American Conference on Chinese Linguistics (NACCL-29). 2017. Volume 1. Edited by Lan Zhang. University of Memphis, Memphis, TN. Pages 141-152.

Checked Tone Merger in the Nanjing Dialect: An Acoustic Analysis

Christopher Oakden
Rutgers, The State University of New Jersey

This study provides new evidence relating to a reported tonal merger for younger speakers of the Nanjing Dialect, a Lower Yangzi Mandarin dialect. Recent impressionistic accounts (Song 2009, Gu 2015) report an in-progress merger of checked tone syllables containing a glottal stop coda, first by coda deletion and then syllable lengthening. In the current study, an acoustic analysis of checked tone syllable realizations was performed on a younger speaker of the dialect. Rime duration and vowel creak data extracted from recordings conflict with earlier characterizations of the merger. Lack of vowel creak (glottal coda residue) on canonically checked syllables indicates outright elimination of glottal stops, while a significant duration difference between checked tone syllables and other lexical tones is retained.

0. Introduction

A topic of enduring interest in discussions of Chinese phonology is the status of checked tone syllables, generally distinguished by the presence of an occlusive coda (Chen 2000:5) rather than a specific tonal melody. Understood as a relic of Middle Chinese phonology, the checked or entering tone (*rusheng*) category is attested across the majority of modern Chinese dialect families, but is mostly absent from the Mandarin dialects. Certain Mandarin subdialect classes do retain it, however, such as Lower Yangzi Mandarin or *jianghuai guanhua.*

One such example is the Nanjing dialect (Fei & Sun 1993, Liu 1995; henceforth NJD); earlier accounts describe it as a five-tone system, with one lexical tone category populated by checked tone syllables containing a glottal stop coda. Recent studies of the dialect (Song 2009; Gu 2015) suggest, though, that checked tone syllables have begun merging into other lexical tone categories, first by coda deletion and subsequent syllabic lengthening. Gu (2015) reports that the merger is not uniform across generations of speakers, and is reported to be more advanced in younger speakers (those between the ages of 16 and 25 years old).

The current study investigates the reported merger in NJD with the aid of a phonetic experiment. Specifically, it quantifies the extent of the merger by narrowing in on two acoustic parameters that distinguish checked tone from other tone categories: rime duration and vowel creak. Two main results of the experiment disagree with earlier impressionistic accounts of the merger; while a significant duration contrast remains

between checked tone rimes and those of other lexical tones, there is no evidence of creak on vowels to suggest residue from a glottal coda.

The remainder of the paper is organized as follows: Section 1 introduces the tonal phonology of the Nanjing dialect and reviews earlier reports of merger, focusing on the predictions these accounts make in regard to rime duration and vowel creak. Section 2 outlines the design of the acoustic experiment and discusses the necessary controls implemented. Data analysis and results from the experiment are presented in Section 3, while their implications and directions for further research are explored in Section 4. Section 5 concludes.

1. Background

1.1 Nanjing Dialect Tone

This study focuses on the Nanjing dialect, a Lower Yangzi Mandarin dialect spoken in Nanjing, the capital of Jiangsu province, and surrounding districts. Its tonal inventory consists of five lexical tones. One is a checked tone, distinguished by high, level pitch, short duration, and the presence of a glottal stop coda. Middle Chinese (MC) categories are given a numeral designation, which will be used throughout the remainder of the paper.

MC Category	**Description**	**Melodic Representation**	**Chao Tone Letter**
Yinping (1)	Falling contour	[HL]	31
Yangping (2)	Rising contour	[LH]	24
Shangsheng (3)	Low level	[LL]	11
Qusheng (4)	High level	[HH]	44
Rusheng (5)	High level checked	[H]	5

Table 1: Tonal Inventory of Nanjing Dialect

Earlier descriptive accounts of NJD make no mention of the instability of the checked tone category (Fei & Sun 1993, Liu 1995); however, recent accounts (Song 2009, Gu 2015) report that checked tone syllables have begun to merge into other lexical tone categories for NJD speakers.

1.2 Checked Tone Merger

What sets checked tone merger apart from other tonal merger processes is that it entails more than a simple change in melodic realization of syllables. Instead, checked tone merger is a complex of three primary contrast neutralizations, described schematically below.

(1) a. syllable shape (presence of occlusive coda vs. absence)
b. duration (short *cusheng* vs. long *shusheng*)
c. melodic realization (5 vs. 31, 24, 11, 44)

As a result, a degree of possible variability exists in terms of how the merger could proceed.

Two previous studies have identified an in-progress merger for NJD checked tone syllables, each decomposing the trajectory of the merger into distinct stages. These stages make predictions about the phonetic realization of checked tones as the merger progresses, which allows for testable hypotheses regarding the correlation of certain acoustic cues.

Song (2009) identifies three stages of merger. The first stage comprises glottal coda weakening accompanied by gradual lengthening of the rime. By the second stage, the glottal stop coda has neutralized completely, and duration of checked syllables becomes indistinguishable from those of non-checked syllables. However, these syllables retain a unique melodic realization. The final stage predicts complete merger into other tonal categories producing a four-tone system. Song argues that the current state of the dialect is at an intermediary phase between stages one and two.

A more recent impressionistic study by Gu (2015) also creates a three-way partition of the merger process, but narrows its focus on both generational differences among speakers and the effects of Standard Mandarin on diachronic change. In the analysis, the first stage represents a pre-merger state characterized by a robust glottal coda and a clear durational contrast between checked and non-checked tones. This is followed by coda weakening/neutralization and rime lengthening. No distinct melodic realization remains at this point of the merger; formerly checked syllables assume the tonal melodies of other lexical tones in the dialect. The final stage proposed in (Gu 2015) involves a total supplanting of NJD melodies on checked tone syllables with Standard Mandarin melodies. In that study, survey participants between the ages of 16 and 25 showed signs of being at a more advanced stage of the merger than older speakers.

Despite differences in the projection of merger, both studies converge on the same crucial generalization regarding the glottal coda and rime duration: lengthening accompanies coda neutralization. Phonetically, the correlation of these two phenomena is expected; lack of full glottal closure results in concomitant creak on the preceding vowel and compensatory lengthening (Garellek 2013).

These generalizations give rise to testable hypotheses. If earlier accounts are on the right track, there should be a clear correlation between glottalization and duration such that checked syllables with no evidence of a glottal coda (or even residue) should be indistinguishable from non-checked syllables with respect to rime duration. In addition, this effect should be particularly evident in younger speakers of the dialect, who are reported to be at a relatively advanced stage of merger.

