Research Paper I:
Jinhua Dialect Tone Sandhi

Lucien Carroll

Main Reader: Sharon Rose
Ancillary Readers:
Amalia Arvaniti and Roger Levy

Abstract

This study documents speaker variation and lexical variation in the phonological tone change patterns of Jinhua Wu Chinese, one dialect in a region known for particularly complicated tone sandhi systems and regional variation. Previous reports of Jinhua tone sandhi have agreed that there is some lexical variation, but they have disagreed considerably about the contour shapes as well as about mergers between tone categories. It was not clear whether these were differences in methodology, stimuli selection, speaker idiolect, or systematic regional or generational differences. Previous reports acknowledged speaker differences but made no attempt to systematically document it or quantify it. An acoustic analysis was performed on the pitch tracks of disyllabic Jinhua Wu words. The list of words was recorded spoken by 15 native speakers from two locations and two generations. The analysis proposes new methods of contour parameterization and tone categorization. The results indicate that there are systematic regional and generational differences in Jinhua sandhi, as well as lexical variation. Also, the space of tone contours is highly constrained, apparently by a combination of coarticulation and an obligatory contour principle (OCP) effect. The generational differences show active tonal evolution consistent with theories of sandhi arising from tonal coarticulation.

Acknowledgements

This study could not have been completed without the support and assistance of numerous people.

First, I need to thank the people of Jinhua: the native speakers who agreed to be recorded, as well as my friends who made me feel welcome, helped me prepare the stimuli, and connected me with the study participants. I am especially grateful to Emily Cen and Betty Hu for their assistance.

I would like to thank the linguistics community at UC San Diego. Though the idea of this project was planted before coming to UCSD, and the data was gathered while on a leave of absence, this study was cultivated and enabled by classes and conversations here. Discussions and readings with the San Diego Phonology Interest Group, the first year phonetics/phonology series, and a directed study with Sharon Rose have been especially formative. Research assistant Eric Hendrickson also helped me evaluate the annotation procedure. I have received great advice and encouragement from Rebecca Colavin, Alex Delgiudice and Leslie Lee, among many others. And finally, this paper's evaluation committee (Sharon Rose, Amalia Arvaniti and Roger Levy) have given me valuable feedback on the study methodology as well as its presentation.

I am also indebted to Matthew Chen, Sun Yizhi, Phil Rose and Ryan Shosted, who have lent their time to answering my questions. I have also benefited from Matthew Chen's legacy in the UCSD library collection, and my interest in Jinhua sandhi began when I found a copy of *Tone Sandhi: Patterns across Chinese Dialects* (Chen, 2000) in a Jinhua bookstore.

Finally, I thank Angie Kim, who helped me like Jinhua more, gave me reason to return, and enabled me to carry out some of the the most difficult aspects of this project. Without her support I would have given up several times over.

Any errors remain my own.

Contents

1 Introduction **1**

2 Background **4**

2.1 Chinese tone categories and contours 4

2.2 Tone sandhi 7

2.2.1 Sandhi is not coarticulation 7

2.2.2 Sandhi is not morphological 9

2.2.3 Sandhi is not intonation 11

2.3 Examples of sandhi systems 11

2.3.1 Standard Mandarin 11

2.3.2 Xiamen/Taiwanese 13

3 Jinhua Sandhi **15**

3.1 The complexity of Southern Wu sandhi 15

3.2 Previous reports 16

3.2.1 Qian Nairong 16

3.2.2 Cao Zhiyun 17

3.2.3 You and Yang 19

3.3 The documentation problem 22

4 Procedure and Results **24**

4.1 Elicitation 24

4.1.1 Sandhi domains 24

4.1.2 Stimuli 25

4.1.3 Speakers 26

4.1.4 Presentation 26

4.1.5 Recording 27

4.2 Annotation 27

4.2.1 Syllable structure 28

4.2.2 The tone domain 29

4.2.3 Tone domain onsets 30

4.2.4 Tone domain offsets 30

4.3 Pitch contours 31

4.3.1 Previous work 31

4.3.2 Extraction of pitch tracks 34

4.3.3 Evaluation of parameterizations . . . 35
4.3.4 Frequency scale and transform . . . 36
4.3.5 Summarized representations . . . 39
4.4 Contour systems . . . 40
4.4.1 Speaker groups . . . 40
4.4.2 Contour categories . . . 41
4.4.3 Group contour systems . . . 46

5 Discussion **52**
5.1 Dimensionality of contours . . . 52
5.2 Differences among words . . . 53
5.3 Differences among speakers . . . 54
5.4 Differences among studies . . . 58
5.5 Conclusions . . . 60

References **62**

A Stimuli **66**

1 Introduction

This paper describes a study which documents generational, geographic, and lexical variation in the tone sandhi of Jinhua Wu, a Chinese dialect in a region known for extraordinarily complicated sandhi systems. Previous reports of Jinhua sandhi disagree considerably, and it was not clear whether these differences were due to differences in methodology, stimuli selection, speaker idiolect, or systematic regional or generational differences. This study replicates and extends those studies, confirming the claims of complicated sandhi systems and showing that each of these sources of study differences must play a role in the disagreements among the previous reports.

