Kansas Working Papers in Linguistics, Vol. 28 (2006): 65-126

Testing the Role of Phonetic Naturalness in Mandarin Tone Sandhi

Jie Zhang
The University of Kansas
zhang@ku.edu

Yuwen Lai
The University of Kansas
yuwen@ku.edu

Abstract

It has long been noted that phonological patterning is influenced by phonetic factors. But phonologists diverge on whether phonetic motivations take effect in synchronic or diachronic phonology. This article aims to tease apart the two theories by investigating native Mandarin speakers' applications of two tone sandhi processes to novel words: the phonetically motivated contour reduction 213→21/__T (T≠213) and the neutralizing 213→35/__213 whose phonetic motivations are less clear. Twenty Mandarin subjects were asked to produce two monosyllables they heard as disyllabic words. Five groups of disyllabic words were tested: AO-AO (AO=actual occurring morpheme) where the disyllable is also a real word, AO-AO' where the disyllable is non-occurring, AO-AG (AG=accidental gap in Mandarin lexicon — legal syllable and tone but non-existent combination), AG-AO, and AG-AG. The first syllable is always 213, and the second syllable has one of the four tones in Mandarin. Results show that speakers apply the phonetically more natural 213→21 sandhi more quickly and with greater accuracy than the 213→35 sandhi. Theoretically, the study supports the direct relevance of phonetics to synchronic phonology by showing that there is a psychological advantage to phonetically natural patterns. Methodologically, it complements existing research paradigms that test the nature of the phonology-phonetics relationship, e.g., the study of phonological acquisition and the artificial language paradigm; when extended to other Chinese dialects, it can also provide insights into the long-standing mystery of how Chinese speakers internalise complicated tone sandhi patterns that sometimes involve opacity, near-neutralization, and syntactic dependency.

1. Introduction

1.1. The relevance of phonetics to phonological patterning

It has long been noted that phonological patterning is influenced by phonetic factors. The influence manifests itself in a number of ways, the staple of which is the comparison between the prevalence of patterns that have articulatory or perceptual bases and the scarcity of those that do not. For example, velar palatalization before high front vowels, postnasal voicing, and regressive consonant place assimilation have clear phonetic motivations and are extremely well-attested, while velar palatalization before low back vowels, postnasal devoicing, and progressive consonant place assimilation are nearly nonexistent. The influence can also be manifested in terms of implicational statements. For example, typological works on consonant place assimilation (Mohanan 1993, Jun 1995, 2004) discovered that if oral stops are targets of assimilation in a language, then *ceteris paribus*, nasal stops are also targets of assimilation; this is to be expected perceptually, as nasal stops have weaker transitional place cues and are thus more prone to losing their contrastive place than oral stops when articulatory economy is of concern (cf. the Production Hypothesis in Jun 1995, 2004).

Evidence for the relevance of phonetics can also be found in the peripheral phonology of a language even when the phonetic effects are not directly evident in its core phonology. Such peripheral phonology may include the phonology of its established loan words (Fleischhacker 2001, Kang 2004, Kenstowicz, to appear), how speakers adapt foreign words on-line (Davidson, to appear), and speakers' judgments on poetic rhyming (Steriade and Zhang 2001). For example, Steriade and Zhang (2001) show that although postnasal voicing is not neutralizing in Romanian, its phonetic effect is crucial in accounting for speakers' preference for /Vnt/~/Vnd/ as a semi-rhyme over /Vt/~/Vd/.

Finally, the parallels between the traditionally conceived categorical/phonological and gradient/phonetic patterns also indicate their close relation. Flemming (2001), for instance, outlines the similarity of patterning between phonological assimilation and phonetic coarticulation as well as a number of other processes present in both the traditional phonological and phonetic domains.

1.2. The synchrony vs. diachrony controversy

Although the observations illustrating *some* form of relationship between phonology and phonetics are relatively uncontroversial, the precise way in which this relationship should be captured is a continuous point of contention among phonologists. Many theories have been proposed within rule-based phonology to encode this relation, from the abbreviation conventions of *SPE* (Chomsky and Halle 1968), the innateness of articulatory-based phonological processes in Natural Phonology (Stampe 1979), to the grounding conditions for universal constraints in Grounded Phonology (Archangeli and Pulleyblank 1994). Optimality Theory (Prince and Smolensky 1993), with its separation of "problem" (markedness constraints) and "solution" (selection of optimal candidate according to the interaction of markedness and faithfulness constraints), and consequently the ability to state phonetic motivations explicitly in the system as markedness constraints, further invites phonetic explanations into phonology (Hayes and Steriade 2004).

What Optimality Theory further brings to the limelight is an old dispute whether the influence of phonetics on phonology is synchronic or diachronic, and both positions have been abundantly explicated in the literature.

The synchronic approach treats phonetic knowledge as part of linguistic knowledge proper and admits phonetic details into the phonological grammar (e.g., Boersma 1998, Steriade 1999, 2001, Kirchner 2000, 2001, 2004, Flemming 2001, Zhang

2002, 2004). The influence of phonetics on phonology is "direct," in that the phonology of the speakers is directly calculated upon their phonetic abilities together with other phonological primitives and principles, and that the phonetic naturalness in phonological patterns is a direct result of the phonetic nature of the phonological grammar.

The diachronic approach, however, considers the phonetically natural phonologies to be the results of historical sound changes, which are often initiated and perpetuated by perceptual "mistakes" made by speakers over generations and are hence phonetically natural. From the synchronic perspective, the speakers are simply learning the patterns handed to them by history; their phonology is composed of abstract representations imposed by the Universal Grammar (Chomsky 1981) with no phonetic teleology, and their phonetic ability is not directly responsible for any phonetic naturalness in the synchronic system (e.g., Anderson 1981, Ohala 1981, 1992, 1993, 1996, Blevins and Garrett 1998, Hale and Reiss 2000, Buckley 2000, 2003, Hansson 2001, Hyman 2001, Blevins 2004, Yu 2004).

To tease apart these two approaches has proved difficult. Proponents of the synchronic approach have been mostly concerned with establishing stringent implicational statements on phonological behavior, discovering the phonetic rationales behind the implications, and proposing Optimality-Theoretic models from which the implicational statements fall out as predictions (e.g., Jun 1995, Steriade 1999, Kirchner 2001, Zhang 2002). These endeavors must be considered theory-making, as the phonetic rationales provide the bases for predictions, and they have contributed to our understanding of phonological patterns. But the implicational statements gathered from data are necessarily inductive — provided that we have not looked at all languages, we cannot be certain that such statements are empirically valid. This leaves open the question "what are the ramifications of counterexamples to the theory?" And most works in the synchronic approach have not been explicit about the answer to this question. A more serious problem with using implicational statements to motivate the synchronic relevance of phonetics is that the statements may in fact be consistent with both the synchronic and diachronic approaches. The fact that in the literature, certain phonetically motivated phonological asymmetries are accounted for synchronically (e.g., Jun 1995 for consonant place assimilation, Zhang 2002 for contour tone distribution), while others are explained diachronically (e.g., Hansson 2001 for consonant harmony, Blevins and Garrett 1998 for CV metathesis), points to this distinct possibility. In other words, explicating a workable synchronic analysis only shows its sufficiency, not its necessity.

Heralds of the diachronic approach, on the other hand, have focused on identifying the relationship between observed phonological patterns and common phonetically motivated sound change (e.g., Blevins and Garrett 1998, Blevins 2004), and Occam's razor is considered a safeguard against the same "necessity" problem outlined above for the synchronic approach (e.g., Hale and Reiss 2000). The existence of unnatural patterns and counterexamples to established implicational statements (e.g., Buckley 2000, Hyman 2001, Yu 2004) and the fact that they are readily learnable by speakers (e.g., Buckley 2003, Seidl and Buckley 2005) are also taken as arguments against the synchronic approach, with the assumption that they cannot be accommodated in this approach. But this logic is faulty in a number of respects. First, no proponents of the synchronic approach would in fact claim that phonology is purely determined by phonetics; phonetics-independent factors may interact with phonetics to produce surface counterexamples to implicational statements or unnatural patterns, and these need to be investigated carefully on a case by case basis. Relatedly, the point of interest in the learnability issue should be whether there is a *difference* between phonetically natural and unnatural patterns, not whether unnatural patterns are learnable. Showing the latter does not prove that phonetics is irrelevant, but showing the former does indicate that there must a cognitive bias towards the natural patterns, and hence dissolves the Occam's razor argument — the cognitive bias can be predicted by the synchronic approach, but not by the diachronic one; thus the two approaches differ beyond "simplicity." Finally, the diachronic approach needs an explicit model for how articulatory and perceptual mistakes culminate in the regularity of sound change, and how the learning mechanism of the speaker interprets the perpetuation of such "mistakes." But this is hardly delivered.

1.3. Experimental approaches to tease apart synchrony and diachrony

As just mentioned, one possible way to tease apart the synchronic and diachronic approaches is to observe in a language with both phonetically natural and unnatural patterns, whether the natural patterns are acquired more quickly and with greater accuracy than the unnatural patterns in language acquisition. A positive answer to this query would support the synchronic approach, as it would indicate that learners must have taken advantage of their phonetic abilities in phonological learning, and the only way in which this is possible is that phonetics is directly relevant to the construction of their phonological grammar. A negative answer would support the diachronic approach, as only this approach predicts that learners should not be sensitive to whether or not the process is phonetically natural — they are simply learning the patterns handed to them by language history. But this kind of research is likely marred by the lack of control between the two types of processes. One can only make safe conclusions about the learning difference between natural and unnatural processes if all confounding factors, such as the regularity of the pattern, lexical frequency, and transitional probability, are controlled for. But in observing language learning in a natural setting, this kind of control is extremely difficult to come by.

The claim that phonetically natural morphophonological processes are acquired earlier and with fewer errors has been made in the literature (e.g., MacWhinney 1978, Slobin 1985, Menn and Stoel-Gammon 1995). E.g., Slobin (1985) compares the effortless acquisition of final devoicing by Turkish children and the error-ridden acquisition of stop-spirant alternation in Modern Hebrew by Israeli children and suggests that there is a hierarchy of acceptable alternation based on universal predispositions that favor assimilation and simplification in the articulatory output (p.1209).

Buckley (2002) correctly points out that the role of such universal predispositions can only be established if the accessibility — the particular distribution and regularity — of the pattern is teased apart from the naturalness of the pattern, and in demonstrating that many unnatural patterns are acquired readily due to their high regularity and high frequency of occurrence while many natural patterns are acquired with much difficulty due to their low accessibility, Buckley argues that accessibility, but not naturalness, determines the ease of learning. But what Buckley fails to show is that when the accessibility is matched, an unnatural process is acquired just as easily as a natural process. The only such comparisons that can be found in Buckley (2002) are in Hungarian — the more natural backness harmony vs. the less natural /a/-lengthening, both of which are highly accessible, and the more natural rounding harmony vs. the less natural /e/-lengthening, both of which have low accessibility. MacWhinney (1978)'s original work, which Buckley cites, shows that backness harmony is acquired earlier than /a/-lengthening, but rounding harmony is acquired later than /e/-lengthening. Therefore, naturalness does seem to affect the order of acquisition, but it is unclear what the precise effect is. Moreover, given that these comparisons are only made under a very crude control of "accessibility," they cannot be deemed conclusive in any way.

Another experimental approach to potentially distinguish the synchronic and diachronic relevance of phonetics is to test the learning of patterns with different degrees of naturalness in an artificial language. The artificial grammar paradigm (see Reber 1967, 1989, Redington and Chater 1996) has been widely used to investigate the learnability of phonological patterns in both adults and children. The paradigm usually involves two stages — the exposure stage, in which the subject is presented with stimuli generated by an artificial grammar, and the testing stage, in which the subject is tested on their learning of the patterns in the artificial grammar, measured by their ability to distinguish legal vs. illegal test stimuli, reaction time, or looking time in the head-turn paradigm for infant studies. For example, Onishi et al. (2002) show that adult English speakers can learn non-English phonotactic regularities, such as "/p/ only occurs in onset position before /ɪ/," upon only a few minutes of familiarization of CVC syllables with the patterns, as indicated by faster reaction times when asked to repeat new syllables that conform to the patterns. The same research group later demonstrates that the same regularities can be acquired by 16.5-month-old infants, who listened longer to new syllables that violated the patterns in a head-turn test (Chambers et al. 2003). Works by

Newport, Aslin, and Saffran (e.g., Saffran et al. 1996, Aslin et al. 1998, Newport and Aslin 2004) show that speakers are capable of learning regularities between adjacent syllables, but not between non-adjacent syllables, in an artificial grammar. Dell et al. (2000) show that artificially induced speech errors conform to not only language-wide (e.g., /h/ only occurs in onset position and /ŋ/ only occurs in coda position in English), but also experiment-wide (e.g., /f/ and /s/ only occur in onset and coda positions, respectively), phonotactic constraints.