2. Experiment

An acoustic experiment was designed and implemented to test the hypotheses regarding NJD checked tone merger generalized from the results of Song (2009) and Gu

(2015). The primary goal was to explore the interaction of duration and glottalization on checked syllables in the speech of younger NJD speakers. To achieve this goal, a native speaker was recorded producing NJD's five lexical tones. Rime duration was compared across lexical tone categories, and canonically checked syllables were examined for evidence of glottalization.

To ensure maximally accurate duration and glottalization results, a number of controls were applied in the process of selecting target words. Targets were limited to [CV] syllable shape to aid in segmentation and minimize potential durational effects (Turk et al. 2006, van Santen 1992). Syllables with onglides [CGV], nasal codas [CVN], or a combination of the two [CGVN] were excluded from consideration due to the difficulty of identifying precise [GV] and [VN] transitions. Additionally, only targets with the low, back vowel [a] as a rime were included to control for inherent durational differences between vowels (Toivonen et al. 2015 and numerous sources cited therein). Table 2 includes the number of tokens for each tonal category and total number of checked/non-checked target words (see Appendix A for a complete list of targets).

Tone Category	**Number of Tokens**		
Tone 1	8		
Tone 2	3		
Tone 3	5		
Tone 4	8	*Non-Checked Tokens*	*Checked Tokens*
Tone 5	16	**24**	**16**

Table 2: Number/Type of Tokens

In addition to the controlled targets, sixteen fillers (eight with checked tones and eight with non-checked tones) of varying syllabic structure and segmental content were added to the list of tokens.

Two frame sentences were used to illicit vernacular speech production. The sentences described an informal classroom setting in which the subject was asked to imagine a casual conversation among students during a break between class periods. When a Chinese character is written on the blackboard, the participants in the conversation observe that the character is written in an ugly manner. Subjects read the target word in both frame sentences:

(2) a. 黑板 上 写 了 一个 ___ 字。
heiban shang xie le yige ___ zi
'On the blackboard is written a ___ character.'

b. 这个 ___ 字 写 滴 有 点儿 丑。
zhege ___ zi xie di you dianr chou
'This ___ character is written a bit ugly.'

The rationale behind selecting a written character is that it allows for isolation of a single syllable in a natural context, and facilitates construction of a frame sentence that equally accommodates any syllable, regardless of part of speech. Targets were placed sentence-medially in each frame sentence to control for phrase-final lengthening.

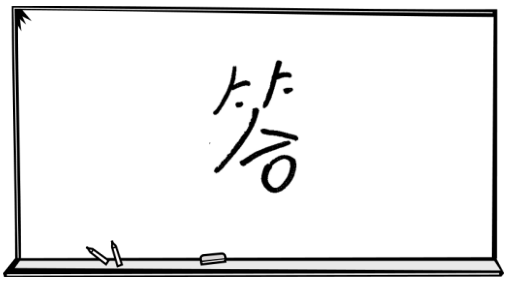
Three repetitions of the targets and fillers within the frame sentences in (2) were recorded for the experiment, totaling 144 non-checked tokens and 96 checked tokens. The order of targets was randomized and counterbalanced in each repetition, with fillers evenly spaced among stimuli. Additionally, five fillers were placed at the beginning of each repetition to accommodate any initial hyper- or mis-articulation in case the subject felt anxious about the task.

The participant in this experiment was a native NJD speaker born in Nanjing, Jiangsu, and grew up speaking the language at home. At the time of recording, the participant—a first-year undergraduate student at Rutgers University—had been living in the United States for approximately two months, but still spoke NJD daily with classmates and family members. The participant had no training in linguistics and no history of speech impairment.

Recording took place inside a sound-attenuated booth at the Phonology and Field Research Laboratory at Rutgers University. The subject was fitted with an AKG C420 head-worn microphone with behind-the-neck headband to maintain a constant distance from the mouth. The recording was made using *GoldWave* v6.10 software at a 44.1k Hz sampling rate and 16-bit quantizing rate in mono. A visual aide to the prompts was projected on a laptop screen that the subject scrolled through manually as the experiment progressed. The prompt contained an ugly hand-written character on an image of a blackboard, with the character in standard font to the left, and the frame sentences at the bottom of the screen for reference, as in (3).

(3)

答



黑板上写了一个__字。这个__字写滴有点儿丑。

All three repetitions were completed during a single recording session with a 10-15 minute break between each repetition. The next section outlines processing and analysis procedures for the data collected, as well as the results from statistical analysis.

3. Data Analysis and Results

After data collection concluded, the sound files were annotated using TextGrids in Praat (Boersma & Weekik 2015). Each file was labeled with four intervals: the vowel of the target syllable in both frame sentences, as well as a fixed interval in each frame sentence for the purpose of duration normalization. The left edge of the target vowel was identified by placing a boundary at the zero crossing of the first periodic, non-deformed waveform of the vowel (Francis et al. 2002); the right edge was determined by inserting a boundary at the end of the second formant on the vowel (Turk et al. 2006: 7). Duration and jitter (as a measure of vowel glottalization; see Koike 1973, Koike & Hirano 1973, Surana & Slifka 2006) figures were extracted from these labeled sound files with the use of customized Praat scripts. Analysis of both parameters and results are discussed separately below.

3.1 Duration Results

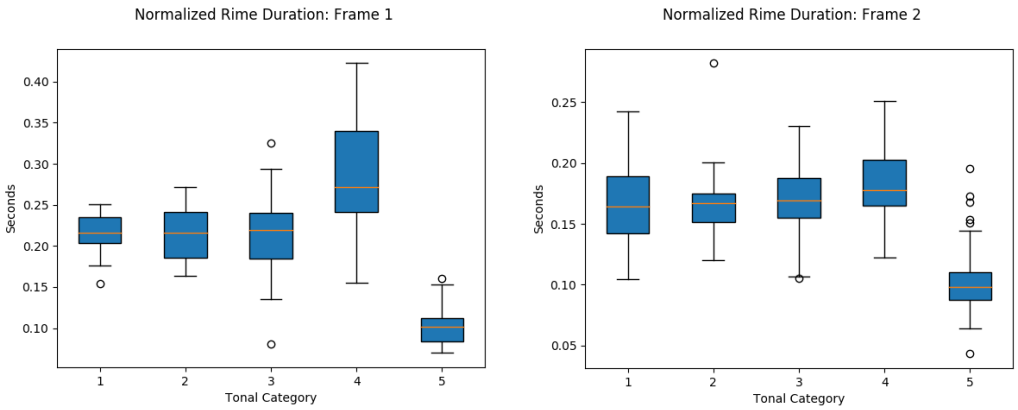
Duration measurements were normalized using a fixed point in each frame sentence to control for variation in speech rate. Table 3 outlines mean rime duration and standard deviation post-normalization.