The case of Jinhua sandhi poses a challenge to the traditional methods of tone documentation, which depend on impressionistic transcriptions and metalinguistic awareness of the language's phonological categories. The linguist's own phonological categories are bound to influence the representation, and it is often unclear to what extent tone transcriptions reflect coarticulation effects and conscious or unconscious efforts to remove coarticulation effects. These problems are by no means unique to tonal phonology (see e.g., Port & Leary, 2005), but they are particularly problematic here. In Jinhua and surrounding areas, the abundance of lexical and sociolinguistic variation makes it quite difficult to know how many tone categories are involved or what constitutes a meaningful transcription distinction. The methodology used in this study builds on previous acoustic tonetics research (e.g., Rose, 1987; Zhu, 1999) to partially overcome these issues by using acoustic data rather than transcriptions, basing representation decisions on objective fitness measures, and allowing Bayesian clustering to determine the number of categories.

The city of Jinhua is the administrative seat of a prefecture-level municipality in the middle of Zhejiang Province, China. The dialects native to Zhejiang Province and a few surrounding regions, including Shanghai, constitute Wu Chinese (see Fig. 1). The dialects of China are surprisingly divergent in mutual intelligibility, particularly so among the southern dialects (Tang, 2009). Syntactic and phonotactic structures are broadly similar from dialect to dialect, but lexical and phonological differences make a dialect of one region completely unintelligible to a speaker from another region. Even dialects which are genetically and geographically close might not be mutually intelligible. Within Wu, there is more regional variation in the southern inland area than elsewhere, such that the dialects of Jinhua, Tangxi, and Yongkang are mutually unintelligible, even though less than 50 km separate one from the next, and they are all members of the same dialect group (B. Xu, 1999). Because dialects are so different from one another, each dialect provides typologically valuable linguistic data. The northern Wu dialects have been fairly well studied, and a fair amount of work has been



Figure 1: Map of region where Wu is spoken. Wu is spoken throughout most of Zhejiang province as well as in Shanghai, the southern end of Jiangsu Province (e.g. Wuxi), and the eastern corner of Anhui Province (Xuanzhou).

done in the Wenzhou area. However, the dialects in between these two groups and inland have been much less studied.

The data from Jinhua needs to be gathered soon, because in 50 years there might be no fluent speakers left, because children are no longer learning it. A study like this one, which compares the tone systems between two generations, will not be possible in even less time. The dialects of the larger cities like Shanghai, Ningbo and Wenzhou enjoy a high degree of local prestige, and many middle-aged people in these larger cities only speak Wu. Those dialects remain the preferred local trade language, and children are still learning them. In Jinhua, however, the local dialect is less prestigious, so though the older generations still use the language for social purposes, all but the very oldest also speak Mandarin, and very few people born after 1990 speak Wu. None of the Wu dialects are used in print media, but many of them, including Jinhua dialect, are now used in television. Just in the last

few years, the local television station has started producing a sitcom and a news show in Jinhua dialect with Mandarin subtitles. Cultural influences like this might raise the prestige of Jinhua dialect, but according to the current trend of language learning in the city, the dialect is threatened.

Just as the complicated and diverse tone systems of other Wu, Hakka, and Min dialects provide rich ground for understanding tonal phonology (e.g. Chen, 2000; Zhang, 2007), Jinhua sandhi provides a valuable data point for typological studies. But because Jinhua sandhi exhibits non-negligible lexical, geographic and generational variation, in addition to being unusually complicated, even the goal of documentation requires new analytic methods. This paper first explains the problem that Jinhua sandhi holds for documentation and linguistic analysis and then describes the procedure by which the acoustic data was acquired and analyzed. Finally, the consequences and outstanding issues are discussed.

2 Background

To understand Jinhua sandhi, it is first necessary to understand Chinese isolation tone categories and sandhi more generally. And because Chinese dialectology is heavily based on relationships to historical forms, studies of Jinhua sandhi refer to historical tone categories as well as contemporary tone categories.

2.1 Chinese tone categories and contours

In many tone systems, in Asia if not more generally, lexical tone arose historically when certain onsets and codas underwent contrast reduction, and the functional load shifted away from the onsets and codas onto related systematic differences in pitch, a process called tonogenesis (Matisoff, 1973). In many of these languages, including Vietnamese and Chinese, there is a register split based on the voicing of the onset, where the historically voiced onsets gave rise to a set of low tones comprising a 'low register' and the historically voiceless onsets gave rise to a set of high tones comprising a 'high register'. Each register is also split into three or four tones based on the manner and place of articulation of historical codas. In the process of tonogenesis, the original segmental contrast may be reduced or lost all together.

In Middle Chinese, there were four coda-based tone classes referred to as *ping*, *shang*, *qu* and *ru*. The *ping* class consisted of open syllables and nasal-coda syllables. The *shang* and *qu* classes are commonly tentatively reconstructed in Old Chinese (OC) as respectively ending with /ʔ/ and /s/, but these codas were lost by Middle Chinese (MC). And the *ru* class had voiceless stop codas (Baxter, 1992, pp. 303–324). The names are mnemonic, in that *ping* (平 'level') is a member of *ping* class, *shang* (上 'rise') is a member of *shang* class, and likewise for *qu* (去 'depart') if of *qu* class and *ru* (入 'enter') of *ru* class. The names apparently also describe the contour shape when the names were given c. 500 AD, though there is disagreement about the appropriate interpretation of the descriptions (Ting, 1996). Similarly, the two registers were mnemonically called *yin* (阴 *ʔim 'dark/female')[1] and *yang* (阳 *jaŋ 'light/male') for the voiceless onset (high) and voiced onset (low) registers, respectively (see Table 1). In western scholarship, the four classes have typically been referred to by the Roman numerals *I*, *II*, *III* and *IV*, and the registers as *A* and *B*. Because Roman numerals are easily confusable with other symbols, this paper uses the circled numbers ①, ②, ③ and ④ for the tone classes, and for the sake of keeping a more mnemonic notation, *H* and *L* are used to refer to the high and low registers, though the original correlation between register and pitch height is not always maintained, particularly in sandhi.