Although the above works all demonstrate that phonological patterns not grounded in phonetics can be acquired by speakers, as we have argued, they do not suffice to disprove the relevance of phonetics to synchronic phonology. To do so, we need to show that there is no difference between the learning of natural and unnatural patterns when all else is equal. The artificial grammar paradigm is particularly suited for this comparison as the natural and unnatural patterns can be designed to have matched regularity, lexical frequency, and transitional probability. This line of research has been actively pursued, and we have conflicting results from different researchers. Seidl and Buckley (2005) report two experiments that test whether nine-month-old infants' learn natural and unnatural patterns differently. The first experiment tests whether the infants prefer a phonetically grounded pattern in which only fricatives and affricates, but not stops, occur intervocalically, or an arbitrary pattern in which only fricatives and affricates, but not stops, occur word-initially. The second experiment tests the difference between a grounded pattern in which a labial consonant is followed by a rounded vowel and a coronal consonant is followed by a front vowel and an arbitrary pattern in which a labial consonant is followed by a high vowel and a coronal consonant is followed by a mid vowel. In both experiments the infants learn both patterns fairly well and show no learning bias towards the phonetically grounded pattern, indicating that phonetic naturalness does not play a role in the learning of synchronic phonological patterns. But in experiment 1, all fricatives and affricates used are stridents, and as Kirchner (2001, 2004) shows, the precise articulatory control necessary for stridents in fact makes them less desirable in intervocalic position. Therefore, at least for this experiment, the validity of the claim that there is no learning bias towards phonetically grounded patterns is open to debate. In Jusczyk et al. (2003), 4.5-month-old infants are presented with sets of three words, or "triads," which consist of two monosyllabic pseudowords in the forms VC_1 and C_2V, followed by a disyllabic word in which either C_1 or C_2 place-assimilates to the adjacent consonant (*an, bi, ambi; an, bi, andi*). The C_1-assimilation pattern is perceptually motivated and crosslinguistically extremely common, while the C_2-assimilation pattern has no clear perceptual grounding and crosslinguistically extremely rare. In a head-turn procedure, infants show no difference in looking time towards the triads with regressive and progressive assimilations, indicating irrelevance of phonetic naturalness in the learning of these patterns. But as Seidl and Buckley (2005) point out, the A, B, AB triad procedure is quite novel in infant research, and the assumption that the infants take the AB string as a concatenation of A and B may not be valid. Pycha et al. (2003) test adult English speakers' learning of three non-English patterns — palatal vowel harmony (stem and suffix vowels agree in [back]), palatal vowel disharmony (stem and suffix vowels disagree in [back]), and palatal arbitrary (arbitrary relation between stem and suffix vowels) — and find that although subjects exhibit better learning of the harmony and disharmony patterns than the arbitrary pattern, there is no difference between harmony and disharmony. Taking harmony as phonetically natural and disharmony as unnatural, they conclude that naturalness is not relevant to the construction of the synchronic grammar. Wilson (2003), in two similar experiments with similar results, interprets the results differently, however. Upon showing that alternations involving both nasal assimilation (l → n / [+nasal]V__) and nasal dissimilation (n → l / [+nasal]V__) are learned with greater accuracy than random alternations (e.g., l → n / [+dorsal]V__), Wilson argues that both assimilation and dissimilation have a privileged cognitive status in phonological grammar, presumably due to their phonetic naturalness. Wilson (2006) shows that when speakers are presented with highly impoverished evidence of a new phonological pattern, they are able to extend the pattern to novel contexts predicted by a phonetically based phonology and linguistic typology, but not to other contexts; e.g., speakers presented with velar palatalization before mid vowels are able to extend the process before high vowels, but not *vice versa*. Wilson shows that a phonology that encodes no substantive bias cannot predict these experimental observations and further develops a substantively biased framework of phonology that captures these behaviors.

The problem with using the artificial language paradigm in adult research is that the learning at best approximates second language acquisition, whose mechanism is arguably very different from first language acquisition (Cook 1969, 1994, Dulay et al. 1982, Bley-Vroman 1988, Ellis 1994, among others); but it is the latter that we are more interested in, as the synchrony vs. diachrony issue more crucially hinges on the relevance of phonetic naturalness during the construction of *native* phonological grammars.[1] The two infant studies that directly compare natural and unnatural patterns with controlled lexical frequency and regularity — Jusczyk et al. (2003) and Seidl and Buckley (2005) — both have methodological uncertainties, as mentioned above. Therefore the extent to which phonetic naturalness is relevant to first language phonological acquisition still awaits further research.

1.4. The present study

The present study aims to overcome some of the difficulties in testing the phonology-phonetics relationship by using the nonce-probe paradigm ("wug" test) (Berko 1958) with adult speakers. In a typical wug test, subjects are taught novel forms in their language and then asked to provide morphologically complex forms, using the novel forms as the base. For instance, the subject is first shown a picture of an imaginary animal and instructed "This is a wug [wʌg];" then a picture with two of the same animal is shown and the subject is prompted "Now there is another one. There are two ___." Upon successful learning of English plural morphophonology, the subject should respond with "[wʌgz]." This paradigm has been widely used to test the productivity of regular and irregular morphological rules (e.g., Bybee and Pardo 1981, Albright 2002, Albright and Hayes 2003, Pierrehumbert, to appear) and morphophonological alternations (e.g., Hsieh 1970, 1975, 1976, Wang 1993, Zuraw 2000, Albright et al. 2001). Our study wug-

[1] Moreover, the artificial language paradigm often involves a heavy dose of explicit learning, while second language acquisition, like first language acquisition, often involves a significant amount of implicit learning. This furthers the distance between artificial language learning and real language acquisition.

tests two patterns of tonal alternation (tone sandhi) that differ in the degree of phonetic naturalness in Mandarin Chinese and compares the reaction time and accuracy with which the sandhi patterns apply to nonce words.

This approach is advantageous to previous attempts to test the role of phonetics in synchronic phonology in the following respects. First, it not only tests the productivity of phonological patterns, natural or unnatural, but also directly compares speakers' behaviors on patterns that differ in naturalness. As we have argued, this comparison provides a sufficient basis for teasing apart the synchronic and diachronic approaches to phonetics-in-phonology. Second, this paradigm uses real phonological patterns that exist in the subjects' native language. This circumvents the learning-strategy problem with the artificial language paradigm. Third, given that the wug forms are created by the researchers, it is easier for us to control for confounding factors such as lexical frequency and transitional probability. This, to a great extent, minimises the control problem in studying phonological learning in a naturalistic setting.

1.5. Organization of the paper

We discuss the details of the tone sandhi patterns under investigation in Mandarin and methodological issues in §2. Results of the experiments are given in §3. Theoretical implications of the results are further discussed in §4. §5 is the conclusion.

2. Methodology

2.1. Tone sandhi in Mandarin Chinese and general hypotheses

Mandarin Chinese is a prototypical tone language, in which the relative pitch on a syllable serves contrastive functions. There are four lexical tones in Mandarin: 55, 35, 213, and 51,[2] as shown in the examples in (1). The representative pitch tracks of the four tones are given in Figure 1. Although Tone 2 is usually transcribed as a high rising tone 35, there is a small pitch dip at the beginning of the tone, creating a turning point. In fact, various researchers have shown that the perceptual difference between Tone 2 and Tone 3 mainly lies in the timing and pitch height of the turning point (Shen and Lin 1991, Shen et al. 1993, Moore and Jongman 1997). We can also notice in Figure 1 that the different tones have different durational properties; in particular, Tone 3 has the longest duration. These observations will become important in the discussion of Mandarin tone sandhi and the experimental results.

(1) Mandarin tones:

Tone 1:	ʂow55	'to collect'
Tone 2:	ʂow35	'ripe'
Tone 3:	ʂow213	'hand'
Tone 4:	ʂow51	'thin'

---

2 Tones are marked with Chao letters (Chao 1948, 1968) here. "5" indicates the highest pitch used in lexical tones while "1" indicates the lowest pitch. Contour tones are marked with two juxtaposed numbers. E.g., 51 indicates a falling tone from the highest pitch to the lowest pitch.

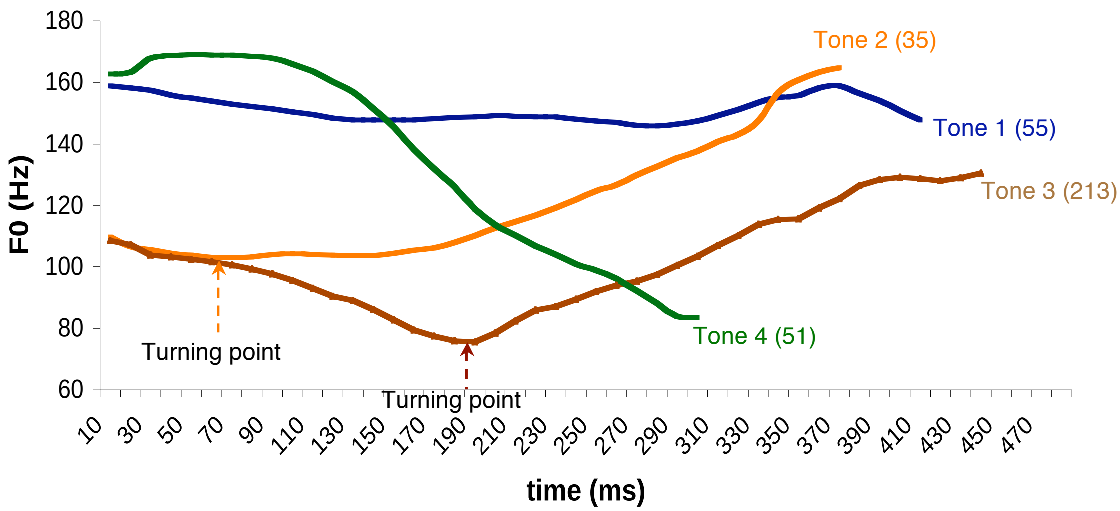


Figure 1. Representative pitch tracks of the four tones in Mandarin.

The tone on a syllable may undergo alternation as conditioned by adjacent tones or the prosodic and/or morphosyntactic position in which the tone occurs. This type of alternation is often referred to as tone sandhi (see Chen 2000 for a comprehensive survey on tone sandhi in Chinese dialects). Mandarin Chinese has two tone sandhi patterns, both of which involve the Third Tone 213. Specifically, 213 becomes 35 (i.e., it neutralises with Tone 2) when followed by another 213 — this is often referred to as the "Third-Tone Sandhi;"[3] but 213 becomes 21 (i.e., the second half of the complex contour is truncated) when followed by any other tone — this is often referred to as the "Half-Third Sandhi." These sandhis are exemplified in (2).

[3] Recent phonetic studies (e.g., Xu 1997, Peng 2000) have shown that this sandhi is in fact non-neutralizing—the sandhi tone is lower in overall pitch than the lexical Tone 2, but this difference cannot be reliably perceived by native adult listeners.

(2) Mandarin tone sandhi:

a. 213 → 35 / ___ 213

xaw213-tɕju213 → xaw35-tɕju213 ‘good wine’

tʂan213-lan213 → tʂan35-lan213 ‘exhibit’

b. 213 → 21 / ___ {55, 35, 51}

xaw213-ʂu55 → xaw21-ʂu55 ‘good book’

xaw213-ɻən35 → xaw21-ɻən35 ‘good person’

xaw213-k^han55 → xaw21-k^han51 ‘good-looking’

Both of these sandhi patterns are fully productive in Mandarin disyllabic words. But we consider them to differ with respect to phonetic naturalness; specifically, we consider the latter pattern to have a stronger phonetic basis than the former one. Our judgment is based on the following three reasons.

First, in the traditional Lexical Phonology sense, the Third-Tone Sandhi has the characteristics of being a lexical rule — it is structure-preserving, and its application to a polysyllabic compound is dependent on syntactic bracketing; but the Half-Third Sandhi is characteristic of a postlexical rule — it is allophonic and applies across the board. The syntactic dependency of the Third-Tone Sandhi can be seen in (3). Examples in (3a) illustrate that for underlying Third Tone sequences, the output tones differ depending on the syntactic branching structure. Examples in (3b) show that there is a special status for prepositions in permitting the blocking of the Third-Tone Sandhi on them.[4] The close

[4] For more discussion on the application of the Mandarin Third-Tone Sandhi, see Shih 1997, Zhang 1997, and Duanmu 2000.

relation between the postlexical status of a phonological rule and its phonetic motivation has been well established in the Lexical Phonology literature (e.g., Kiparsky 1982, 1985, Mohanan 1982), and we take it as one piece of evidence that the Half-Third Sandhi has a stronger phonetic motivation than the Third-Tone Sandhi.