Tone Category	Frame 1	Frame 2
Tone 1	212.7 (33.7)	161.8 (22.1)
Tone 2	214.1 (33.9)	156.8 (24.7)
Tone 3	212.8 (56.4)	167.4 (25.6)
Tone 4	292.1 (64.8)	182.5 (20.6)
Tone 5	101.8 (23.7)	105.6 (28.5)

Table 3: Mean Normalized Target Rime Durations (ms) and Standard Deviation

Tone 5 (checked) rimes were shorter than those of the other lexical tones; this is particularly stark in the first frame sentence, where checked rimes' duration was on average less than half that of other tones. The distribution of durational data across both frame sentences is represented graphically in (4).

(4)



Two-sample Student's t-tests between checked tone rime duration measurements and those of other tonal categories confirmed impressionistic judgments about durational differences, showing significant effects ($p < 0.01$) for each test between checked tone and the non-checked tones. Since duration values were normalized by frame sentence, cross-sentential comparisons were not conducted. Within each frame sentence, rimes on checked tone syllables were shown to be significantly shorter than those of each of the other four lexical tones.

3.2 Jitter Results

To determine the degree of glottalization on the vowel in target syllables, Praat's Jitter(ddp) function values were extracted. This function measures frequency perturbations on the vowel (vocal creak) as differences in the durations of consecutive glottal periods. To calculate the degree of perturbation, Jitter(ddp) averages the absolute durational differences between consecutive intervals, then divides by the average interval, yielding a numerical value between 1 and 6. A higher value corresponds to a higher degree of vocal creak.

Table 4 below presents the mean jitter results and standard deviations for each NJD lexical tone across both frame sentences.

Tone Category	Frame 1	Frame 2
Tone 1	3.198 (1.47)	2.181 (1.06)
Tone 2	3.346 (1.35)	2.947 (1.22)
Tone 3	3.217 (1.84)	2.175 (1.33)
Tone 4	1.932 (1.23)	2.551 (2.23)
Tone 5	1.988 (1.39)	2.246 (1.77)

Table 4: Mean Jitter(ddp) Values and Standard Deviation

Jitter values for checked tone syllables (as well as Tone 4) are noticeably lower than Tones 1, 2, and 3 in the first frame sentence. Student's t-tests performed individually between checked tone and lexical tones 1, 2, and 3 each reveal differences that are statistically significant ($p < 0.01$). The same results were obtained between Tone 4 and lexical tones 1, 2, and 3. In the second frame sentence, no significant effects were observed; degree of vocal creak was comparable on vowels of all five lexical tones.

This result indicates that vowels on checked tone syllables do not show evidence of glottalization; they exhibit less creak than uncontroversially open CV syllables. If residue of glottal stop codas was still present on these vowels, a higher degree of vocal creak compared to non-checked syllables would be expected. Rather, the inverse was observed in the current study. The jitter values corroborate an advanced stage of glottal coda weakening in checked tone syllables, and are compatible with a total neutralization analysis.

The generalization that emerges from combining duration and jitter results is problematic for earlier hypotheses regarding the trajectory of NJD checked tone merger. Instead of a clear correlation between durational and glottalization effects, there is a durational contrast but no evidence of a robust glottal coda (or even residue of a glottal coda). The next section will discuss these results in greater detail as they relate to previous hypotheses, as well as suggest possible explanations for the results.

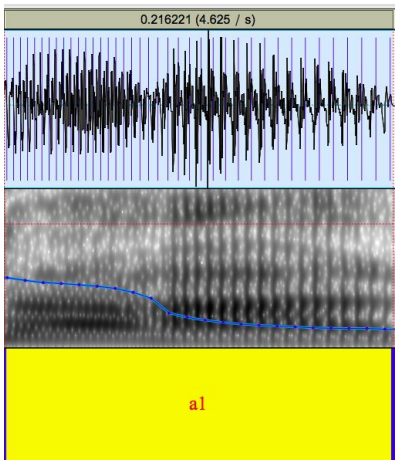
4. Discussion

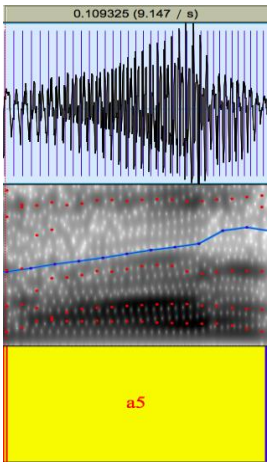
The acoustic profile of checked tone syllable rimes established in this study does not confirm the hypothesis that predicts a correlation between glottal weakening and rime lengthening as a component of NJD checked tone merger. No evidence of glottal coda residue was found on checked tone syllables, suggesting outright glottal deletion; however, a clear durational contrast between checked and non-checked syllables was observed.

Among the properties which distinguish NJD checked tone syllables from non-checked tone syllables, the durational contrast was robust in the speech of the experiment's participant. Checked rimes in the first frame sentence were on average less than 50% of the duration of those from other lexical tone categories. Though this ratio was less extreme in the second frame sentence, the differences evident in both frames were determined to be statistically significant.

Jitter measures provided no evidence of vocal creak on checked tone rimes relative to non-checked tones. In fact, jitter values on tones 1, 2, and 3 were found to be significantly higher (indicating a higher degree of vocal creak) than both tones 4 and 5 in the first frame sentence. One possible explanation for this result is a correlation between vocal creak and [L] tone; precisely what distinguishes these tones is the presence or absence of [L] tone in the melodic representation. Examination of the spectrograms of tones of each type corroborates this hypothesis, with typical examples given in (5).

(5) Tone 1 [HL] Tone 5 [H]





Striations in the spectrogram indicating vocal creak are apparent in the [L]-tone portion of tone 1 in (5); no such striations appear in the spectrograms of checked tone syllables. In terms of creak, then, checked tone syllables pattern with tone 4 [HH], an uncontroversially non-checked tone category.

A number of open questions remain regarding mid-merger NJD checked tone syllables and the nature of tone 5 as a lexical tone category in general. Chief among them is their melodic realization; earlier studies differ in their predictions about the pitch realization of NJD checked tone at different stages of merger. Song (2009) argues for initial retention of a unique melody after rime lengthening followed by merger into other NJD tonal categories. Gu (2015) predicts merger terminating in full adoption of Standard Mandarin melodies. The current study suggests a stage in merger—neutralization of syllabic shape contrast only—that does not fit either model. This obscures the possible predictions each hypothesis might make about the melodic realization of checked tone syllables, thus complicating any evaluation of the earlier predictions. I leave analysis of melodic realization, then, to future work.