[1] Reconstructed forms in this paper are according to Baxter (n.d.), unless otherwise noted

Table 1: Tone categories in Middle Chinese

	< OC *[+son]# ① (ping)	< OC *ʔ# ② (shang)	< OC *s# ③ (qu)	MC *{ptk}# ④ (ru)
$[-vc]$ onset (yin)	$①_{\mathrm{H}}$	$②_{\mathrm{H}}$	$③_{\mathrm{H}}$	$④_{\mathrm{H}}$
$[+vc]$ onset (yang)	$①_{\mathrm{L}}$	$②_{\mathrm{L}}$	$③_{\mathrm{L}}$	$④_{\mathrm{L}}$

Asian tone contours are conventionally represented with Chao tone numbers: a string of 1 to 3 numerals between 1 and 5, where 1 represents the lowest pitch, and 5 the highest (Chao, 1930). A contour consisting of a single numeral is short (for a reduced tone or with a stop coda), and contours of two and three numerals are typically medium and long, though some are short. To distinguish tone categories from tone contours, contours are written as superscripts in this paper, and categories are written as subscripts or in-line. For example, Beijing Mandarin (Table 2) has four tone categories, plus a reduced or neutral tone. The first tone (T_1^{55}) corresponds to MC $①_{\mathrm{H}}$ and has a high pitch and a long level contour. The second tone (T_2^{35}) corresponds to MC $①_{\mathrm{L}}$, and has a contour that starts in the middle of the pitch range and rises. The third tone (T_3^{214}), whose contour prototypically starts mid-low and falls further before rising quickly, includes the syllables of MC $②_{\mathrm{H}}$ as well as the sonorant-onset syllables from $②_{\mathrm{L}}$, while the obstruent-onset syllables from $②_{\mathrm{L}}$ joined with class ③ syllables to become Mandarin fourth tone (T_4^{51}), whose contour starts high and falls very quickly. (Syllables in MC class ④ were scattered among Mandarin tone categories in a complicated

Table 2: Beijing tone categories compared to Middle Chinese

	MC ①	MC ②	MC ③
MC $[-vc]$ onset	T_1^{55}	T_3^{214}	T_4^{51}
MC $[+vc,+son]$ onset	T_2^{35}	T_3^{214}	T_4^{51}
MC $[+vc,-son]$ onset	T_2^{35}	T_4^{51}	T_4^{51}

pattern, and are left off.) Note that the onset voicing in the table is that of Middle Chinese onsets, as Mandarin has lost the obstruent voicing contrast, and the contrasts between MC$①_{\mathrm{H}}$ and MC$①_{\mathrm{L}}$ are now only maintained in the tone categories. Finally, the Mandarin neutral tone (a by-product of grammaticization and morphological processes) is always short, but its pitch height depends on its environment, and so it may appear as T^1, T^2, T^3 or T^4, depending on its context.

Compared to Mandarin, Southern Wu dialects maintain a tone inventory more like that of Middle Chinese, with typically 7 or 8 tone categories, including a clear register split (Cao, 2002). According to Chao's survey (Chao, 1928) which included one Jinhua speaker, Jinhua had an 8 tone system at that time, but in Qian's survey (Qian, 1992), conducted in 1985 with 12 Jinhua speakers born between 1912 and 1968, the younger speakers had no distinct ②$_{\text{L}}$ tone and the oldest speakers had almost lost the category as well. The tone categories and contours of the middle generation are shown in Table 3. The syllables of MC ②$_{\text{L}}$ had mostly merged with ②$_{\text{H}}$, though some (mostly voiced obstruent onset syllables) converted to ③$_{\text{L}}$, as in Mandarin. Cao (2002), differing slightly from Qian, reports that sonorant onset syllables of ②$_{\text{L}}$ have merged with ②$_{\text{H}}$, and obstruent onset syllables of ②$_{\text{L}}$ have merged with ②$_{\text{H}}$ in informal speech but converted to ③$_{\text{L}}$ in literary use. In any case, that tone category is now lost, but Jinhua still otherwise retains the register split, and H register tones are higher in pitch than their corresponding L register tones. In Qian's survey (Table 3), the class ① tones are dipping contours, the class ② tone is slightly falling, the class ③ tones are rising, and the class ④ tones are level with glottal stop codas. The contours that Cao reports are slightly different from Qian's, and they do not have one shape for each tone class like those from Qian do. Like Mandarin, Jinhua does not have an obstruent voicing

Table 3: Tone categories of Jinhua (of generation born ca. 1930) compared to Middle Chinese, according to Qian (1992)