(3) Syntactic dependency of the Mandarin Third-Tone Sandhi:

a. Left- vs. right-branching compounds (from Duanmu 2000: 238):

[213 [213 213]] → 35-35-213 or 21-35-213:

[maj	[xaw	tɕjow]]	
buy	good	wine	'buy good wine'
213	213	213	UR
35	35	213	SR1
21	35	213	SR2

[[213 213] 213] → 35-35-213 only:

[[maj	xaw]	tɕjow]	
buy	done	wine	'finished buying wine'
213	213	213	UR
35	35	213	SR
*21	35	213	

[213 [213 [213 213]]] → 35-35-35-213, 21-35-35-213, or 35-21-35-213:

[ɕjaw	[tʂɨ	[law	[xu]]]	
little	paper	old	tiger	'little paper tiger'
213	213	213	213	UR
35	35	35	213	SR1
21	35	35	213	SR2
35	21	35	213	SR3

[[[213 213] 213] 213] → 35-35-35-213 only:

[[[tʂan	lan]	kwan]	li]	
show	see	hall	inside	'inside the exhibition hall'
213	213	213	213	UR
35	35	35	213	SR
*21	35	35	213	
*35	21	35	213	

b. Special status of prepositions (from Zhang 1997: 294-295):

[213 [[213_{prep} 213] 213]] → 35-35-35-213, 21-35-35-213, or 35-21-35-213:

[ma	[[**wɑŋ**	pej]	tsow]]	
horse	to	north	walk	'The horse walks to the north.'
213	213	213	213	UR
35	35	35	213	SR1
21	35	35	213	SR2
35	21	35	213	SR3

[ma	[[**kej**	kow]	jaw]]	
horse	by	dog	bite	'The horse was bitten by the dog.'
213	213	213	213	UR
35	35	35	213	SR1
21	35	35	213	SR2
35	21	35	213	SR3

[213 [[$213_{\text{non-prep}}$ 213] 213]] → 35-35-35-213 or 21-35-35-213:

[ma	[[**xən**	ʂaw]	xow]]	
horse	very	rarely	roar	'Horses rarely roar.'
213	213	213	213	UR
35	35	35	213	SR1
21	35	35	213	SR2
*35	21	35	213	

[ma	[[**ta**	san]	tsow]]	
horse	take	umbrella	walk	'The horse walked with an umbrella.'
213	213	213	213	UR
35	35	35	213	SR1
21	35	35	213	SR2
*35	21	35	213	

Second, in terms of pure phonetics, although both the Third-Tone and Half-Third sandhis involve simplification of a complex contour in prosodic nonfinal position, which has articulatory and perceptual motivations, as Zhang (2002, 2004) has established, the Third-Tone Sandhi also involves a raising of the pitch, which cannot be accounted for by the phonetic motivation of reducing pitch contours on syllables with insufficient duration. The Half-Third Sandhi, however, only involves truncation of the second half of the contour. We take this as another argument for the stronger phonetic motivation for the Half-Third Sandhi.

Third, the Third-Tone Sandhi corresponds to a historical sandhi pattern in Chinese, namely, *shang* → *yang ping* / __ *shang*, where *shang* and *yang ping* refer to the historical tonal categories from which 213 and 35 descended respectively. This historical sandhi pattern dates back to at least the 16^{th} century (Mei 1977). And according to Mei's reconstruction, the pitch values for *shang* and *yang ping* in 16^{th} century Mandarin were low level (22) and low rising (13) respectively. The present-day rendition of the sandhi in Mandarin is the result of historical tone changes that morphed *shang* into low-falling-rising and *yang ping* into high rising. Therefore, the Mandarin Third-Tone Sandhi was not originally motivated by the phonetic rationale of avoiding complex pitch contour on short duration. The same point is made by the variable synchronic renditions of the same historical sandhi in related Mandarin dialects (Court 1985). E.g., in Tianjin, it is 13 → 45 / __ 13 (Yang et al. 1999, Chen 2000); in Jinan, it is 55 → 42 / __ 55 (Qian and Zhu 1998); in Taiyuan, it is 53 → 11 / __ 53 (Wen and Shen 1999). The Half-Third Sandhi, on the other hand, does not have a similar historical origin; and due to the different tonal shapes of the historical *shang* tone in different present-day dialects, it apparently does not have comparable synchronic renditions.

We therefore consider the Half-Third Sandhi to be phonetically more natural than the Third-Tone Sandhi. Let us note that we do not need to commit to an absolute cut-off point for what is natural and what is unnatural — we in fact do not believe that such a criterion exists. In order to identify patterns that are useful for testing of the synchronic relevance of phonetics, we only need to be able to compare two patterns on their naturalness, as we have done for the two types of sandhi in Mandarin.

The general question we pursue is whether Mandarin speakers exhibit different behaviors on the two sandhi patterns in a wug test. Specifically, we test whether there are differences in reaction time and accuracy of application between the two sandhis. In line with the synchronic approach, we hypothesize that, for the more natural Half-Third Sandhi, the reaction time will be faster, and the sandhi tone on wug words will be acoustically identical to that on real words; for the less natural Third-Tone Sandhi, the reaction time will be slower, and the sandhi tone on wug words will not be identical to that on real words. In particular, we expect some degree of non-application of the Third-Tone Sandhi to wug words. In light of the discussion on the difference between Tone 2 and Tone 3, we specifically expect a lower and later turning point and a longer duration for the sandhi tone in wug words than in real words for the Third-Tone Sandhi.

2.2. Methods

2.2.1. Stimuli construction

Following Hsieh (1970, 1975, 1976)'s experimental design for a Taiwanese wug test, we constructed five sets of disyllabic words in Mandarin. The first set includes real words, denoted by AO-AO (AO = actual occurring morpheme). This set serves as the control for the experiment and is the set with which results of wug words are compared. The other four sets are wug words: *AO-AO, where both syllables are actual occurring morphemes, but the disyllable is non-occurring; AO-AG (AG = accidental gap), where the first syllable is actual occurring, but the second syllable is an accidental gap in Mandarin syllabary; AG-AO, where the first syllable is an accidental gap and the second syllable is actual occurring; and AG-AG, where both syllables are accidental gaps. The AG's were hand-picked by both authors, who are native speakers of Mandarin Chinese. In each AG, both the segmental composition and the tone of the syllable are legal in Mandarin, but the combination happens to be missing. E.g., [p^han] is a legal syllable, but it accidentally does not occur with the Third Tone 213. Therefore, [p^han213] is a possible AG.

For each set of words, we used four different tonal combinations: the first syllable always has the Third Tone 213, and the second syllable can have Tone 1, 2, 3, or 4. Therefore, each tonal combination is in the environment to undergo either the Third-Tone or the Half-Third Sandhi. Eight words for each tonal combination were used, making a total of 160 test words (8×4×5). For half of the Tone 3 + Tone 3 combinations, the sandhi form (Tone 2) in the first syllable is an actual occurring morpheme; for the other half, the sandhi form is an accidental gap.

For AO-AO words, we controlled both the frequency and the Mutual Information Score for the disyllables using the Feng Hua Yuan character and digram frequency corpus (1998 version) compiled by Dr. Jun Da at Middle Tennessee State University.[5] The corpus contains 4,718,131 characters and 4,159,927 digrams. All disyllables fall

[5] http://lingua.mtsu.edu/chinese-computing/statistics/index.html

within the raw frequency range of 31-62. The Mutual Information Score is calculated as in (4), where *p(x,y)* represents the digram frequency, and *p(x)* and *p(y)* represent the frequencies of the two characters respectively. A higher Mutual Information Score indicates a higher likelihood for the two characters to cooccur, and hence form real words. All AO-AO words fall within the range of 8-12 for the Mutual Information Score, which is a range that indicates that all these digrams are common words.[6]

(4) $$I(x,y) = \log_2 \frac{p(x,y)}{p(x)p(y)}$$

For the four wug sets, the digram frequencies are all zero, and we used the same first syllable to combine with the four different tones in the second syllable, e.g., in AG-AG, we use [**pʰɪŋ213** ʂwən55], [**pʰɪŋ213** tʰɤ35], [**pʰɪŋ213** tsɤŋ213], and [**pʰɪŋ213** tʂa51], along with seven other such sets. The identity of the first syllable allows the comparison between the two types of sandhi that the first syllable may undergo. One caveat to the frequency control is that in Mandarin, syllables with Tone 3 are rarer than those with Tones 1, 2, and 4, in terms of both type (except for Tone 2) and token frequencies. Calculations based on a syllable frequency corpus on Dr. Jun Da's webpage[7] indicate that

---

6 Dr. Jun Da's website (see footnote 5) provides the following guidelines on how to interpret Mutual Information Scores: a Mutual Information Score greater than 3 indicates that the two words have a strong collocation; a Mutual Information Score less than 1 indicates that they are unlikely to be related; and a Mutual Information Score between 1 and 3 is in the gray area. For more information on Mutual Information Scores, see Oakes (1998).

7 http://lingua.mtsu.edu/chinese-computing/phonology/syllabletone.php

the numbers of legal syllable types with Tones 1 through 4 are 258, 224, 254, and 318 respectively, and syllables with Tones 1 through 4 account for 16.7%, 18.4%, 14.8%, and 42.5% of all syllables in the corpus, which has 192,647,157 syllables.[8] Therefore, the ideal control for this frequency effect would be to use those syllables whose concatenations with a Third-Toned σ_2 are of comparable frequency with their concatenations with a First-, Second-, or Fourth-Toned σ_2 combined. But this type of control is extremely difficult to achieve. Moreover, even if this could be achieved, we would still face the problem that the speakers simply have more practice with the Half-Third Sandhi in general. Therefore, we opted for what we considered the "next best thing" in frequency control for the wug words — identical σ_1 for all the tonal combinations.

To avoid neighborhood effects in wug words, at least to some extent, we ensure that the disyllabic combinations are not real words with any tones on either of the syllables. We specifically control for the tonal neighbors because research on homophony judgment (Taft and Chen 1992, Cutler and Chen 1997), similarity rating (Vitevitch and Zhang, in progress), phoneme (toneme) monitoring (Ye and Connine 1999), and legal-phonotactic judgment (Myers 2002) has all shown that phonemic tonal differences are perceptually less salient than segmental differences, which entails that tonal neighbors, in a sense, make closer neighbors.

Finally, to disguise the purpose of the experiment, we also used 160 filler words, half of which were real words, and the other half were wug words. The fillers also included tones other than Tone 3 in σ_1 position.

---

[8] The other 7.6% are syllables with neutral tone.

The entire set of test stimuli is given in the Appendix.

2.2.2. Experimental set-up

The experiment was conducted with SuperLab by Cedrus in the Phonetics and Psycholinguistics Laboratory at The University of Kansas. There were 320 stimuli in total (160 test items + 160 fillers). Each stimulus consisted of two monosyllabic utterances read by the first author, separated by an 800ms interval. The stimuli were played through a headphone worn by the subjects. For each stimulus, the subjects were asked to put the two syllables together and pronounce them as a disyllabic word in Mandarin as soon as they heard the second syllable. Their response was collected by a Sony PCM-M1 DAT recorder through a 33-3018 Optimus dynamic microphone placed on the desk in front of them, and also by a head-mounted microphone connected to an SV-1 Voice Key by Cedrus, which collected the reaction time. The sampling rate for the DAT recorder was 44.1kHz. The digital recording was then down-sampled to 22kHz onto a PC hard-drive using Praat — a free speech analysis software package by Boersma and Weenink.[9] The recorded reaction time was the duration between the end of audio file for the second syllable and the time at which the subject's response reached a level preset on the Voice Key. This preset level was kept consistent for all subjects. There was a 2000ms interval between stimuli. If the subject did not respond within 2000ms after the second syllable played, the next stimulus would begin. The stimuli were divided into two same-sized blocks (A and B) with matched stimulus types, and there was a five-minute break between the blocks. Half of the subjects took block A first, and the other half took

[9] http://www.praat.org/

block B first. Within each block, the stimuli were automatically randomised by SuperLab. Before the experiment began, there was a short introduction in Chinese that the subjects heard through the headphone and also read on a computer screen in front of them, which explained their task both in prose and through examples. There was then a practice session that consisted of 14 words from the experiment (two of each from AO-AO, *AO-AO, AO-AG, AG-AO, and AG-AG; two real-word fillers, two wug fillers). The experiment began after a verbal confirmation from the subjects that they were ready. The entire experiment took around 45 minutes.

2.2.3. Subjects

Twenty native speakers of Mandarin Chinese (12 male, 8 female) participated in the experiment. They were recruited through flyers on KU campus and word of mouth. Except for one speaker who was 45 years old and had been in the US for 20 years, all speakers ranged from 23 to 35 years in age and had been in the US for less than four years at the time of the experiment. Each subject was paid $10 for participating in the study.

2.2.4. Data analyses

The raw reaction time returned from the SuperLab Voice Key is the duration from the end of second syllable to beginning of the subjects verbal response. But the canonical reaction time measure in auditory lexical decision and psycholinguistic production studies is from the onset of the stimulus to the onset of the verbal response. Given that the first syllables in the stimuli are matched across the tone conditions, we measured the durations from the onset of the second syllables to the end of the stimuli and added these durations to the raw reaction time data. Therefore, the reaction time data reported here are as in Figure 2. The tokens for which the subjects did not provide a response within the 2000ms inter-stimuli interval were not included in the reaction time analysis. 89 tokens (out of a total of 3,200) were discarded this way. To reduce the effect of outliers and ensure the normal distribution of the RT data for the Analysis of Variance, we took the log of the RT data in *ms* (Ratcliff 1993).