5. Conclusion

This study has provided new evidence regarding checked tone merger in the Nanjing Dialect. An acoustic analysis performed on the recorded speech of a younger NJD native speaker found a robust durational difference between checked and non-checked tone syllables, but no evidence of glottal residue in checked tone syllables. These results conflict with previous accounts of the merger process, which predict a correlation between glottal coda weakening and rime lengthening in checked tone syllables. Thus, among the three contrast neutralizations inherent in checked tone merger, there is only evidence for the neutralization of contrast in syllable structure between checked and non-checked tone syllables.

REFERENCES

BOERSMA, PAUL and DAVID WEENINK. 2016. *Praat: doing phonetics by computer* (version 6.0.25). http://praat.org.

CHEN, MATTHEW C. 2000. *Tone sandhi: Patterns across Chinese dialects*. Cambridge, UK: Cambridge University Press.

FEI, JIA and SUN LI (eds.) 1993. *Nanjing fangyan zhi* [*Survey of the Nanjing dialect*]. Nanjing: Nanjing Publishing House.

FRANCIS, ALEXANDER; VALTER CIOCCA; and JOJO MAN CHING YU. 2002. Accuracy and variability of acoustic measures of voicing onset. *The Journal of the Acoustical Society of America* 113.2: 1025-1032.

GARELLEK, MARC. 2013. Production and perception of glottal stops. PhD Dissertation, University of California Los Angeles.

GU, QIAN. 2015. Nanjing fangyan dangjiang liangshe rushengyun de gongshi bianyi ji shili yanbian yanjiu [On synchronic variation and diachronic evolution of the entering tone in the dang and jiang classes in the Nanjing dialect]. *Linguistic Sciences* 14.4: 384-393.

KOIKE, YASUO. 1973. Application of some acoustic measures for the evaluation of laryngeal dysfunction. *Studia Phonologica* 7: 17-23.

KOIKE, YASUO and HIRANO MINORU. 1973. Glottal-area time function and subglottal pressure variation. *The Journal of the Acoustical Society of America* 54: 1618.

LIU, DANQING. 1995. *Nanjing fangyan cidian* [Dictionary of the Nanjing dialect]. Nanjing: Nanjing Publishing House.

SONG, YIDAN. 2009. Nanjing fangyanzhong de rusheng housaiwei shiyan yanjiu [Experimental research on entering tone in the Nanjing dialect]. *Journal of the School of Chinese Language and Culture Nanjing Normal University* 2009.2: 166-172.

TOIVONEN, IDA; LEV BLUMENFELD; ANDREA GORMLEY; LEAH HOITING; JOHN LOGAN; NALINI RAMLAKHAN; and ADAM STONE. Vowel height and duration. Proceedings of the 32nd West Coast Conference on Formal Linguistics, edited by Ulrike Steindl et al., 64-71. Somerville, MA: Cascadilla Proceedings Project.

TURK, ALICE; SATSUKI NAKAI; AND MARIKO SUGAHARA. 2006. Acoustic segment durations in prosodic research: a practical guide. *Methods in empirical prosody research*, edited by Stefan Sudhoff, Denisa Lenertova, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter and Johannes Schlieber. Berlin: Mouton de Gruyter.

SURANA, KUSHAN and JANET SLIFKA. 2006. Acoustic cues for the classification of regular and irregular phonation. Proceedings of INTERSPEECH 2006 and the 9th International Conference on Spoken Language Processing, 693-696. International Speech Communication Association.

VAN SANTEN, JAN P. 1992. Contextual effects on vowel duration. *Speech Communication* 11: 513-546.

Appendix A: Target List

Character	IPA	Tone	Melody	Gloss
八	[pa]	5	H	‘eight’
拔	[pa]	5	H	‘pluck’
罚	[fa]	5	H	‘punish’
法	[fa]	5	H	‘law’
发	[fa]	5	H	‘launch’
乏	[fa]	5	H	‘lack’
达	[ta]	5	H	‘arrive’
搭	[ta]	5	H	‘put up’
答	[ta]	5	H	‘answer’
塌	[t^ha]	5	H	‘collapse’
杀	[sa]	5	H	‘kill’
杂	[tsa]	5	H	‘mix’
撒	[sa]	5	H	‘sprinkle’
擦	[ts^ha]	5	H	‘scrub’
扎	[tsa]	5	H	‘tie up’
踏	[t^ha]	5	H	‘step on’
巴	[pa]	1	HL	*used in names*
他	[t^ha]	1	HL	‘he’
渣	[tsa]	1	HL	‘crumbs’
叉	[ts^ha]	1	HL	‘fork’
沙	[sa]	1	HL	‘sand’
纱	[sa]	1	HL	‘yarn’
莎	[sa]	1	HL	*used in names*
砂	[sa]	1	HL	‘grit’

爬	[p^{h}a]	2	LH	‘climb’
茶	[ts^{h}a]	2	LH	‘tea’
查	[ts^{h}a]	2	LH	‘check’
打	[ta]	3	LL	‘hit’
傻	[sa]	3	LL	‘stupid’
把	[pa]	3	LL	‘grasp’
塔	[t^{h}a]	3	LL	‘pagoda’
卡	[k^{h}a]	3	LL	‘get stuck’
爸	[pa]	4	HH	‘father’
霸	[pa]	4	HH	‘bully’
怕	[p^{h}a]	4	HH	‘fear’
大	[ta]	4	HH	‘big’
榨	[tsa]	4	HH	‘press’
差	[ts^{h}a]	4	HH	‘bad’
耙	[pa]	4	HH	‘shovel’
坝	[pa]	4	HH	‘dam’

Proceedings of the 29th North American Conference on Chinese Linguistics (NACCL-29). 2017. Volume 1.
Edited by Lan Zhang. University of Memphis, Memphis, TN. Pages 153-165.

Onset Voicing and Tone Distribution in the Taiwanese Lexicon

Sheng-Fu Wang
New York University

This study examines the relationship between onset voicing and distribution of tones in the Taiwanese (Taiwan Southern Min) lexicon by looking at whether onset-tone association still follows the prediction of the phonetics of tonogenesis: syllables with voiceless onsets are more likely to have a high tone, and syllables with voiced onsets are more likely to have a low tone. By using the observed-over-expected metric (e.g. Frisch & Zawaydeh, 2001) from counting lexical items in a dictionary, the analysis reveals that for the high-level, high-falling, and the mid-rising tones, the tone-onset association follows the prediction of the phonetics of tonogenesis, although the mid-falling tone does not exhibit the predicted association. The low-level tone, which used to be a high tone, is more positively associated with voiced onset, showing a consistent association with its synchronic tonal value. On the other hand, the checked tones, which also had a tonal flip since earlier stages of Min, exhibit a tone-onset association that conform to what their diachronic tonal values would predict.