	MC ①	MC ②	MC ③	lit. MC ④	cas. MC ④
MC $[-vc]$ onset	①$_{\text{H}}^{435}$	②$_{\text{H}}^{544}$	③$_{\text{H}}^{45}$	③$_{\text{H}}^{45}$	④$_{\text{H}}^{44\text{?}}$
MC $[+vc, +son]$ onset	①$_{\text{L}}^{213/324}$	②$_{\text{H}}^{544}$	③$_{\text{L}}^{24}$	③$_{\text{L}}^{24}$	④$_{\text{L}}^{22\text{?}}$
MC $[+vc, -son]$ onset	①$_{\text{L}}^{213/324}$	②$_{\text{H}}^{544}$ ③$_{\text{L}}^{24}$	③$_{\text{L}}^{24}$	③$_{\text{L}}^{24}$	④$_{\text{L}}^{22\text{?}}$

contrast, though many neighboring dialects do. Cao (2002) indicates that historically voiced obstruents are devoiced in urban Jinhua and are inconsistently voiced in the surrounding rural areas. He still transcribes most historically voiced obstruents as voiced, but this paper assumes voiceless obstruents except when citing data from the literature. According to Qian, literary vocabulary from MC ④ have merged with the ③ class in Jinhua, while casual vocabulary from MC ④ remains in distinct ④ tone categories, pronounced with a short vowel and a glottal stop coda. Cao also reports that some MC ④ syllables have merged with ③ while others have remained in a distinct ④ class, but according to him the syllables that merged with ③ are only those that ended in *-ap or *-at in Middle Chinese. These two factors of MC tone class and vocabulary register remain important for distinguishing types of Jinhua sandhi patterns.

2.2 Tone sandhi

Tone sandhi can be defined as context-dependent phonological changes to lexically specified tone. It is useful to distinguish three common patterns of sandhi in Chinese languages. In tone spreading sandhi, the tone change can be analyzed as the extension of one syllable's underlying tone over that syllable and neighboring syllables. For example, in Shanghai, the typical sandhi tones are determined by the isolation tone of the first syllable (Duanmu, 1999).

(1) a. *se*$^{\mathbf{52}}$ *pe*^{52} → *se*$^{\mathbf{55}}$ *pe*$^{\mathbf{21}}$
three cups

b. *si*$^{\mathbf{34}}$ *pe*^{52} → *si*$^{\mathbf{33}}$ *pe*$^{\mathbf{44}}$
four cups

In tone reduction sandhi, the sandhi tone is a simplification of the underlying contour. For example, in certain contexts Standard Mandarin T_3^{214} reduces to T_3^{21} (Chao, 1948).

(2) a. *ʂwej*$^{\mathbf{214}}$ *kʰu*^{51} → *ʂwej*$^{\mathbf{21}}$ *kʰu*^{51}
water storage reservoir

b. *ljaŋ*$^{\mathbf{214}}$ *pej*^{55} → *ljaŋ*$^{\mathbf{21}}$ *pej*^{55}
two cups

In tone replacement sandhi, the sandhi tone can be an arbitrary contour, featurally unrelated to the underlying tonal contour. When followed by another T_3, Standard Mandarin T_3^{214} converts to T_2^{35}.

(3) a. *ʂwej*$^{\mathbf{214}}$ *kwo*^{214} → *ʂwej*$^{\mathbf{21}}$ *kwo*^{214}
water fruit fruit

b. *ljaŋ*$^{\mathbf{214}}$ *wan*^{214} → *ljaŋ*$^{\mathbf{35}}$ *wan*^{214}
two bowls

The boundaries of the concept of sandhi are unclear in a couple ways, but sandhi should be distinguished from several other kinds of pitch change.

2.2.1 Sandhi is not coarticulation

Tone sandhi and tonal coarticulation are both "tonal variations triggered by tonal context" (Shen, 1992). Shen attempts to distinguish them from each other by three criteria:

- Tone sandhi is an effect of language-specific phonology, whereas tonal coarticulation is an effect of biomechanics.

- Tone coarticulation was said to only involve assimilation, while tone sandhi may involve assimilation or dissimilation
- Tone sandhi makes a categorical change in tone, whereas tonal identity is essentially unchanged in coarticulation.

However, in his discussion of tonal coarticulation, mostly reviewing Shih's work (Shih, 1988), Chen (2000) points out that the line between sandhi and coarticulation is not so clear. Each of Shen's criteria sometimes fails, even just within Taiwan Mandarin. First, a certain effect ($T_2^{35}T_3^{21} \rightarrow T_2^{53}T_3^{21}$), which is considered coarticulation because it is only detectable instrumentally, is sensitive to syntax, applying across word boundaries but not word internally. Second, there is a slight pitch raising effect on T_2^{35} and T_4^{51} before the low T_3^{214}, i.e. coarticulatory dissimilation from the following tone onset. And finally, though speakers cannot hear the difference between an underlying T_2^{35} and the result of Mandarin's textbook third-tone sandhi ($T_3^{214} \rightarrow T^{35}/_\ T_3$), the one that is underlyingly a third tone has a phonetically lower pitch than the normal second tone, and they are thus not fully merged tonemes.