Using SPSS, we conducted a two-way Repeated-Measures Analysis of Variance with Tone and Group as independent variables for RT. The Tone variable has four levels — Tones 1, 2, 3, and 4 in σ_2 position; the Group variable has five levels — AO-AO, *AO-AO, AO-AG, AG-AO, and AG-AG. To restate the crucial hypothesis, we expect the reaction time to be shorter for cases involving the Half-Third Sandhi (3+1, 3+2, 3+4) than those involving the Third-Tone Sandhi (3+3).

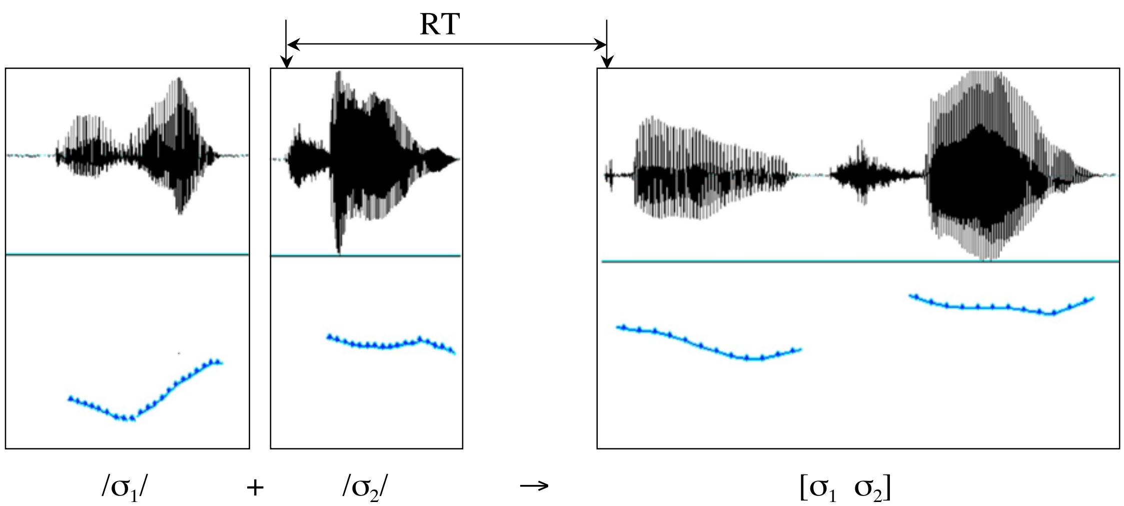


Figure 2. Reaction Time (RT) measurements

A confound with using this RT measure is that this does not take into account the possible differences in "uniqueness point" (Marslen-Wilson 1990, Goldinger 1996) of the second syllable, which refers to the point at which all competitors of the syllable are ruled out during processing, among different tones and different word types. The uniqueness point of a Mandarin syllable depends on both the segmental and tonal compositions of the syllable. To parcel out this effect while lacking the uniqueness point information, we conducted two Analyses of Covariance, with the RT as the dependent variable, Tone and Group as the independent variables, and the duration of the second syllable as the covariate for both analyses. See Goodman and Huttenlocher (1988) for discussions of analytical issues of RT and the recommendation that the stimulus duration be included as a covariate to statistical analyses.

To investigate the accuracy of sandhi application, we extracted the f_0 of the rhyme in σ_1 in the subjects' response using Praat. We then took a f_0 measurement every 10% of the duration of the rhyme, giving eleven f_0 measurements for each rhyme. For each tonal combination (3+1, 3+2, 3+3, 3+4), we did two comparisons. The first is between AO-AO and the rest of the tone groups *AO-AO, AO-AG, AG-AO, and AG-AG; i.e., real disyllables vs. wug disyllables. The other is between AO-AO, *AO-AO, AO-AG and AG-AO, AG-AG; i.e., real σ_1's vs. wug σ_1's. The rationale for the two comparisons is that lexical listing could be at the disyllabic word or monosyllabic morpheme level. By doing both comparisons, it may give us a better idea of how the listing of surface tones, if any, works in Mandarin. For 3+3 combinations, we separated the cases in which the sandhi syllable is an actual occurring morpheme from those in which the sandhi syllable is an accidental gap. It turned out that these two types of cases have no significant differences from each other. Therefore, in §3, we only report results in which they were pooled together. Our hypothesis for these comparisons is that the sandhi tones in real words and wug words are identical for cases of Half-Third Sandhi due to its clear phonetic basis, but they are different for cases of Third-Tone Sandhi because of its less clear phonetic motivation. In particular, we expect a certain degree of non-application of the Third-Tone Sandhi in wugs, i.e., Tone 3 in σ_1 will resist the change to Tone 2. Again, given what we know about the acoustic characteristics of these tones, this translates into a lower and later turning point and a longer duration for the sandhi tone in wug words than in real words.

Among the twenty speakers, there were two speakers (one male and one female) whose f_0 values could not be reliably measured by Praat due to high degrees of creakiness in their voice. We discarded these speakers' data in the f_0 analysis.

Figure 3 illustrates how we compared two f_0 curves. We conducted a two-way Repeated Measures ANOVA, with Curve and Point as independent variables. The Curve variable has two levels — Curve 1 and Curve 2, with Curve 1 representing real disyllables for 3+3 and Curve 2 representing wug disyllables, for example, and the significance of this variable for the f_0 value would indicate that the two curves have different average pitches. The Point variable has eleven levels, representing the eleven points where f_0 data are taken. A significant interaction between Curve and Point would indicate that the two curves have different shapes. This method of comparing two f_0 curves has been used by Peng (2000).

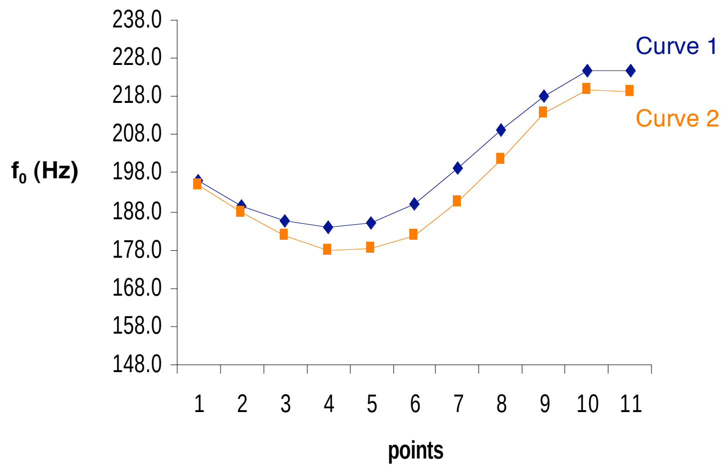


Figure 3. Comparing two f_0 curves.

For σ_1 in 3+3 combinations, we also measured the f_0 drop and the duration from the beginning of the rhyme to the pitch turning point, as shown in Figure 4. Comparisons between real and wug disyllables and between real and wug σ_1's on these measurements were made using one-way Repeated-Measures ANOVA's. We expect the f_0 drop to be greater and the TP duration to be longer for wug words than real words.

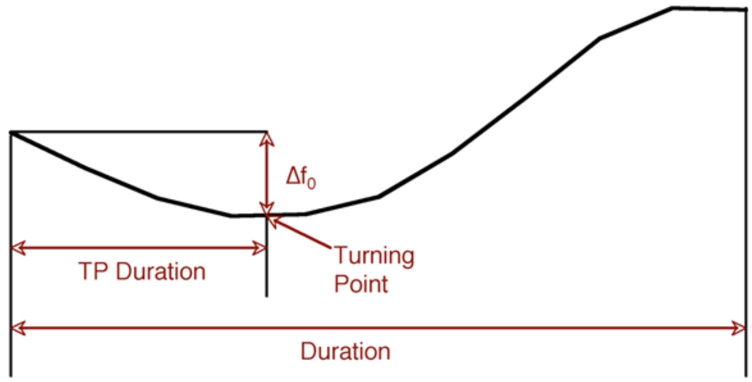


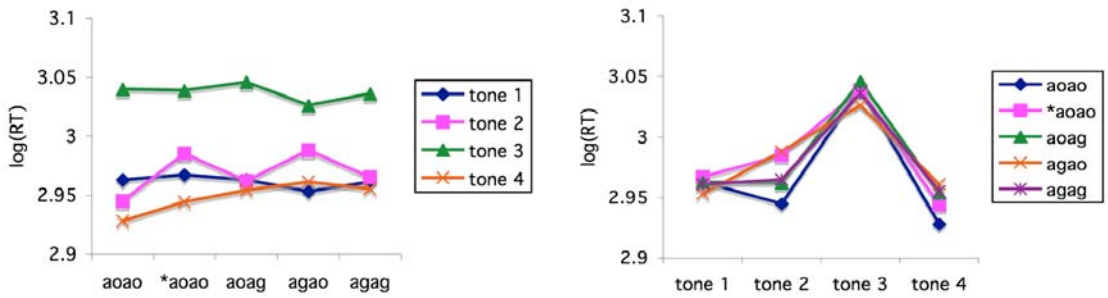
Figure 4. Schematic of the measurements taken from the pitch curve of the rhyme in σ_1 in 3+3 combinations. "Δf_0" and "TP Duration" are the pitch drop and duration from the beginning of the rhyme to the turning point, respectively. "Duration" is the entire rhyme duration.

Finally, we measured the σ_1 rhyme duration for all the disyllabic combinations and compared real and wug disyllables and real and wug σ_1's for each tonal combination using one-way Repeated-Measures ANOVA's. We expect to find a longer rhyme duration for the wug words in 3+3 combination, but no difference between wug and real words in other combinations.

3. Results and discussions

3.1. Reaction time

The reaction time data are given in Figure 5. Figure 5a graphs the five groups on the *x*-axis, and Figure 5b graphs the four tones on the *x*-axis. The *y*-axis represents the natural log value of the RT in ms.



(a) (b)

Figure 5. Reaction time data. (a) graphs the five groups on the *x*-axis; (b) graphs the four tones on the *x*-axis. The *y*-axis represents the natural log value of the RT in ms.

With Tone and Group as independent variables, a two-way Huynh-Feldt ANOVA shows that the effect of Tone on RT is significant ($F(2.682, 50.964)=183.802$, $p<0.001$), the effect of Group is not significant ($F(2.349, 44.638)=2.817$, $p>0.05$), and the interaction between Tone and Group is significant ($F(11.670, 221.729)=6.119$, $p<0.001$). Pairwise comparisons on Tone with Bonferroni corrections show that Tone 3 has significantly longer RTs than all other tones (all at $p<0.001$), and Tone 1 has significantly longer RTs than Tone 4 ($p<0.05$). Pairwise comparisons on Group show no significant differences in any pairs.

The ANCOVAs with the second syllable duration as a covariate indicate that the effect of Tone on RT is still significant ($F(3, 395)=8.832$, $p<0.001$), and the effect of Group is also significant ($F(4, 394)=6.990$, $p<0.001$). Pairwise comparisons on Tone with Bonferroni corrections indicate that Tone 3 has significantly longer RTs than all other tones (all at $p<0.001$), and there are no other significant effects among Tones 1, 2, and 4. Pairwise comparisons on Group indicate that AO-AO has significantly shorter RTs than AG-AG and AO-AG (both at $p<0.001$), but there are no significantly differences in other pairs.

Therefore, the RT results support our hypothesis that a Tone 3 in σ_2 position induces a longer reaction time than other tones due to the lesser phonetic motivation that a 3+3 sandhi involves. The reaction time is shorter for real words (AO-AO) than some wug words, but not all of them (only AG-AG and AO-AG, not *AO-AO and AG-AO). It is not clear why AO-AO does not consistently give rise to shorter reaction time than all wugs, and the exact influence of the nature of disyllables on reaction time awaits further research.