1. Introduction

This paper aims to examine the relationship between onset voicing and tone distribution in the Taiwanese (Taiwan Southern Min) lexicon, as one step towards a better understanding of the synchronic phonotactics of this language. In particular, predictions made by diachronic and synchronic tonal values in connection to onset voicing are compared.

Taiwanese is a tone language featuring seven lexical tones. In the "checked syllables", which are syllables closed by a plosive coda such as /p/, /t/, /k/, or /ʔ/, there are the high-checked tone and the low-checked tone. In other syllable types, there are five tones, namely the high-level tone, the high-falling tone, the mid-falling tone, and the mid-rising tone, and the low-level tone. There are gaps for the combination of tones and syllable types: not all syllable types in Taiwanese carry all possible tones they can carry. For example, /ba/ only carries the mid-rising and low-level tones.

In the present study, I focus on the distribution of attested combinations between syllable types and tones, following both phonetic observations about F0 perturbations after voiceless and voiced onsets (higher F0 after voiceless stops, e.g., in Ohde 1984) and the related claims on tonogenesis (e.g., Hombert et al. 1979): It is believed that lexical tones arise through tonogenesis driven by F0 perturbations following different types of

onsets. Voiceless stops cause a higher F0 at the beginning of the vowel than voiced onsets do. Lexical tones, at least the difference in tonal registers, then arise through a reanalysis of the phonetic side effect as a phonological contrast.

Two claims about the tonal development in Min are of interest. First, voiced stops in early Min have become voiceless, potentially as a by-product of the reanalysis of a voicing contrast into a tonal contrast, and the present-day voiced stops arose through other sources such as nasal stops (Norman, 1988, 1973). Assuming no drastic change has happened to alter the tone distribution on syllable types, syllables with voiceless onsets in the synchronic Taiwanese lexicon should have a higher probability of having high tones than syllables with voiced onsets, which through the same process of tonogenesis should only acquire low tones. In other words, syllable types with voiceless onsets should be more likely to carry high tones than syllable types with voiced onsets. An investigation on how well onset voicing predicts the distribution of tones could shed light on the issue of whether the result of phonetically motivated diachronic changes maintains the relationship between synchronic phonological patterns and phonetic motivations.

Another interesting claim is that in two of the tonal classes[1], the high and the low tones underwent a change in tonal register: the tone that has been described as high tone historically has become a low-level tone in contemporary Taiwanese. The same is also true for the two checked tones, as there is a flip in tonal values from the historically high and low tones (Norman, 1973, 1974; Handel, 2003), the first two columns of Table 1.

Table 1. Diachronic and synchronic tones in Taiwanese

Diachronic class & tonal register	Tonal value in Taiwanese	Prediction on tone-onset association	
		Diachronic	Synchronic
*A, high	High-level	[−voice]	[−voice]
*A, low	Mid-rising	[+voice]	[+voice]
*B, high	High-falling	[−voice]	[−voice]
*B, low	(absent)	NA	NA
*C, high	Low-level	[−voice]	[+voice]
*C, low	Mid-falling	[+voice]	[+voice]
*D, high	Low-checked	[−voice]	[+voice]
*D, low	High-checked	[+voice]	[−voice]

[1] The tonal classes refer to the four categories that describe the shapes of tones in Middle Chinese. *A refers to level tones, *B refers to rising tones, *C refers to falling tones, and *D refers to checked tones.

Following these two claims, if the distribution of tones in Taiwanese still follows the diachronic development, there should be a positive association between voiceless onsets and the diachronic high tones: the high-level, high-falling, and, crucially, the low-level and low checked tones. As shown in Table 1, diachronic and synchronic tonal values would make different predictions on onset-tone associations. Whether the Taiwanese lexicon supports the prediction made by diachronic tonal values or synchronic ones is the main research question of this paper.

The wider implication of this paper concerns the role of synchronic and diachronic explanations in phonological grammars or lexicons. Two different views make two different predictions concerning the results of the present study. One of them sees the synchronic phonological grammar as results of language changes that are motivated by channel bias, which includes phonetic and perceptual factors (e.g., Blevins, 2004; Mielke, 2008). Under this view, the current relationship of the voicing feature and the tonal feature in Taiwanese should be independent of the tonal development other than the remnant of sound change. In other words, it is possible for low tones to associate more with voiceless stops. A different view is the relationship between onset voicing and tone may be somehow encoded in the Universal Grammar, which in turns constrains sound change. Under this view, high tones should be associated with voiceless onsets in the contemporary lexicon no matter regardless of the diachronic facts. If this is the case, then the onset-tone association may be seen as one instance of markedness asymmetry that not totally explainable by diachronic changes caused by performance factors (see de Lacy and Kingston, 2013), and it is thus motivated to posit markedness constraints to account for the asymmetry (e.g., *[+voice]$V^{\text{high-tone}}$ and *[−voice]$V^{\text{low-tone}}$).

2. Method

The data used in this study were extracted from a dictionary of Taiwanese downloaded from the Github repository[2] of the MoeDict project[3]. The MoeDict project aimed to build better interfaces to the government-provided language resources, namely the official online dictionaries of Mandarin Chinese, Taiwanese, and Hakka from the Ministry of Education in Taiwan. The downloaded dictionary contained a list of lexical entries written in Chinese characters along with their pronunciations coded in the Romanized writing system for Taiwanese. The pronunciations listed in the dictionary were taken as the attested syllable-tone combinations in the Taiwanese lexicon.

The dictionary was transformed by tallying the tonal categories each syllable type (e.g., pa, ba, nĩ, etc) can carry. Six syllable-level phonological features were annotated, including onset voicing, onset aspiration, onset nasality, vowel nasality, vowel length, and coda nasality. For the present analysis, only onset voicing was used.

[2] https://github.com/g0v/moedict-data-twblg/tree/master/raw
[3] https://www.moedict.tw

The aim of the analysis was to see whether the presence of each tone on a given syllable type is associated with onset voicing. This was carried out by investigating the observed/expected (O/E) value of the frequency of syllable types that contains a voiced onset and lacks a given tone. An O/E value higher or lower than 1 indicates that the number of syllable types having a particular tone is higher or lower than expected given independent distribution of onset voicing and tone gaps (see e.g. Coetzee and Pater, 2008; Frisch and Zawaydeh, 2001). The statistical significance of the deviation from prediction made by chance was evaluated with Chi-Square tests.

3. Results

For each tonal category, the association between whether the tone is attested and whether the onset is voiced presented with the O/E measurement and a Chi-Square test. Syllable types without onset were excluded from the analysis, as the prediction about their interaction between tonal distribution is less clear. In addition, since the phonetic motivation for the association between onset voicing and F0 arises through interactions between for stops and F0 perturbations, I also ran analysis with syllable types whose onset is a stop so that I could see whether this restricted data set behave differently.