The criterion of language-specific effects versus (language-independent) biomechanical effects is not reliable. Language-specific effects and even word-specific phonetic effects (Pierrehumbert, 2002) are well-known, and the effects of biomechanics can be seen in both phonetically 'natural' sandhi (like rightward spreading) and phonetically 'unnatural' sandhi (like tone replacement) (Hyman, 2007). The typological abundance of rightward spreading can easily be seen as a phonologization of biomechanically grounded assimilatory coarticulation (J. Myers & Tsay, 2003). And both 'natural' and 'unnatural' sandhi often involve tonal simplification on light syllables and non-final syllables (Zhang, 2002), since it is easier to produce multiple pitch targets on longer, more sonorant syllables (Gordon, 2001).

Nor can we distinguish sandhi from coarticulation on the basis of dissimilation, at least in lab speech. As Shih's work showed for Taiwan Mandarin, carryover coarticulation effects are assimilatory, whereas anticipatory coarticulation can be assimilatory or dissimilatory. Strong carryover assimilation effects and inconsistent or weak anticipatory dissimilation effects have also been observed in the lab for Beijing Mandarin (Y. Xu, 1997), Cantonese (Gu & Lee, 2007), Taiwanese (Peng, 1997) and Thai (Gandour & Potisuk, 1994). In casual speech however, anticipatory coarticulation effects are overwhelmingly assimilatory (Shih, Kochanski, & Yoon, 2007).

The criterion of tonal identity may be of some value if we can determine in what sense tonal identity is maintained in coarticulation and changed in sandhi. The magnitude of the pitch alternations that are considered sandhi does make them more perceptible to native

speakers and thus more accessible to routine category learning — as well as to metalinguistic awareness — but there is no clear dividing line. Chen (2000) suggests that the difference between sandhi and coarticulation may simply be two ends of a spectrum, distinguished primarily by what linguists can hear.

2.2.2 Sandhi is not morphological

There are also certain kinds of morphological tone change that we want to exclude from the definition of sandhi, and here the boundary is not much clearer. The intuition is that phenomena known as floating tonal morphemes do not count as sandhi, but sandhi is well known to be influenced by the morphosyntactic structure of tone sandhi domains, and the distinction between tonal morphemes and morphosyntactic contexts is somewhat dependent on the analysis.

As an example of how sandhi is influenced by morphosyntactic structure, Pingyao Mandarin (Chen, 1990; Zhang, 1999) has one pattern 'Type A' (Ex. 4) that appears in sandhi domains composed of subject+predicate or verb+object, and 'Type B' (Ex. 5) that appears in sandhi domains composed of modifier+noun and many other constructions[2].

(4) a. *tsjaw*^{13} *hwa*^{13} → *tsjaw*^{13} *hwa*^{13}
water(V) flower

b. *maj*$^{\mathbf{35}}$ *hwa*^{13} → *maj*$^{\mathbf{13}}$ *hwa*^{13}
sell flower

c. *kʌŋ*$^{\mathbf{13}}$ *ti*^{35} → *kʌŋ*$^{\mathbf{31}}$ *ti*^{35}
plough(V) ground

(5) a. *ts*^{h}*oŋ*$^{\mathbf{13}}$ *hwa*$^{\mathbf{13}}$ →*ts*^{h}*oŋ*$^{\mathbf{31}}$ *hwa*$^{\mathbf{35}}$
onion flower

b. *kwej*^{35} *hwa*$^{\mathbf{13}}$ → *kwej*^{35} *hwa*$^{\mathbf{53}}$
cassia flower

c. *tsjaŋ*^{13} *tow*$^{\mathbf{35}}$ → *tsjaŋ*^{13} *tow*$^{\mathbf{13}}$
cowpea

In verb+object domains such as in (Ex. 4), the changes are on the first syllable, whereas in domains like (Ex. 5), either tone or both tones can change, and the changes seen in the two paradigms are not the same. Similar splits between sandhi paradigms are known in many Wu dialects, including Shanghai (Zhang, 2007), Huzhou's Tangxi (Kennedy, 1953),

[2]Data is from Chen (1990), following him in using standard Mandarin segments, but using IPA and Chao tone numbers rather than pinyin and [L, M, H] to be consistent with the rest of the notation here.

and Shaoxing (Yue-Hashimoto, 1987). Cao (2002) reports three distinct domains in Jinhua (discussed later) and seven for Tangxi, a town about 25 km west of Jinhua.

But it also sometimes happens that a language will have purely morphological tone change, in the form of tonal morphemes, which in the simplest case supplete the original tone of the syllable they associate to. For example, in Cantonese a meaning of smallness or disdain is sometimes conveyed by changing the tone of a word from its underlying tone to a high rising tone (T^{35}) or a high level tone (T^{55}) (Wong, 1982). This is typically analyzed as a 'floating' tone that replaces the original tone.

Another place the Cantonese tonal morpheme appears is in verb reduplication. In one form of verb reduplication, the first copy of the verb takes high rising tone.

(6) a. *si*$^{\mathbf{35}}$ *si*^{33}
try+DIM try
'to give a try'

b. *haːŋ*$^{\mathbf{35}}$ *haːŋ*^{21}
walk+DIM walk
'to take a walk'

c. *tʊk*$^{\mathbf{35}}$ *tʊk*^{2}
read+DIM read
'to read casually'

But in other Chinese languages, tone change in this same verb reduplication construction is more simply analyzed as a kind of sandhi. In Tangxi, tone change in reduplicated verbs is similar but not identical to the prototypical sandhi of noun compounds (Cao, 2002, p. 120). With tones ②$_{\text{H}}$, ②$_{\text{L}}$ and ④$_{\text{L}}$, the second syllable changes to ③$_{\text{H}}$ and the first syllable tone simplifies (7a,b). With the other tone categories, the first syllable retains its citation tone, and the second syllable is reduced (7c).