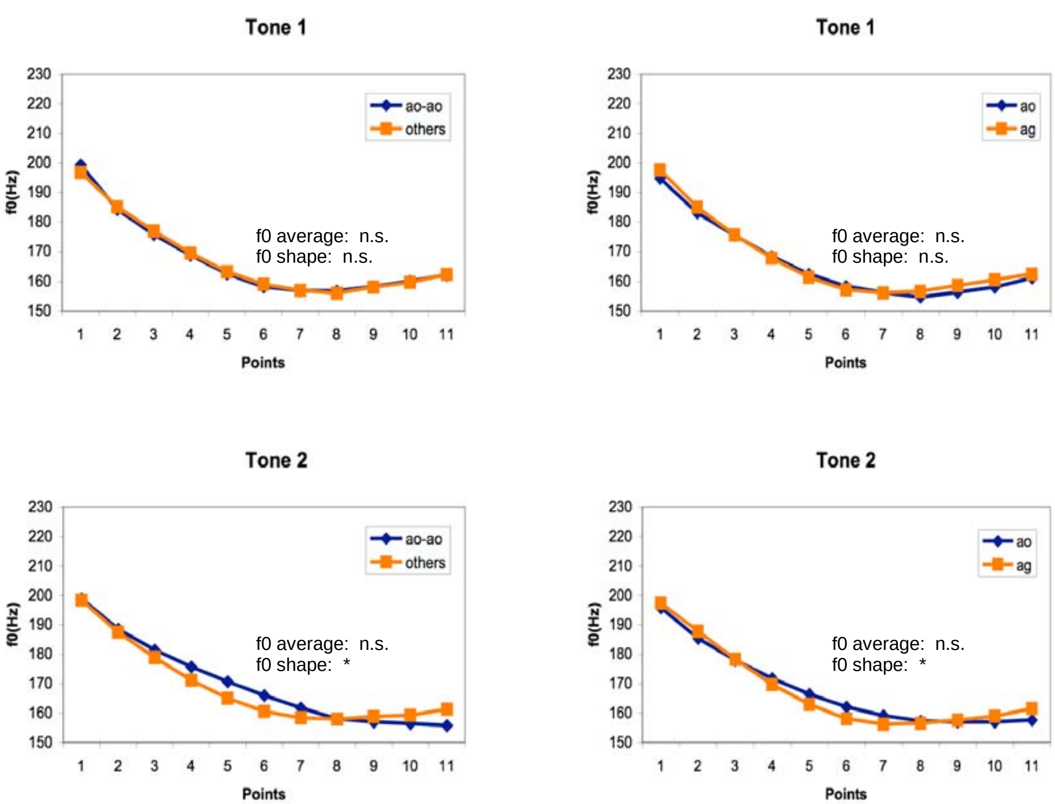
3.2. f_0 contour

3.2.1. Average f_0 and contour shape

In this section, we first report results of the two comparisons (real disyllables vs. wug disyllables and real σ_1's vs. wug σ_1's) for the natural Half-Third Sandhi (σ_2 = Tone 1, 2, or 4), then report results for the less natural Third-Tone Sandhi (σ_2 = Tone 3). As a reminder, our hypothesis is that the speakers should perform the Half-Third Sandhi in wug words identically to real words, but for the Third-Tone Sandhi, there should be a certain degree of non-application in the wug words, as indicated by a lower and later turning point and a longer overall duration for the sandhi tone.

The results from the Half-Third Sandhi comparisons are given in Figure 6. The three graphs in Figure 6a represent the real disyllable vs. wug disyllable comparisons (AO-AO vs. *AO-AO, AO-AG, AG-AO, AG-AG) for Tones 1, 2, and 4. The three graphs in Figure 6b represent the real σ_1 vs. wug σ_1 comparisons (AO-AO, *AO-AO, AO-AG vs. AG-AO, AG-AG) for Tones 1, 2, and 4. As discussed in §2.2.4, the average f_0 comparisons are indicated by the main effect of Curve, which has two levels representing the two curves, in a two-way Repeated-Measures ANOVA, with Curve and Point as independent factors, and the f_0 shape comparisons are determined by the interaction between the two factors Curve and Point. We again used the Huynh-Feldt Repeated-Measures ANOVA to correct for sphericity violations.

(a) AO-AO vs. others (b) AO vs. AG in σ_1



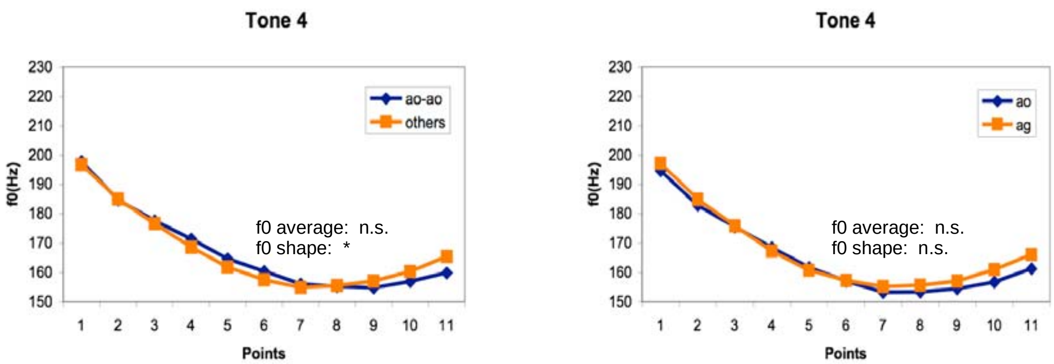


Figure 6. f_0 curves of the first syllable for the Half-Third Sandhi. The three graphs in the (a) column represent the real disyllable vs. wug disyllable comparisons for Tones 1, 2, and 4. The three graphs in the (b) column represent the real σ_1 vs. wug σ_1 comparisons for Tones 1, 2, and 4. "n.s." indicates no significant difference. "*" indicates significant difference at $p<0.05$.

As we can see in Figure 6, for both Tone 1 and Tone 4, the Half-Third Sandhi, subjects' performance on wug words is generally identical to that on real words in terms of both the average f_0 and the f_0 contour shape. This is true for both the disyllabic and σ_1 comparisons for Tone 1 and the σ_1 comparisons for Tone 4. When σ_2 has Tone 2, the f_0 contour on σ_1 has a significantly different shape between real and wug words for both comparisons. The statistical results for these comparisons are given in Table 1.

Real disyllable vs. wug disyllable			
	Tone 1	Tone 2	Tone 4
Curve (f_0 average)	F(1.000, 17.000) =0.005, p=0.945	F(1.000, 17.000) =0.805, p=0.382	F(1.000, 17.000) =0.000, p=1.000
Point	F(3.187, 54.180) =125.614, p=0.000	F(2.119, 36.023) =168.840, p=0.000	F(2.663, 45.263) =133.073, p=0.000
Curve × Point (f_0 shape)	F(3.574, 60.750) =0.880, p=0.472	F(2.824, 48.012) =13.036, p=0.000	F(3.436, 58.409) =3.535, p=0.016
Real σ_1 vs. wug σ_1			
	Tone 1	Tone 2	Tone 4
Curve (f_0 average)	F(1.000, 17.000) =0.061, p=0.808	F(1.000, 17.000) =0.000, p=0.997	F(1.000, 17.000) =0.189, p=0.670
Point	F(3.275, 55.680) =167.524, p=0.000	F(2.143, 36.439) =178.423, p=0.000	F(2.651, 45.059) =117.356, p=0.000
Curve × Point (f_0 shape)	F(2.545, 43.265) =2.178, p=0.113	F(3.150, 53.546) =9.072, p=0.000	F(2.942, 50.011) =2.265, p=0.093

Table 1. Two-way Huynh-Feldt Repeated-Measures ANOVA results for the first syllable f_0 curves in Half-Third Sandhi. The two factors are Curve and Point, with two and eleven levels respectively. For real disyllables vs. wug disyllables, the two levels for Curve represent (a) AO-AO, and (b) all other groups (*AO-AO, AO-AG, AG-AO, and AG-AG). For real σ_1 vs. wug σ_1, the two levels for Curve represent (a) AO-AO, *AO-AO, AO-AG, and (b) AG-AO, AG-AG. The eleven levels for Point represent the measurement points on the raw f_0 curves.

The results from the Third-Tone Sandhi comparisons are given in Figure 7. As we already mentioned in §2.2.4, we pooled together cases in which the sandhi syllable is an actual occurring morpheme and those in which the sandhi syllable is an accidental gap, as they present no significant differences from each other. The two graphs in Figure 7 represent the real disyllable vs. wug disyllable (AO-AO vs. *AO-AO, AO-AG, AG-

AO, AG-AG) and real σ_1 vs. wug σ_1 (AO-AO, *AO-AO, AO-AG vs. AG-AO, AG-AG) comparisons, respectively. Two-way Huynh-Feldt Repeated-Measures ANOVA's indicate that although the average f_0 is the same for both comparisons, the f_0 contour shape is significantly different between the real words and wug words for both comparisons. The ANOVA results are summarised in Table 2.

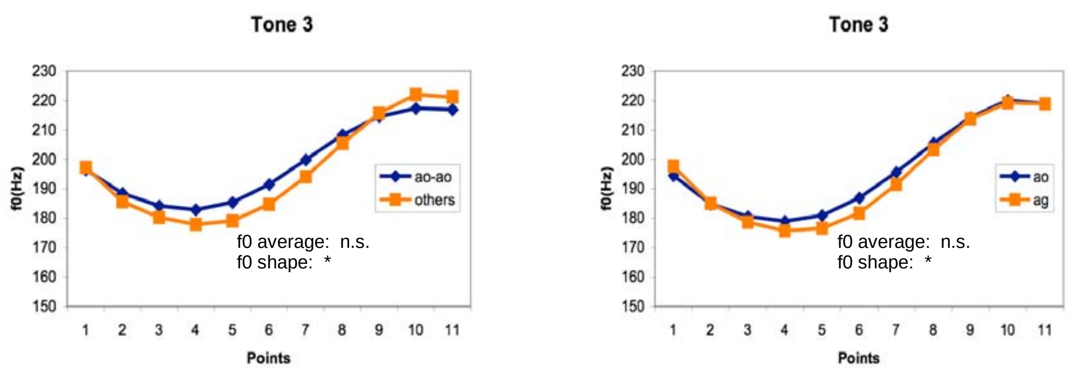


Figure 7. f_0 curves of the first syllable for the Third-Tone Sandhi. The two graphs represent the real disyllable vs. wug disyllable and real σ_1 vs. wug σ_1 comparisons, respectively. "n.s." indicates no significant difference. "*" indicates significant difference at p<0.01.

Real disyllable vs. wug disyllable	
	Tone 3
Curve (f_0 average)	F(1.000, 17.000)=1.351, p=0.261
Point	F(2.371, 40.312)=73.135, p=0.000
Curve × Point (f_0 shape)	F(2.414, 41.031)=9.537, p=0.000
Real σ_1 vs. wug σ_1	
	Tone 3
Curve (f_0 average)	F(1.000, 17.000)=0.000, p=0.997
Point	F(2.143, 36.439)=178.423, p=0.000
Curve × Point (f_0 shape)	F(3.150, 53.546)=9.072, p=0.000

Table 2. Two-way Huynh-Feldt Repeated-Measures ANOVA results for the first syllable f_0 curves in Third-Tone Sandhi. The two factors are Curve and Point, with two and eleven levels respectively. For real disyllables vs. wug disyllables, the two levels for Curve represent (a) AO-AO, and (b) all other groups (*AO-AO, AO-AG, AG-AO, and AG-AG). For real σ_1 vs. wug σ_1, the two levels for Curve represent (a) AO-AO, *AO-AO, AO-AG, and (b) AG-AO, AG-AG. The eleven levels for Point represent the measurement points on the raw f_0 curves.

From Figure 7, we can also see that for the curves representing wug words ("others" in the first graph, "ag" in the second graph), the turning points are both lower and later than their counterparts for the curves representing real words, indicating that there may be a certain degree of non-application of the sandhi. We report quantitative results on these turning points in the next section.

3.2.2. Characteristics of the turning point

For σ_1 in 3+3 combinations, Δf_0 is defined as the difference between the f_0 of the beginning of the rhyme and the f_0 turning point in the rhyme; TP duration is the duration from the beginning of the rhyme to the turning point. Results of comparisons between real and wug disyllables and between real and wug σ_1's on f_0 are given in Figure 8 and Figure 9. One-way Repeated-Measures Huynh-Feldt ANOVA's with Curve as the independent factor (two levels) indicate that for Δf_0, AO-AO is significantly different from other groups ($F(1.000, 17.000)=8.543$, $p<0.01$), so is σ_1=AO from σ_1=AG (F(1.000, 17.000)=48.254, $p<0.001$); for TP duration, AO-AO is significantly different from other groups ($F(1.000, 17.000)=19.561$, $p<0.001$), so is σ_1=AO from σ_1=AG ($F(1.000, 17.000)=21.343$, $p<0.001$). These results support our hypothesis: with a lower and later turning point, the sandhi tone on wug words is more similar to the original Tone 3 than the sandhi tone on real words, indicating a certain degree of non-application of the sandhi in wug words.