This section is broken down into four subsections, according to whether the diachronic and the synchronic accounts have the same prediction and whether the tones are considered high tones or low tones in a synchronic sense. The results will be reported along with contingency tables showing the distribution of syllable types regarding whether they have voiced or voiceless onsets and whether they can carry a particular tone.

3.1 High-level and high-falling tones:

The high-level and high-falling tones were described as high tones in earlier Min. They also start with a high target in contemporary Taiwanese. This section examines whether they are more likely to co-occur with voiceless onset. Table 2 shows that for the overall dataset, syllables with voiceless onsets are more likely to carry high level tone (O/E = 1.19), whereas syllables with voiced onsets are less likely to carry high-level tone (O/E = 0.44). The Chi-Square test yielded statistical significance [$\chi^2(1, N = 468) = 121.87, p < .00001$]. The positive association between high-level tone and voiceless onsets follows the prediction of both the diachronic and the synchronic account.

Similar results were obtained when the analysis focused on the set of syllable types whose onsets are stops. As shown in Table 3, syllables with voiceless onsets are more likely to carry high level tone (O/E = 1.38), whereas syllables with voiced onsets are less likely to carry high level tone (O/E = 0.20). The association between onset voicing and the distribution of high-level tones may be considered stronger as these two O/E values become more different. The Chi-Square test yielded statistical significance [$\chi^2(1, N = 151) = 71.60, p < .00001$].

Table 2. Association between the high-level tone and onset voicing: overall data[4]

	With high-level tone	Without high-level tone
Voiced onset	38 (85.64) 8.11% / 18.29% O/E = **0.44**	82 (34.35) 17.52% / 7.34 % O/E = 2.38
Voiceless onset	296 (248.36) 63.25% / 53.07% O/E = **1.19**	52 (99.64) 11.11% / 21.29% O/E = 0.52

Table 3. Association between the high-level tone and onset voicing: syllables with plosive onsets

	With high-level tone	Without high-level tone
Voiced onset	6 (30.18) 3.97% / 19.98 % O/E = **0.20**	43 (18.82) 28.48% / 12.46% O/E = 2.28
Voiceless onset	87 (62.82) 57.61% / 41.60% O/E = **1.38**	15 (39.18) 9.93% / 25.95% O/E = 0.38

As for the high-falling tone, the analysis showed that it occurs slightly more than expected with voiceless onsets (O/E = 1.01) and less than expected with voiced onsets (O/E = 0.96). The Chi-Square test did not yield statistical significance [$\chi^2(1, N = 468) = 0.29$, $p = 0.59$]. The full results can be seen in Table 4.

Table 4. Association between the high-falling tone and onset voicing: overall data

	With high-level tone	Without high-level tone
Voiced onset	81 (83.84) 17.30% / 11.91% O/E = 0.96	39 (36.15) 8.33% / 7.72% O/E = 1.07
Voiceless onset	246 (243.15) 52.56% / 51.95% O/E = 1.01	102 (104.85) 21.79% / 22.40% O/E = 0.97

[4] For each cell in the table, the number in the first row without the parenthesis indicates the number of syllable types that (1) has a voiced onset or not and (2) carries a particular tone or not in the Taiwanese lexicon. The number in the parenthesis indicates the expected number given the overall distribution, if there is no dependency between onset voicing and whether a syllable can carry a particular tone. The second row shows the observed and the expected values in percentages. The third row shows the observed-over-expected value.

When the analysis was conducted with the syllable types whose onsets are stops, a stronger positive association between the occurrence of the high-falling tone and onset voicing was observed. The high-falling tone is more likely to occur with voiceless onsets (O/E = 1.09) and less likely with voiced onset (O/E = 0.81). The Chi-Square test yielded statistical significance [$\chi^2(1, N = 151) = 5.02$, $p < 0.05$]. The full results can be seen in Table 5. This positive association of high-falling tone and voiceless stops follows the prediction of both the diachronic and the synchronic account.

Table 5. Association between the high-falling tone and onset voicing: syllables with plosive onsets

	With high-falling tone	Without high-falling tone
Voiced onset	28 (34.40) 18.54% / 22.78% O/E = **0.81**	21 (14.60) 13.91% / 9.67% O/E = 1.43
Voiceless onset	78 (71.60) 51.66% / 47.42% O/E = **1.09**	24 (30.40) 15.89% / 20.13% O/E = 0.79

3.2. Mid-falling and mid-rising tones

This subsection presents analyses the contemporary mid-falling and mid-rising tones, who were also categorized as low-register tone diachronically. Table 6 shows that for the overall dataset, syllables with voiceless onsets are more likely to carry mid-falling tone (O/E = 1.22), whereas syllables with voiced onsets are less likely to carry high-level tone (O/E = 0.37). The Chi-Square test yielded statistical significance [$\chi^2(1, N = 468) = 111.22$, $p < .00001$]. The positive association between mid-falling tone and voiceless stops does not follow the prediction of both the diachronic and the synchronic account.

Table 6. Association between the high-falling tone and onset voicing: overall data

	With mid-falling tone	Without mid-falling tone
Voiced onset	28 (76.41) 5.98% / 16.32% O/E = **0.37**	92 (43.58) 19.66% / 9.31% O/E = 2.11
Voiceless onset	270 (221.58) 57.69% / 47.34% O/E = **1.22**	78 (126.41) 16.67% / 27.01% O/E = 0.62

When the analysis was conducted with the syllable types that have a stop as the onset, similar association between the occurrence of high-falling tone and onset voicing was observed. The mid-falling tone is more likely to occur with voiceless onset (O/E = 1.34) and less likely with voiced onset (O/E = 0.30). The Chi-Square test yielded statistical significance [$\chi^2(1, N = 151) = 54.61$, $p < .00001$]. The full results can be seen

in Table 7. Again, the positive association between mid-falling tone and voiceless stops does not follow the prediction of both the diachronic and the synchronic account.

Table 7. Association between the high-falling tone and onset voicing: syllables with plosive onsets

	With mid-falling tone	Without mid-falling tone
Voiced onset	9 (30.18) 5.96% / 19.99% O/E = 0.30	40 (18.82) 26.49% / 12.46% O/E = 2.13
Voiceless onset	84 (62.82) 55.62% / 41.60% O/E = 1.34	18 (39.18) 11.92% / 25.95% O/E = 0.46

As for the mid-rising tone, the analysis showed that the high-falling tone occurs more than expected with voiced onsets (O/E = 1.15) and less than expected with voiceless onsets (O/E = 0.95). The Chi-Square test yield statistical significance [χ^2(1, N = 468) 7.44, $p < .01$]. The full results can be seen in Table 8. The positive association between mid-falling tone and voiced stops follows the prediction of both the diachronic and the synchronic account.