(7) a. *zɣ*$^{\mathbf{113}}{}_{②\mathbf{L}}$ *zɣ*$^{\mathbf{113}}{}_{②\mathbf{L}}$ → *zɣ*$^{\mathbf{11}}{}_{①\mathbf{L}}$ *sɣ*$^{\mathbf{52}}{}_{③\mathbf{H}}$
sit sit
'to sit a while'

b. *kuo*$^{\mathbf{535}}{}_{②\mathbf{H}}$ *kuo*$^{\mathbf{535}}{}_{②\mathbf{H}}$ → *kuo*$^{\mathbf{52}}{}_{③\mathbf{H}}$ *kuo*$^{\mathbf{52}}{}_{③\mathbf{H}}$
talk talk
'to talk a bit'

c. *bo*$^{11}{}_{①L}$ *bo*$^{\mathbf{11}}{}_{①\mathbf{L}}$ → *bo*$^{11}{}_{①L}$ *bo*$^{\mathbf{3}}$
climb climb
'to climb a bit'

With Cantonese diminutives and verb reduplication, the altered tone is not sandhi, since it is most simply analyzed as suppletion by a tonal morpheme. Tone change in Tangxi reduplicated verbs could be analyzed as the interaction of an abstract tonal morpheme with the underlying tones, but considering the variety of morphosyntactically defined sandhi domains known in Tangxi, it is simpler to consider reduplicated verbs a distinct sandhi domain.

2.2.3 Sandhi is not intonation

Finally, we need to distinguish effects of lexical context from effects of phrasal accent and intonation. One such effect is that of declination (Ladd, 1984; Shih, 1997). This is the cross-linguistic tendency for average F0 and pitch range to decrease throughout the length of an intonation phrase. It is attributed to decreasing air pressure below the vocal folds, but it may also be used to signal discourse structure, since the rate of declination depends on the length of the intonation phrase and the discourse relationships to the intonation phrases before and after (Ladd, 1988).

Another gradient, long range effect is that of focus. Focus-marking is different from language to language, but often involves manipulation of F0. For example, Beijing Mandarin marks constrastive focus primarily by expanding the pitch range over the focused syllable and compressing the pitch range in later syllables (Y. Xu, 1999). In Cantonese, the pitch range begins to expand prior to the focus, reaches a maximum on the focused region, and then gradually decreases to the end of the intonation phrase (Gu & Lee, 2007). In New Chongming (Chen, 2000), sandhi and focus are inextricable because focus results in tone deletion in the post-focus region, which then permits sandhi. Except for cases like New Chongming, we want to exclude the effect of focus from sandhi phenomena.

2.3 Examples of sandhi systems

Before turning to the Jinhua sandhi system, a couple of simpler systems may help orient the reader to Chinese sandhi generally. The two most studied systems are those of Standard Mandarin and Xiamen.

2.3.1 Standard Mandarin

In Standard Mandarin, the sandhi observable in two-syllable sandhi domains consists primarily of the previously mentioned 'third-tone sandhi' (Chao, 1948). It is so called because it involves changes to T_3 only. The sandhi tone contours for all possible two-syllable combinations are displayed in Table 4, again ignoring ④ syllables, as in Table 2. The first syllable

Table 4: Standard Mandarin Sandhi Table. The first syllable tone categories are along the left margin and the second syllable tone categories are along the top margin. The tones are marked according to both Middle Chinese (MC) and contemporary Beijing (BJ) categories. The body of the table contains the two-syllable tone contours, and the 'row/col trend' is a generalization of the first syllable contours in the rows and of the second syllable contours in the columns. The tone changes are highlighted.

					2nd Syllable						
				MC class	①		②			③	
				MC reg.	**H**	**L**	**H**	**L (son)**	**L (obs)**	**H**	**L**
				BJ cat.	**T_1**	**T_2**	**T_3**		**T_4**		
	MC class	**MC reg.**	**BJ cat.**	**row/col trend**	**55**	**35**	**214**	**214**	**51**	**51**	**51**
1st Syllable	①	**H**	**T_1**	**55**	55.55	55.35	55.214	55.214	55.51	55.51	55.51
		L	**T_2**	**35**	35.55	35.35	35.214	35.214	35.51	35.51	35.51
	②	**H**	**T_3**	**21**	21.55	21.35	35.214	35.214	21.51	21.51	21.51
		L (son)		**21**	21.55	21.35	35.214	35.214	21.51	21.51	21.51
		L (obs)	**T_4**	**51**	51.55	51.35	51.214	51.214	51.51	51.51	51.51
	③	**H**		**51**	51.55	51.35	51.214	51.214	51.51	51.51	51.51
		L		**51**	51.55	51.35	51.214	51.214	51.51	51.51	51.51

tones with their Middle Chinese and modern categories are on the left side, and the second syllable categories across the top. Generalizations of first syllable contours are written beside the first syllable categories, and generalizations about second syllable contours are written below the second syllable categories.