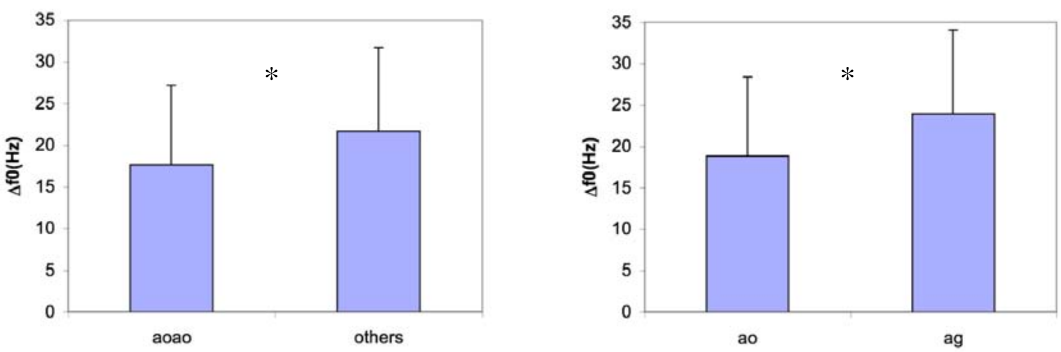


Figure 8. Δf_0 — f_0 drop from the beginning of the rhyme to the f_0 turning point. The two graphs represent the real disyllable vs. wug disyllable and real σ_1 vs. wug σ_1 comparisons, respectively. Error bars are one standard deviations. "*" indicates significant difference at $p<0.05$.

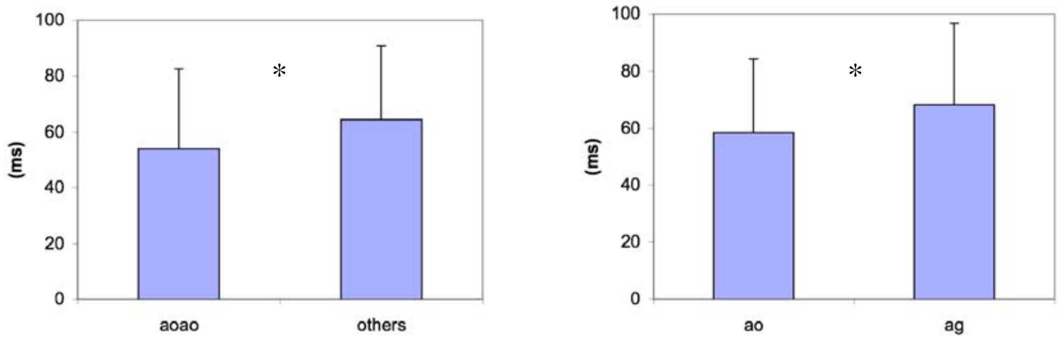


Figure 9. TP duration — duration from the beginning of the rhyme to the f_0 turning point. The two graphs represent the real disyllable vs. wug disyllable and real σ_1 vs. wug σ_1 comparisons, respectively. Error bars are one standard deviations. "*" indicates significant difference at $p<0.05$.

3.3. Rhyme duration

The results for σ_1 rhyme duration for all the tonal combinations are given in Figure 10, and the statistical results are summarised in Table 3. One-way Repeated-Measures Huynh-Feldt ANOVA's with Curve as the independent factor (two levels) show that there are no significant differences between AO-AO and other groups for any of the tonal combinations. But for 3+3, the difference approaches significance at $p<0.05$ ($F(1.000, 17.000)=4.218$, $p=0.056$), and the difference is in the direction that we expected, i.e., wug > real. For AO vs. AG, 3+3 is the only combination in which the wug words have a significantly longer σ_1 rhyme duration than the real words ($F(1.000, 17.000)=5.653$, $p<0.05$). To a great extent, these results support our hypothesis: the durational property for the sandhi tones is identical between real and wug words for the Half-Third Sandhi, but for the Third-Tone Sandhi, the sandhi tone on wug words is longer than that on real words, indicating again a certain degree of non-application of the sandhi in wug words.

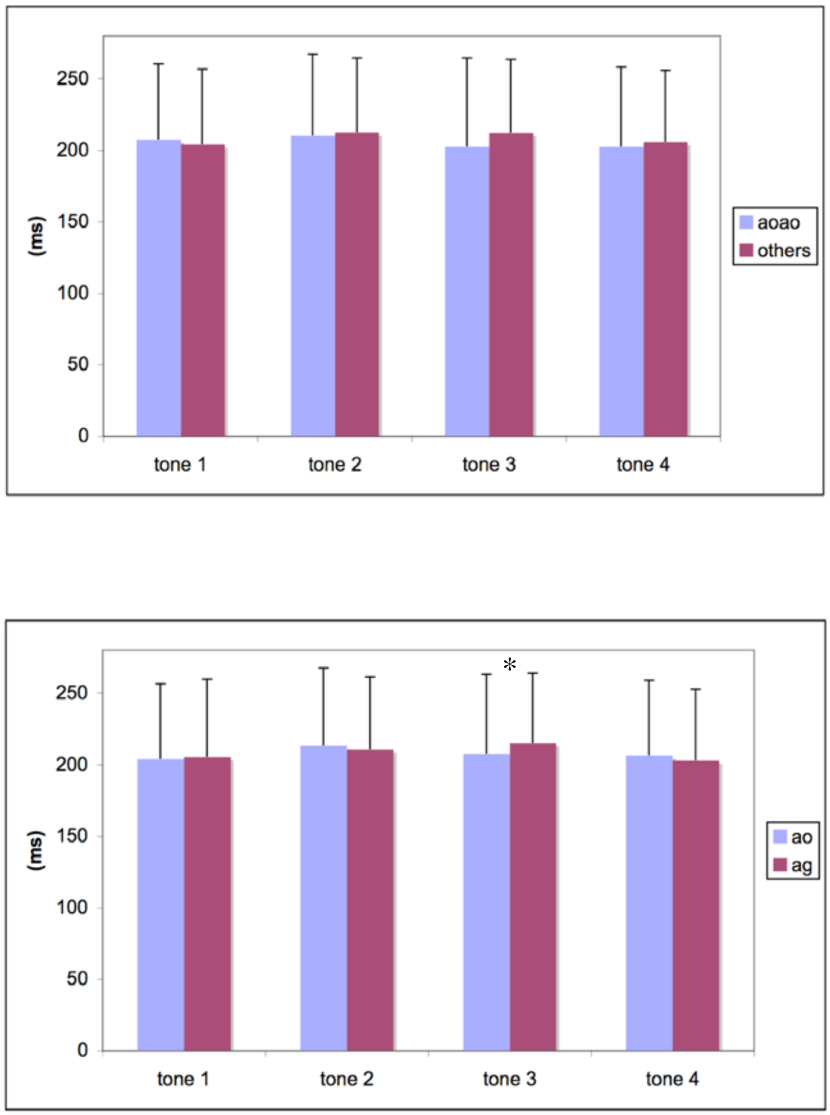


Figure 10. Rhyme duration of σ_1 for all tonal combinations. The two graphs represent the real disyllable vs. wug disyllable and real σ_1 vs. wug σ_1 comparisons, respectively. Error bars represent one standard deviation. "*" indicates significant difference at $p<0.05$.

Real disyllable vs. wug disyllable	
Tone 3 + Tone 1	F(1.000, 17.000)=0.660, p=0.428
Tone 3 + Tone 2	F(1.000, 17.000)=0.206, p=0.656
Tone 3 + Tone 3	**F(1.000, 17.000)=4.218, p=0.056**
Tone 3 + Tone 4	F(1.000, 17.000)=0.620, p=0.442
Real σ_1 vs. wug σ_1	
Tone 3 + Tone 1	F(1.000, 17.000)=0.097, p=0.759
Tone 3 + Tone 2	F(1.000, 17.000)=0.559, p=0.465
Tone 3 + Tone 3	**F(1.000, 17.000)=5.653, p=0.029**
Tone 3 + Tone 4	F(1.000, 17.000)=1.118, p=0.305

Table 3. One-way Huynh-Feldt Repeated-Measures ANOVA results for the σ_1 rhyme duration in all tonal combinations. The independent factor is Curve. The two levels of the factors are AO-AO and "others" for the real vs. wug disyllabic comparisons, and σ_1=AO and σ_1=AG for the real vs. wug σ_1 comparisons.

4. Discussion

With respect to RT, our wug test results support the hypothesis that the sandhi that has a stronger phonetic basis — the Half-Third Sandhi — applies faster in wug words than the sandhi with a weaker phonetic basis — the Third-Tone Sandhi. With respect to the accuracy of sandhi application, our results of the Third-Tone Sandhi indicate a clear difference between real words and wug words in the contour shape of the sandhi tone; in particular, the contour shape of the sandhi tone in wug words shares a greater similarity with the original Tone 3 by having a lower and later turning point and a longer tone duration. The real vs. wug comparison for the Half-Third Sandhi, however, showed identical contour shape of the sandhi tone for Tone 1, a marginal contour shape difference for Tone 4 — marginal because the difference is only significant at the 0.05

level for the disyllabic comparison, but all but disappears for the σ_2 comparison, and a significant contour shape difference for Tone 2. Furthermore, there is an independent reason for the significant difference we found for Tone 2, namely, subjects' possible confusion between Tone 2 and Tone 3. The Tone 2 graphs in Figure 6 show that the sandhi tones for the wug words have an obvious low turning point, indicating that they resemble the original Tone 3 or its sandhi tone in a 3+3, not 3+2 context. Therefore, it is likely that some speakers have mistaken the Tone 2 in σ_2 position as Tone 3 in some tokens of wug words. Given that real words are not expected to have this problem, this may have caused the difference between real and wug words in 3+2 combination. From perceptual and acquisition studies of Mandarin tones, this is also a likely scenario. Perceptual studies have shown that Tone 2 and Tone 3 are the most easily confusable tonal pairs in Mandarin for both native speakers (Chuang et al. 1972, Zue 1976, Shen and Lin 1991, Whalen and Xu 1992) and second language learners (Kiriloff 1969); and the acquisition study by Li and Thompson (1977) has shown that the Tone 2 ~ Tone 3 contrast is the tonal contrast that emerges the latest in children's acquisition of Mandarin phonology.

Therefore, we conclude that our wug tests of the two tone sandhi patterns in Mandarin Chinese generally support the relevance of phonetics to synchronic phonology: the sandhi that has a stronger phonetic basis — the Half-Third Sandhi — generally applies faster and with greater accuracy in wug words than the sandhi with a weaker phonetic basis — the Third-Tone Sandhi. Moreover, the way in which the contour shape in 3+3 wug words differs from real words can be characterised by a greater similarity between the sandhi tone and the original Tone 3. This further suggests that phonological patterns with different degrees of phonetic basis are not only different in how they are encoded in synchronic phonology, but that the difference is in the direction predicted by a phonetically-informed synchronic phonology — the process that has a stronger phonetic basis is more productive.

In our statistics, we specifically made two comparisons between wug words and real words — wug disyllables vs. real disyllables and wug σ_1 vs. real σ_1 — in the hopes that they would reveal whether sandhi tones are lexically listed in Mandarin, and if so, whether the listing includes the whole disyllabic word or just the monosyllabic morphemes in different positions. But the two comparisons turned out extremely similar. Therefore, it is difficult for us to make any claims on the nature of lexical listings of sandhi tones beyond the statement that there is a difference between sandhi patterns that differ in the degree of their phonetic basis in either lexical listing of the sandhi tones or how the phonological grammar derives the sandhi tones from the underlying tones.

Finally, although the main goal of this research is to show that there *is* a difference in productivity to novel forms between phonological patterns that differ in phonetic naturalness, and hence the phonetic motivations must be relevant to synchronic phonological grammar in *some* way, we recognise that we have not provided a solid answer to the question "what must the grammar look like to produce such subtle differences in productivity?" We hope to pursue this question in our future research. In the meantime, we would like to point to Wilson (2006)'s substantive biased model of phonology as a possible direction for answering this question.

5. Conclusion

In this paper, we have proposed a novel research paradigm to test the relevance of phonetics to synchronic phonology — wug testing of patterns differing in phonetic naturalness that coexist in the same language. We believe that this is an improvement over the other research paradigms that have been used to test this issue, e.g., the study of phonological acquisition in a first language and the artificial language paradigm, in that it directly addresses existing native patterns, and in the meantime allows a better control of confounding factors such as lexical frequency. The language we used was Mandarin Chinese, which has two tone sandhi patterns that differ in their degrees of phonetic motivation, and our wug tests showed that Mandarin speakers applied the phonetically more natural sandhi — the Half-Third Sandhi — to wug words with a faster reaction time and a greater accuracy than the phonetically less natural sandhi — the Third-Tone Sandhi, thus supporting the direct relevance of phonetics to synchronic phonology. A diachronic-only view of phonetics-phonology relationship assumes that speakers treat phonological patterns equally as patterns handed to them by language history irrespective of their phonetic nature. This assumption is clearly contradicted by our findings.