Table 8. Association between the mid-rising tone and onset voicing: overall data

	With mid- rising tone	Without mid- rising tone
Voiced onset	95 (82.56) 20.29% / 17.64% O/E = **1.15**	25 (37.44) 5.34% / 8.00% O/E = 0.66
Voiceless onset	227 (239.43) 48.50% / 51.16% O/E = **0.95**	121 (108.56) 25.85% /23.19% O/E = 1.11

When the analysis was conducted with the syllable types that have a stop as the onset, similar association was observed. The mid-falling tone is more likely to occur with voiced onset (O/E = 1.11) and less likely with voiceless onset (O/E = 0.95), although the Chi-Square test did not yield statistical significance [χ^2(1, N = 151) = 2.20, $p = .16$]. The full results can be seen in Table 9.

Table 9. Association between the mid-rising tone and onset voicing: syllables with plosive onsets

	With mid- rising tone	Without mid-rising tone
Voiced onset	41 (36.99) 27.15% / 24.50% O/E = 1.11	8 (12.01) 5.29% / 7.95% O/E = 0.66
Voiceless onset	73 (77.01) 48.34% / 50.10% O/E = 0.95	29 (24.99) 19.20% /16.55% O/E = 1.16

3.3. High-checked tone

For the high-checked tone, a synchronic phonetic account would predict that it should be more likely to occur after a voiceless onset. However, this tone is described as having a low tone in early stages, so a diachronic account should predict an association with voiced onsets.

Analysis on the overall data seemed to suggest that the association between the high-checked tone and onset voicing goes in the direction of the diachronic account: it occurs with voiced onsets more than expected (O/E = 1.24) and less than expected with voiceless onsets (O/E = 0.92). The Chi-Square test yielded statistical significance [$\chi^2(1, N = 278) = 9.06$, $p < .01$]. The full results can be seen in Table 10.

Table 10. Association between the high-checked tone and onset voicing: overall data

	With high-checked tone	Without high-checked tone
Voiced onset	55 (44.34) 19.78% / 15.95% O/E = **1.24**	12 (22.65) 4.32% / 8.15% O/E = 0.53
Voiceless onset	129 (139.65) 46.40% / 50.23% O/E = **0.92**	82 (71.35) 29.50% /25.66% O/E = 1.15

Similar results were obtained with analysis on a smaller dataset consisting of only syllables with stop onsets. The high-checked tone occurs with voiced onsets more than expected (O/E = 1.22) and less than expected with voiceless onsets (O/E = 0.90). The Chi-Square test yielded statistical significance [$\chi^2(1, N = 86) = 4.42$, $p < .05$]. The full results can be seen in Table 11.

Table 11. Association between the high-checked tone and onset voicing: syllables with plosive onsets (wrong numbers)

	With high-checked tone	Without high-checked tone
Voiced onset	24 (19.65) 27.9% / 22.85% O/E = **1.22**	2 (6.35) 2.32% / 7.38% O/E = 0.31
Voiceless onset	41 (45.34) 47.68% / 52.73% O/E = **0.92**	19 (14.65) 29.50% /25.66% O/E = 1.15

3.4. Low-level and low-checked tones

Contemporary low-level and low-checked tones were described as high tones in early Min where their tonal value supposedly reflect the phonetic environment on tonogenesis. Following the diachronic prediction, they should be more likely to occur with voiceless stops. On the other hand, a synchronic phonetic account would predict a positive association between these tones and voiced onsets.

For the low-level tone, the synchronic phonetic account made the right prediction on the actual data: low-level tones occur more than expected with voiced onsets (O/E = 1.19) and less than expected with voiceless onsets (O/E = 0.93). The Chi-Square test yielded statistical significance [χ^2(1, N = 468) = 4.42, = 10.09, $p < .01$]. The full results can be seen in Table 12.

Table 12. Association between the low-level tone and onset voicing: overall data

	With high-checked tone	Without high-checked tone
Voiced onset	93 (82.56) 19.87% / 16.71% O/E = **1.19**	27 (37.44) 5.77% / 8.93% O/E = 0.65
Voiceless onset	212 (239.43) 45.29% / 48.46% O/E = **0.93**	136 (108.56) 29.06% /25.90% O/E = 1.12

However, when the analysis was done with the set of syllable types whose onsets are stops, the there was a change of the O/E values: the low-level tone is slightly more likely to occur with voiceless stop onsets (1.01) and less likely with voiced onsets (0.98), although the result of the statistical test was far from reaching significance [χ^2(1, N = 151) = 0.00, $p = .95$]. The results are shown in Table 13.

Table 13. Association between low-level tone and onset voicing: syllables with plosive onsets

	With high-checked tone	Without high-checked tone
Voiced onset	37 (37.64) 24.50% / 24.93% O/E = 0.98	12 (11.35) 7.94% / 7.52% O/E = 1.06
Voiceless onset	79 (78.35) 52.31% / 51.89% O/E = 1.01	23 (23.64) 15.23% / 15.66% O/E = 0.97

The distribution of low-checked tones, just like that of the high-checked tones, follows the prediction of a diachronic account: it occurs more than expected with voiceless onsets (O/E = 1.16) and less than expected with voiced onsets (O/E = 0.49). The Chi-Square test yielded statistical significance [$\chi^2(1, N = 278) = 15.64$, $p < .00001$]. The full results can be seen in Table 14. Similar patterns can be observed with the subset of data where the onsets of syllables are stops, but the Chi-Square test did not yield statistical significance. [$\chi^2(1, N = 86) = 2.59$, $p = .11$]. The results can be seen in Table 15.

Table 14. Association between the low-checked tone and onset voicing: overall data

	With low-checked tone	Without low-checked tone
Voiced onset	14 (28.43) 5.03% / 10.22% O/E = **0.49**	53 (38.56) 19.06% / 13.87% O/E = 0.65
Voiceless onset	104 (89.56) 37.41% / 32.22% O/E = **1.16**	107 (121.44) 38.49% / 43.68% O/E = 0.88

Table 15. Association between low-checked tone and onset voicing: syllables with plosive onsets

	With low-checked tone	Without low-checked tone
Voiced onset	7 (10.88) 8.13% / 12.65% O/E = 0.64	19 (15.12) 22.09% / 17.58% O/E = 1.26
Voiceless onset	29 (25.12) 33.72% / 29.20% O/E = 1.15	31 (34.88) 36.05% / 40.56% O/E = 0.89

4. Discussion and Conclusion

Table 16 summarizes the results and the corresponding predictions that are born out. The high-level, mid-rising, and high-falling tones have not undergone flip of tonal values, and the association with onset voicing follows the prediction based on tonal values. In the other case where diachronic and synchronic predictions converge, the prediction was not borne out: the mid-falling tone is more positively associated with voiceless onsets. One potential account based on synchronic grammar is tone sandhi: when a syllable with mid-falling tone occurs in a constituent-medial position, it undergoes tone sandhi and surface with a high-falling tone. However, the same account cannot easily explain the association of voiceless onsets with the high-level tone, since high-level tone also changes its tone register and became a low-level tone in the sandhi form. An investigation of the frequency of base and sandhi forms of these syllables may be able to shed light on this issue.