Note that isolation tone contours are equivalent to the second syllable in the sandhi contours, and the isolation tone categories are sufficient to predict the sandhi contours. The MC tone categories are shown only for comparison with the data from Xiamen and Jinhua. There are two kinds of tone change for T_3: the first syllable generalization T_3^{21} differs from the second syllable generalization T_3^{214}, and when both first and second syllables are T_3, the first syllable takes a third contour T^{35}, that is, T_2. The sandhi can be succinctly summarized

by the following rules.

(8)

	Rule	Example	
T_3	$\rightarrow T^{214}/_\#$	foη^{55}ʂwej^{214}	'feng-shui'
	$\rightarrow T_2{}^{35}/_T_3$	ʂwej^{35}kwo^{214}	'fruit'
	$\rightarrow T^{21}$/elsewhere	ʂwej^{21}k^{h}u^{51}	'reservoir'

Standard Mandarin does have other kinds of sandhi with clitics and in a few other high-frequency morphemes, and with sandhi domains longer than two syllables, but the generic disyllabic sandhi is limited to this. Note that the sandhi is context-specific with respect to the category of the following tone, and involves both structure-preserving change ($T_3 \rightarrow T^{35}$ merges with T_2) and non-structure-preserving change ($T_3 \rightarrow T^{21}$ does not merge with another category).

2.3.2 Xiamen/Taiwanese

The sandhi of Xiamen (Chen, 1987) and Taiwanese (Peng, 1997), whose tones differ only in pitch heights, is characterized by the well-known 'tone-circle'. Each tone category has one contour in final position (Ex. 9a,c), and changes to another contour in non-final position (Ex. 9b,d), forming a circular counterfeeding chain (Figure 2). We ignore ④ tones, which follow a separate similar swap pattern.

(9)

a. $p^{h}e^{\mathbf{24}}$ we^{24} → $p^{h}e^{\mathbf{22}}$ we^{24}
leather shoe

b. $we^{\mathbf{24}}$ twa^{21} → $we^{\mathbf{22}}$ twa^{21}
shoe lace

c. $k^{h}i^{\mathbf{53}}$ $ts^{h}u^{21}$ → $k^{h}i^{\mathbf{44}}$ $ts^{h}u^{21}$
build house

d. $ts^{h}u^{\mathbf{21}}$ $tiŋ^{53}$ → $ts^{h}u^{\mathbf{53}}$ $tiŋ^{53}$
house top

In contast to the Mandarin sandhi, Xiamen sandhi is across-the-board and context-general. To visualize this, the disyllabic sandhi contours of ①, ② and ③ are displayed in Table 5, with the tone changes highlighted. For every category, the contour generalizations in the rows are different than the contour generalizations in the columns, but there are no exceptions to those generalizations in the table. Furthermore, all the changes are structure-preserving.

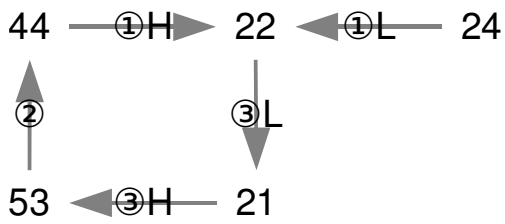


Figure 2: The Xiamen tone circle. The tone contour in initial position and the tone contour in final position are different for each tone category, with $①_H$ and $①_L$ merging in initial position, and four of the categories forming a counter-feeding chain.

Table 5: Xiamen Sandhi Contours. The first syllable tone categories are along the left margin and the second syllable tone categories are along the top margin. The tones are marked according to Middle Chinese (MC) categories. The body of the table contains the two syllable tone contours, and the 'row/col trend' is a generalization of the first syllable contours in the rows and of the second syllable contours in the columns. The tone changes are highlighted in yellow.

				2nd Syllable				
			MC class	①		②	③	
			MC reg.	**H**	**L**		**H**	**L**
	MC class	**MC reg.**	**row/col trend**	**44**	**24**	**53**	**21**	**22**
1st Syllable	①	**H**	**22**	22.44	22.24	22.53	22.21	22.22
		L	**21**	21.44	21.24	21.53	21.21	21.22
	②		**44**	44.44	44.24	44.53	44.21	44.22
	③	**H**	**53**	53.44	53.24	53.53	53.21	53.22
		L	**21**	21.44	21.24	21.53	21.21	21.22

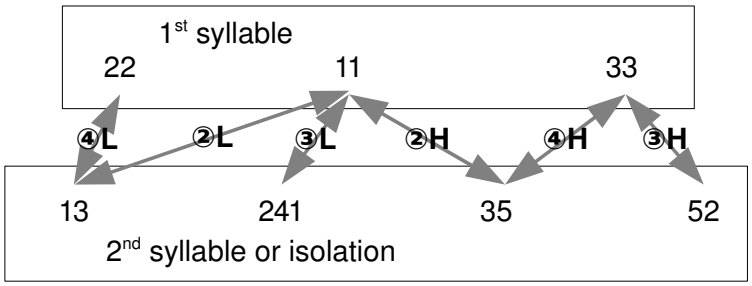


Figure 3: Yongkang sandhi tone type recovery. The initial tone contour and the final tone contour are different for each tone category. Both the initial contours and the final contours have mergers of the tone categories, but the mergers in the two positions are not equivalent.