The significance of our research also lies in the identification of a set of languages that can serve as a rich test-bed for the phonetics-phonology relationship. There are many other Chinese dialects, especially the Wu and Min dialects, that have considerably more intricate patterns of tone sandhi than Mandarin. Invariably in these dialects, some sandhi patterns have stronger phonetic bases than others. This will allow repeated testing of the results that we gathered for Mandarin and improvement, possibly elimination, of the frequency control problem that we grappled with in Mandarin. Moreover, we surmise that the tone sandhi patterns in Chinese are likely the *only* patterns in the world's languages that will allow us to test the synchronic relevance of phonetics using the wug-test paradigm. The two patterns under comparison need to satisfy the following conditions: (a) they have comparable triggering environments; (b) they are of comparable productivity in the native lexicon; (c) they have comparable frequencies of occurrence in the native lexicon; and (d) they differ in phonetic naturalness. Aside from the tone sandhi patterns that the Chinese dialects are laden with, this type of comparisons is extremely difficult to come by. We hope our study on Mandarin has opened the door for similar research in other Chinese dialects, which will make further contributions to the phonetics-phonology interface debate.

Finally, starting from Hsieh's seminal works on wug-testing Taiwanese tone sandhi (Hsieh 1970, 1975, 1976), the psychological reality of complicated tone sandhi patterns, especially those that involve phonological opacity (e.g., the tone circle in Southern Min, where Tone A → Tone B, Tone B → Tone C, Tone C → Tone D, but Tone D → Tone A in nonfinal positions; see Chen 2000 for examples) and syntactic dependency (e.g., the different sandhi patterns that Subject-Predicate and Verb-Object compounds undergo in Pingyao; see Hou 1980), has been an outstanding question in Chinese phonology. We hope that the research method that we have developed, together with our results on Mandarin, will inspire more psycholinguistic testing of these patterns that will answer this long-standing question.

References

Albright, Adam (2002). Islands of reliability for regular morphology: Evidence from Italian. *Language* 78.4: 684-709.

Albright, Adam, Argelia Andrade, and Bruce Hayes (2001). Segmental environments of Spanish diphthongization. In Adam Albright and Taehong Cho (eds.), *UCLA Working Papers in Linguistics* 7 (*Papers in Phonology* 5): 117-151.

Albright, Adam and Bruce Hayes (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90: 119-161.

Anderson, Stephen (1981). Why phonology isn't "natural." *Linguistic Inquiry* 12: 493-539.

Archangeli, Diana and Douglas Pulleyblank (1994). *Grounded phonology*. Cambridge, MA: The MIT Press.

Aslin, Richard N., Jenny R. Saffran, and Elissa L. Newport (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9: 321-324.

Berko, Jean (1958). The child's learning of English morphology. *Word* 14: 150-177.

Blevins, Juliette (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge, UK: Cambridge University Press.

Blevins, Juliette and Andrew Garrett (1998). The origins of consonant-vowel metathesis. *Language* 74: 508-556.

Bley-Vroman, Robert (1988). The fundamental character of foreign language learning. In William E. Rutherford and Michael Sharwood-Smith (eds.), *Grammar and* *second language teaching: A book of readings*. New York: Newbury House/Harper and Row. 19-30.

Boersma, Paul (1998). *Functional phonology: Formalizing the interaction between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Buckley, Eugene (2000). On the naturalness of unnatural rules. *Proceedings from the Second Workshop on American Indigenous Languages, UCSB Working Papers in Linguistics*, Vol. 9.

Buckley, Eugene (2002). *Rule naturalness and the acquisition of phonology*. Talk presented at the 2nd North American Phonology Conference, Concordia University, Montréal.

Buckley, Eugene (2003). Children's unnatural phonology. *Proceedings of the Berkeley Linguistics Society* 29: 523-534.

Bybee, Joan and Elly Pardo (1981). On lexical and morphological conditioning of alternations: A nonce-probe experiment with Spanish verbs. *Linguistics* 19: 937-968.

Chambers, Kyle E., Kristine H. Onishi, and Cynthia Fisher (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition* 87: B69-B77.

Chao, Yuen Ren (1948). *Mandarin primer: An intensive course in spoken Chinese*. Cambridge, MA: Harvard University Press.

Chao, Yuen Ren (1968). *A grammar of spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, Matthew Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge, UK: Cambridge University Press.

Chomsky, Noam and Morris Halle (1968). *The sound pattern of English*. New York: Harper and Row.

Chomsky, Noam (1981). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Chuang, C.-K., S. Hiki, T. Sone, and T. Nimura (1972). The acoustical features and perceptual cues of the four tones of standard colloquial Chinese. In *Proceedings of the 7th International Congress of Acoustics*, vol. 3, 297-300. Budapest: Akademial Kiado.

Cook, Vivian J. (1969). The analogy between first and second language learning. *International Review of Applied Linguistics* 7.3: 207-216.

Cook, Vivian J. (1994). The metaphor of access to Universal Grammar. In Nick C. Ellis (ed.), *Implicit and explicit learning of languages*. London: Academic Press. 477-502.

Court, Christopher (1985). Observations on some cases of tone sandhi. In Graham Thurgood, James A. Matisoff, and David Bradley (eds.), *Linguistics of the Sino Tibetan area: The state of the art*. Canberra: Department of Linguistics, Research School of Pacific Studies, Australian National University. 125-137.

Cutler, Anne and Hsuan-Chih Chen (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics* 59.2: 165-179.

Davidson, Lisa (to appear). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. To appear in *Journal of Phonetics*.

Dell, Gary S., Kristopher D. Reed, David R. Adams, and Antje S. Meyer (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 1355-1367.

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

Dulay, Heidi C., Marina Burt, and Stephen Krashen (1982). *Language two*. Oxford: Oxford University Press.

Ellis, Rod (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Fleischhacker, Heidi (2001). Cluster-dependent epenthesis asymmetries. In Adam Albright and Taehong Cho (eds.), *UCLA Working Papers in Linguistics* 7, *Papers in Phonology* 5: 71-116.

Flemming, Edward (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18.1: 7-44.

Goldinger, Stephen D. (1996). Auditory lexical decision. *Language and Cognitive Processes* 11: 559-567.

Goodman, Judith C. and Janellen Huttenlocher (1988). Do we know how people identify spoken words? *Journal of Memory and Language* 27: 684-698.

Hale, Mark and Charles Reiss (2000). Substance abuse and dysfunctionalism: current trends in phonology. *Linguistic Inquiry* 31: 157-169.

Hansson, Gunnar (2001). *Theoretical and Typological Issues in Consonant Harmony*. Ph.D. dissertation, University of California, Berkeley.

Hayes, Bruce and Donca Steriade (2004). Introduction: The phonetic bases of phonological markedness. In Hayes et al. (2004). 1-33.

Hayes, Bruce, Donca Steriade, and Robert Kirchner (2004). *Phonetically based phonology*. Cambridge, UK: Cambridge University Press.

Hou, Jing-Yi (1980). Pingyao fangyan de liandu biandiao (Tone sandhi in the Pingyao dialect). *Fangyan* (Dialects) 1980.1: 1-14.

Hsieh, Hsin-I (1970). The psychological reality of tone sandhi rules in Taiwanese. In *Papers from the 6th Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society. 489-503.

Hsieh, Hsin-I (1975). How generative is phonology. In E. F. Koerner (ed.), *The transformational-generative paradigm and modern linguistic theory*. Amsterdam: John Benjamins. 109-144.

Hsieh, Hsin-I (1976). On the unreality of some phonological rules. *Lingua* 38: 1-19.

Hyman, Larry (2001). On the limits of phonetic determinism in phonology: *NC revisited. In Elizabeth Hume and Keith Johnson (eds.), *The Role of Speech Perception in Phonology*. New York: Academic Press. 141-186.

Jun, Jongho (1995). *Perceptual and articulatory factors in place assimilation: An Optimality-Theoretic approach*. Ph.D. dissertation, UCLA.

Jun, Jongho (2004). Place assimilation. In Hayes et al. (2004). 58-86.

Jusczyk, Peter W., Paul Smolensky, Karen Arnold, and Elliott Moreton (2003). Acquisition of nasal place assimilation by 4.5-month-old infants. In Derek Houston, Amanda Seidl, George J. Hollich, Elizabeth Johnson, and Ann Marie Jusczyk (eds.), *Jusczyk Lab Final Report*. Retrieved from http://hincapie.psych.purdue.edu/Jusczyk

Kang, Yoonjung (2004). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20.2: 173-218.

Kenstowicz, Michael (to appear). Salience and similarity in loanword adaptation: A case study from Fijian. To appear in *Language Sciences*.

Kiparsky, Paul (1982). Lexical morphology and phonology. In I. S. Yang (ed.), *Linguistics in the morning calm*. Seoul: Hanshin. 3-91.

Kiparsky, Paul (1985). Some consequences of lexical phonology. *Phonology Yearbook* 2: 85-138.

Kirchner, Robert (2000). Geminate inalterability and lenition. *Language* 76.3: 509-545.

Kirchner, Robert (2001). *An effort-based approach to consonant lenition.* New York: Routledge.

Kirchner, Robert (2004). Consonant lenition. In Hayes et al. (2004). 313-345.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica* 20: 63-67.

Li, Charles N. and Sandra A. Thompson (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language* 4: 185-199.

MacWhinney, Brian (1978). The acquisition of morphophonology. *Monographs of the Society for Research in Child Development*, Serial No. 174, Vol. 43, Nos. 1-2.

Marslen-Wilson, William (1990). Activation, competition, and frequency in lexical access. In Gerry T. M. Altmann (ed.), *Cognitive models of speech processing — psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press.

Mei, Tsu-lin (1977). Tones and tone sandhi in 16^{th} century Mandarin. *Journal of Chinese Linguistics* 5: 237-260.

Menn, Lise and Carol Stoel-Gammon (1995). Phonological development. In Paul Fletcher and Brian MacWhinney (eds.), *The handbook of child language*. Oxford: Blackwell. 335-359.

Mohanan, K. P. (1982). *Lexical phonology*. Ph.D. dissertation, MIT. Distributed by Bloomington: Indiana University Linguistics Club.

Mohanan, K. P. (1993). Fields of attraction in phonology. In John Goldsmith (ed.), *The last phonological rule: Reflections on constraints and derivations*. Chicago & London: The University of Chicago Press. 61-116.

Moore, Corinne B. and Allard Jongman (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America* 102: 1864-1877.

Myers, James (2002). An analogical approach to the Mandarin syllabary. *Chinese Phonology* 11, *Special Issue — Glide, Syllable and Tone*: 163-190.

Newport, Elissa L. and Richard N. Aslin (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48: 127-162.

Oakes, Michael P. (1998). *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.

Ohala, John J. (1981). The listener as a source of sound change. *Papers from the Parasession on Language and Behavior, Chicago Linguistic Society* 18: 178-203.

Ohala, John J. (1992). What's cognitive, what's not, in sound change. In Gunter Kellermann and Michael D. Morrissey (eds.), *Diachrony within synchrony: Language history and cognition*. *Duisburger Arbeiten zur Sprach- und Kulturwissenschaft* 14. Frankfurt: Peter Lang Verlag. 309-355.

Ohala, John J. (1993). Coarticulation and phonology. *Language and Speech* 36: 155-170.

Ohala, John J. (1996). The relation between phonetics and phonology. In William Hardcastle and John Laver (eds.), The *handbook of phonetic sciences*. Oxford: Blackwell. 674-694.

Onishi, Kristine H., Kyle E. Chambers, and Cynthia Fisher (2002). Learning phonotactic constraints from brief auditory experience. *Cognition* 83: B13-B23.

Peng, Shu-Hui (2000). Lexical versus 'phonological' representations of Mandarin sandhi tones. In Michael B. Broe and Janet B. Pierrehumbert (eds.), *Language acquisition and the lexicon: Papers in laboratory phonology V*. Cambridge, UK: Cambridge University Press. 152-167.

Pierrehumbert, Janet (to appear). An unnatural process. To appear in Louis M. Goldstein, Douglas H. Whalen, and Catherine T. Best (eds.) *Papers in Laboratory Phonology* 8. Berlin: Mouton de Gruyter.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Ms., Rutgers University and University of Colorado, Boulder.

Pycha, Anne, Pawel Nowak, Eurie Shin, and Ryan Shosted (2003). Phonological rule-learning and its implications for a theory of vowel harmony. In Gina Garding and Mimu Tsujimura (eds.), *WCCFL 22 Proceedings*. Somerville, MA: Cascadilla Press. 423-435.

Qian, Zeng-Yi, Guang-Qi Zhu (1998). *Jinanhua yindang* (*A record of the Jinan dialect*). Shanghai: Shanghai Jiaoyu Chubanshe (Shanghai Education Press).

Ratcliff, Roger (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin* 114: 510-532.

Reber, Arthur S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6: 855–863.

Reber, Arthur S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General* 118: 219-235.

Redington, Martin and Nick Chater (1996). Transfer in artificial grammar learning: Methodological issues and theoretical implications. *Journal of Experimental Psychology: General* 125: 123–138.

Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport (1996). Statistical learning by 8-month old infants. *Science* 274: 1926-1928.

Seidl, Amanda and Eugene Buckley (2005). On the learning of arbitrary phonological rules. *Language Learning and Development* 1: 289-316.

Shen, Xiaonan Susan and Maocan Lin (1991). A perceptual study of Mandarin tones 2 and 3. *Language and Speech* 34: 145-156.