Table 16. Overall results on onset-tone association in Taiwanese

Diachronic class & tonal register	Tonal value in Taiwanese	Result (positive association with…)	consistent with which prediction?
*A, high	High-level	voiceless onsets (and stops)	diachronic and synchronic
*A, low	Mid-rising	voiced onsets	diachronic and synchronic
*B, high	High-falling	voiceless stops	diachronic and synchronic
*B, low	(absent)	NA	NA
*C, high	Low-level	voiced onsets	synchronic
*C, low	Mid-falling	voiceless onsets (and stops)	neither
*D, high	Low-checked	voiceless onsets	diachronic
*D, low	High-checked	voiced onsets and stops	diachronic

In the cases where the diachronic and synchronic accounts diverge, the results were mixed. In the case of checked tones, the diachronic prediction proves to be correct: the association between tonal distribution and onset voicing is the reverse of what a synchronic account would explain, and corresponds to the diachronic description on a tonal flip. One potential confound is that the actual pronunciation of checked tones varies widely from dialect to dialect (e.g., Chen, 2009). Given the description based on some other dialect/accent of Taiwanese, the synchronic account may be consistent with the diachronic pattern. The same level of inconsistency is not found for the tones on plain syllables. However, the variations on synchronic account does not discredit the fact the

onset voicing is connected to the diachronic tone values for checked tones. Such tighter connection to diachronic sound change instead of a synchronic phonetic motivation would be related to the wider discussion of how unnatural sound patterns arise in languages, as in the work of Mielke (2008) and Blevins (2004), who argued that phonetic motivations for sound patterns only should be regard diachronic sound change.

However, for the low-level tone, the synchronic account makes the correct prediction that its occurrence is more associated with voiced onsets. This finding either shows that, to fully account for synchronic patterns, the diachronic phonetic account must be more precise about the pathway of sound change. Another interpretation to this finding is that synchronic phonology directly incorporates or reflect phonetic or perceptual motivation, as suggested in frameworks such as the Dispersion Theory (e.g., Flemming, 2004): One hypothesis might be a need to contrast these two tones: as these two tones occupy similar tonal range, mid-falling tone behaves like a high tone in this pair, at least in its initial tonal target, so that it contrasts better with the low-level tone. This hypothesis would also predict that the mid-falling tone's initial target should be reliably higher than that of the low-level tone, which is supported by acoustic measurements shown in Pan (2017).

As for the question on diachronic and synchronic explanations in phonological grammar, the mixed results make it difficult to assess whether the tone-onset association in Taiwanese is better explained by performance-based diachronic change or a synchronic grammar that affects the makeup of the lexicon. One direction to explore this further is to examine the distribution of onset and tone categories in different lexical strata within the Taiwanese lexicon (e.g., Tu, 2013, 2011): different lexical strata reflect pronunciations of certain syllable onsets at different stages of the Min language, which may shed light on the path of historical change on the onset-tone association.

To conclude, the present study has shown that there exists a certain connection between onset voicing and distribution of tones in the synchronic Taiwanese lexicon, although tonal categories behave differently as to whether such connection is better predicted by diachronic or synchronic tonal values. The quantitative measures of such connections make contribution to the understanding of Taiwanese lexicon, and the results have wider implications on the locus of phonetic motivation of sound patterns in languages and on the historical phonology of Chinese languages. Future directions of this project include looking at the relationship between onset and tone in other languages in the Min family, as well as testing whether speakers are aware of the lexical statistics about the distribution of tone and onset voicing across syllable types.

REFERENCES

BLEVINS, JULIET. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

CHEN, SHU-CHUAN. 2009. 台灣閩南語元音系統及陰，陽入聲調的變異與變化—台灣閩南語的字表調查分析 [The variation and change of the vowel system and the checked tones in Taiwan Southern Min – A wordlist-based investigation]. *Journal of Taiwanese Languages and Literature*, 3, 157-178.

COETZEE, ANDRIES W. and JOE PATER. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language & Linguistic Theory*, *26*(2), 289-337.

DE LACY, PAUL and JOHN KINGSTON. 2008. Synchronic explanation. *Natural Language & Linguistic Theory*, 31(2), 287-355.

FLEMMING, EDWARD. 2004. Contrast and perceptual distinctiveness. In *Phonetically-based phonology*. 232-276. Cambridge: Cambridge University Press.

FRISCH, STEFAN A. and BUSHRA ADNAN ZAWAYDEH. 2001. The psychological reality of OCP-place in Arabic. *Language*, *77*(1), 91-106.

HANDEL, ZEV. 2003. Northern Min tone values and the reconstruction of "softened initials". *Language and Linguistics*, *4*(1), 47-84.

OHALA, JOHN J. and WILLIAM G. EWAN. 1979. Phonetic explanations for the development of tones. *Language*, *55*(1), 37-58.

PAN, HO-HSIEN. 2017. Glottalization of Taiwan Min checked tones. *Journal of the International Phonetic Association*, *47*(1), 37-63.

MIELKE, JEFF. 2008. *The emergence of distinctive features*. New York: Oxford University Press.

NORMAN, JERRY. 1973. Tonal Development in Min. *Journal of Chinese Linguistics*, *1*(2), 222-238.

NORMAN, JERRY. 1974. The initials of Proto-Min. *Journal of Chinese Linguistics*, *2*(1), 27–36.

NORMAN, JERRY. 1979. *Chinese*. Cambridge: Cambridge University Press.

OHDE, RALPH N. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, *75*(1), 224-230.

TU, CHIA-LUN. 2011. *閩語歷史層次分析與相關音變探討 [The Diachronic Strata Analysis of Min Phonology]*. Taipei: National Taiwan University dissertation.

TU, CHIA-LUN. 2013. 閩語古全濁聲類的層次分析 [The Diachronic Strata Analysis of the Traditional Voiced Initials in Min]. *Language and Linguistics*, *14*(2), 409-456.