Shen, Xiaonan Susan, Maocan Lin, and Jingzhu Yan (1993). F0 turning point as an F0 cue to tonal contrast: A case study of Mandarin tones 2 and 3. *Journal of the Acoustical Society of America* 93: 2241-2243.

Shih, Chilin (1997). Mandarin Third Tone Sandhi and prosodic structure. In Wang Jialing and Norvin Smith (eds.), *Studies in Chinese phonology*. Berlin: Mouton de Gruyter. 81-123.

Slobin, Dan Isaac (1985). *The crosslinguistic study of language acquisition, Vol. 2: Theoretical issues*. Hillsdale, NJ: Lawrence Erlbaum Associates. 1157-1256.

Stampe, David (1979). *A dissertation on Natural Phonology*. New York: Garland Publishing.

Steriade, Donca (1999). Phonetics in phonology: The case of laryngeal neutralization. In Matthew K. Gordon (ed.), *UCLA Working Papers in Linguistics, Papers in Phonology* 3: 25-146.

Steriade, Donca (2001). Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume and Keith Johnson (eds.), *The Role of Speech Perception in Phonology*. New York: Academic Press. 219-250.

Steriade, Donca and Jie Zhang (2001). *Context-dependent similarity and the Romanian Semi-rime*. Invited talk at the 37th annual meeting of the Chicago Linguistic Society. Chicago, IL.

Taft, Marcus and Hsuan-Chih Chen (1992). Judging homophony in Chinese: The influence of tones. In Hsuan-Chih Chen and Ovid J. L. Tzeng (eds.), *Language processing in Chinese*. North-Holland: Elsevier Science Publishers. 151-172.

Vitevitch, Michael and Jie Zhang (in progress). *Phonological neighbors in Mandarin Chinese: What role does lexical tone play?* Ms., The University of Kansas.

Wang, H. Samuel (1993). Taiyu biandiao de xinli texing (On the psychological status of Taiwanese tone sandhi). *Tsinghua Xuebao* (*Tsinghua Journal of Chinese Studies*) 23: 175-192.

Wen, Duan-Zheng, Ming Shen (1999). *Taiyuanhua yindang* (*A record of the Taiyuan dialect*). Shanghai: Shanghai Jiaoyu Chubanshe (Shanghai Education Press).

Whalen, Douglas and Yi Xu (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica* 49: 25-47.

Wilson, Colin (2003). Experimental investigation of phonological naturalness. In Gina Garding and Mimu Tsujimura (eds.), *WCCFL 22 Proceedings*. Somerville, MA: Cascadilla Press. 533-546.

Wilson, Colin (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30: 945-982.

Xu, Yi (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61-83.

Yang, Zi-Xiang, He-Tong Guo, and Xiang-Dong Shi (1999). *Tianjinhua yindang* (*A record of the Tianjin dialect*). Shanghai: Shanghai Jiaoyu Chubanshe (Shanghai Education Press).

Ye, Yun and Cynthia M. Connie (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes* 14: 609-630.

Yu, Alan C. L. (2004). Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80.1: 73-97.

Zhang, Jie (2002). *The effects of duration and sonority on contour tone distribution*. New York: Routledge.

Zhang, Jie (2004). The role of contrast-specific and language-specific phonetics in contour tone distribution. In Hayes et al. (2004). 157-190.

Zhang, Ning (1997). The avoidance of the Third Tone Sandhi in Mandarin Chinese. *Journal of East Asian Linguistics* 6: 293-338.

Zue, Victor W. (1976). Some perceptual experiments on the Mandarin tones. *Journal of the Acoustical Society of America* 60, *Supplement* no. 1, S45.

Zuraw, Kie (2000). *Patterned exceptions in phonology*. Ph.D. dissertation, UCLA.

Appendix: Stimuli Lists[10]

I. AO-AO

Base tones	Chinese digram	IPA	Digram freq.	Mutual info. score	Morpheme 1 freq.	Morpheme 2 freq.
3+1	鼓吹	kuː tʂʰwej	45	9	710	543
	锦标	tɕin pjaw	44	10	183	1560
	陕西	ʂan ɕiː	39	9	65	6405
	崭新	tʂan ɕin	38	9	53	7074
	脑筋	ŋaw tɕin	34	10	1494	131
	眼眶	jan kʰwɑŋ	34	9	4324	68
	纺织	fɑŋ tʂɨː	33	11	49	1137
	洒脱	saː tʰwɔ	31	9	357	817
3+2	沈阳	ʂən jɑŋ	45	9.89	147	1526
	谎言	hwɑŋ jɛn	41	8.95	116	3375
	赌博	tuː pwɔ	39	10.30	189	773
	补偿	puː tʂɑŋ	38	10.46	572	222
	礼仪	liː jiː	36	8.79	770	499
	减肥	tɕjɛn fej	35	10.00	620	261
	野蛮	jɛː man	32	10.89	828	96
	饮食	jin ʂɨː	32	9.50	266	786
3+3	展览	tʂan lan	60	8.90	3227	180
	检讨	tɕjen tʰaw	62	9.20	475	1031
	苦恼	kʰuː naw	41	8.62	2080	236
	拇指	muː tʂɨː	34	10.50	35	3162
	甲板	tɕjaː pan	41	8.50	426	1251
	阻挡	tsuː tɑŋ	39	10.48	399	324
	洗碗	ɕiː wan	34	9.15	659	429
	蚂蚁	maː jiː	33	16.69	36	41
3+4	拯救	tʂən tɕju	41	12.15	46	925
	粉碎	fən swej	39	10.32	272	528
	掩护	jɛn xuː	38	8.53	311	1563
	忍耐	ɻən naj	36	9.28	899	304
	巧妙	tɕʰjaw mjaw	35	9.56	415	527
	绑架	pan tɕja	34	11.08	86	863
	饮料	jin ljaw	33	8.82	266	1296
	尺寸	tʂʰɨː tsʰwən	31	10.98	277	262

[10] Digram and monogram frequencies are based on total numbers of 4,159,927 and 4,718,131 respectively.

II. *AO-AO

Base tones	Status of sandhi σ_1	Chinese digram	IPA
3+1	sandhi σ_1 = AO	尺仓	tʂʰɨː tsʰɑŋ
		宇章	ɥyː tʂɑŋ
		写终	ɕjɛː tʂuŋ
		拢叉	luŋ tʂʰaː
	sandhi σ_1 = AG	榜中	pɑŋ tʂuŋ
		拇村	muː tsʰwən
		井披	tɕiᵊŋ pʰiː
		减苍	tɕjan tsɑŋ
3+2	sandhi σ_1 = AO	尺玩	tʂʰɨː wan
		宇零	ɥyː liᵊŋ
		写拳	ɕjɛː tɕʰɥɛn
		拢宅	luŋ tʂai
	sandhi σ_1 = AG	榜连	pɑŋ ljɛn
		拇挪	muː nwɔ
		井菩	tɕiəŋ pʰuː
		减和	tɕjan xɤː
3+3	sandhi σ_1 = AO	尺洒	tʂʰɨː saː
		宇览	ɥyː lan
		写五	ɕjɛː wuː
		拢法	luŋ faː
	sandhi σ_1 = AG	榜洒	pɑŋ saː
		拇饮	muː jin
		井免	tɕiəŋ mjɛn
		减也	tɕjan jɛː
3+4	sandhi σ_1 = AO	尺葬	tʂʰɨː tsɑŋ
		宇耀	ɥyː jau
		写逆	ɕjɛː niː
		拢料	luŋ ljau
	sandhi σ_1 = AG	榜报	pɑŋ pau
		拇葬	muː tsɑŋ
		井妙	tɕiəŋ mjau
		减会	tɕjan xwei

III. AO-AG

Base tones	Status of sandhi σ_1	Chinese digram	IPA
3+1	sandhi σ_1 = AO	闯 shun	tʂʰwɑŋ ʂwən
		火 mu	xwoː muː
		领 lan	liᵊŋ lan
		巧 re	tɕʰjau ɻɤː
	sandhi σ_1 = AG	本 mai	pən mai
		苦 liang	kʰuː ljɑŋ
		款 lang	kʰwan lɑŋ
		损 rao	swən ɻau
3+2	sandhi σ_1 = AO	闯 te	tʂʰwɑŋ tʰɤː
		火 ka	xwoː kʰaː
		领 pie	liᵊŋ pʰjɛː
		巧 jiu	tɕʰjau tɕjuː
	sandhi σ_1 = AG	本 mie	pən mjɛː
		苦 geng	kʰuː kəŋ
		款 dui	kʰwan twei
		损 duan	swən twan
3+3	sandhi σ_1 = AO	闯 zeng	tʂʰwɑŋ tsəŋ
		火 suan	xwoː swan
		领 huai	liᵊŋ xwai
		巧 hang	tɕʰjau xɑŋ
	sandhi σ_1 = AG	本 xun	pən ɕɥyn
		苦 heng	kʰuː xəŋ
		款 pan	kʰwan pʰan
		损 cuo	swən tswɔː
3+4	sandhi σ_1 = AO	闯 zhua	tʂʰwɑŋ tʂwaː
		火 sen	xwoː sən
		领 dei	liᵊŋ tei
		巧 shua	tɕʰjau ʂwaː
	sandhi σ_1 = AG	本 dei	pən tei
		苦 keng	kʰu kʰəŋ
		款 mang	kʰwan mɑŋ
		损 diu	swən tjəu

IV. AG-AO

Base tones	Status of sandhi σ_1	Chinese digram	IPA
3+1	sandhi σ_1 = AO	ping 八	pʰiᵊŋ paː
		pan 昭	pʰan tʂau
		xia 凶	ɕjaː ɕjuŋ
		cang 黑	tsʰɑŋ xei
	sandhi σ_1 = AG	zhui 咪	tʂwej miː
		chua 单	tʂʰwaː tan
		run 邱	ɻwən tɕʰjəu
		shuan 君	ʂwan ɕɥyn
3+2	sandhi σ_1 = AO	ping 豪	pʰiᵊŋ xau
		pan 胡	pʰan xuː
		xia 林	ɕjaː lin
		cang 原	tsʰɑŋ ɥyɛn
	sandhi σ_1 = AG	zhui 伦	tʂwej lwən
		chua 林	tʂʰwaː lin
		run 盘	ɻwən pʰan
		shuan 葵	ʂwan kʰwei
3+3	sandhi σ_1 = AO	ping 马	pʰiᵊŋ maː
		pan 海	pʰan xai
		xia 哪	ɕjaː naː
		cang 尺	tsʰɑŋ tʂʰɨː
	sandhi σ_1 = AG	zhui 法	tʂwej faː
		chua 轨	tʂʰwaː kwei
		run 起	ɻwən tɕʰiː
		shuan 老	ʂwan lau
3+4	sandhi σ_1 = AO	ping 套	pʰiᵊŋ tʰau
		pan 玉	pʰan ɥyː
		xia 类	ɕjaː lei
		cang 率	tsʰɑŋ lyː
	sandhi σ_1 = AG	zhui 半	tʂwej pan
		chua 路	tʂʰwaː luː
		run 费	ɻwən fei
		shuan 怒	ʂwan nuː

V. AG-AG

Base tones	Status of sandhi σ_1	Chinese digram	IPA
3+1	sandhi σ_1 = AO	ping shun	pʰiᵊŋ ʂwən
		pan mai	pʰan mai
		xia mei	ɕjaː mei
		cang re	tsʰɑŋ ɻɤː
	sandhi σ_1 = AG	zhui mai	tʂwej mai
		chua liang	tʂʰwaː ljɑŋ
		run lang	ɻwən lɑŋ
		shuan kuo	ʂwan kʰwɔː
3+2	sandhi σ_1 = AO	ping te	pʰiᵊŋ tɤː
		pan ka	pʰan kʰaː
		xia kong	ɕjaː kʰuŋ
		cang mie	tsʰɑŋ mjɛː
	sandhi σ_1 = AG	zhui mie	tʂwej mjɛː
		chua geng	tʂʰwaː gəŋ
		run dui	ɻwən twei
		shuan ta	ʂwan tʰaː
3+3	sandhi σ_1 = AO	ping zeng	pʰiᵊŋ tsəŋ
		pan seng	pʰan səŋ
		xia lue	ɕjaː lɥɛː
		cang xia	tsʰɑŋ ɕjaː
	sandhi σ_1 = AG	zhui kuang	tʂwej kʰwɑŋ
		chua heng	tʂʰwaː xəŋ
		run pan	ɻwən pʰan
		shuan sai	ʂwan sai
3+4	sandhi σ_1 = AO	ping zhua	pʰiᵊŋ tʂwaː
		pan sen	pʰan sən
		xia dei	ɕjaː tei
		cang shua	tsʰɑŋ ʂwaː
	sandhi σ_1 = AG	zhui dei	tʂwej tei
		chua keng	tʂʰwaː kʰəŋ
		run mang	ɻwən mɑŋ
		shuan sen	ʂwan sən