3 Jinhua Sandhi

The area around Jinhua is known for complicated sandhi systems as well as lexical variation and regional variation. Generational differences are also acknowledged but poorly documented. These issues, combined with the typical difficulties of representing and comparing tone contours, make the documentation of Jinhua sandhi challenging.

3.1 The complexity of Southern Wu sandhi

The dialects of Southern Wu have particularly complicated sandhi (Cao, 2002). Most tone categories change in most sandhi contexts, in patterns that are often highly context-specific. The sandhi tones are sometimes structure-preserving and sometimes not, but there are enough mergers that the number of disyllable contours is generally less than half as many as are possible from the two-syllable combination of the 7 or 8 tone categories. The area around Jinhua is in many ways the most complicated.

In several dialects, including Jinhua, Tangxi and Yongkang, there are so-called 'tone type recovery' phenomena, where historical tone categories that have merged in isolation are treated differently in sandhi. For example, in Yongkang (Ballard, 1988), $④_{\mathrm{H}}$ and $②_{\mathrm{H}}$ have merged in isolation form, as have $④_{\mathrm{L}}$ and $②_{\mathrm{L}}$. In disyllabic sandhi, these keep their isolation form in the final syllable, but in the first syllable, $②_{\mathrm{L}}$ and $②_{\mathrm{L}}$ merge with $③_{\mathrm{L}}$, while $④_{\mathrm{H}}$ merges with $③_{\mathrm{H}}$, and $④_{\mathrm{L}}$ takes a third contour (Figure 3). Since final and non-final contours represent non-equivalent mergers, neither the final nor non-final contours alone can serve as underlying tone identities. (There is some context-dependence when ① tones are involved, but that data does not make things any clearer.)

Another kind of complexity is that sometimes there can be multiple sandhi patterns for a single combination of tone categories. Cao (2002, p. 111) reports the example of three patterns for the combination ③$_{H}$+①$_{L}$ in Jinhua.

(10) ③$_{H}{}^{55}$ ①$_{L}{}^{313}$ →

a. $T^{33}T^{14}$

pu.a	*tɕju.dʑjoŋ*
cloth shoe	bedbug

b. $T^{33}T^{55}$

su.ju	*sje.maw*
veggie oil	delicate

c. $T^{55}T^{3}$

t^{h}ja.məŋ	*sje.dʑju*
steel door	snowball

Cao (2002) explains that the tone reduction pattern in (Ex. 10c) occurs if the first syllable is one of the Middle Chinese *-ap and *-at syllables that have migrated from ④$_{H}$ to ③$_{H}$ in Jinhua. (This sandhi pattern is an example of tone type recovery.) But he knows no explanation for the difference between (Ex. 10a) and (Ex. 10b).

3.2 Previous reports

Previous reports on Jinhua sandhi agree that it is complicated, but disagree considerably on sandhi neutralizations and contours. There are three previous reports of sandhi contours from Jinhua.

3.2.1 Qian Nairong

The Wu survey of Qian (1992) included Jinhua among the 17 locations where sandhi data was gathered. The study followed the dialectology methods developed by Chao (1928) and colleagues, depending on their own ears to classify pitch contours, with the assistance of the native speakers in distinguishing categorical differences. They did make audio recordings, and they did have access to a Visi-Pitch machine, but found transcription by ear more convenient. In Jinhua, they collected isolation tone data from 12 speakers: six high school students, one 34 year-old, one 48 year-old, and four speakers over 65.

The isolation tone data shows differences between young, middle and old generations (just the middle generation is in Table 6), but there is no such annotation for the sandhi data, so it appears that the sandhi data may have been collected from just one or two individuals.

Table 6: Tone categories of Jinhua middle generation compared to Middle Chinese, according to Qian (1992).

	MC ①	MC ②		MC ③	lit. MC ④	cas. MC ④
MC $[-vc]$ onset	①$_{\mathrm{H}}^{435}$	②$_{\mathrm{H}}^{544}$		③$_{\mathrm{H}}^{45}$	③$_{\mathrm{H}}^{45}$	④$_{\mathrm{H}}^{44?}$
MC $[+vc, +son]$ onset	①$_{\mathrm{L}}^{213/324}$	②$_{\mathrm{H}}^{544}$		③$_{\mathrm{L}}^{24}$	③$_{\mathrm{L}}^{24}$	④$_{\mathrm{L}}^{22?}$
MC $[+vc, -son]$ onset	①$_{\mathrm{L}}^{213/324}$	②$_{\mathrm{H}}^{544}$	③$_{\mathrm{L}}^{24}$	③$_{\mathrm{L}}^{24}$	③$_{\mathrm{L}}^{24}$	④$_{\mathrm{L}}^{22?}$

Because there were so many locations in the survey, the researchers were not able to adapt their word list to each dialect. As a result, the sandhi tables reported there are based on the Middle Chinese tone categories, without regard to splits and mergers in the individual dialects. The sandhi table for Jinhua (Table 7), which has the sandhi effects highlighted, includes what at first glance appears to be a great deal of unexplained variation, but at least some of it is actually traceable to Jinhua's redefined tone categories. The contour generalization about the ②$_{\mathrm{L}}$ row is that there are two forms, T^{54} and T^{24}, and similarly the ②$_{\mathrm{L}}$ column has two forms, T^{51} and T^{24}. The isolation tone data indicates that for all speakers, many of the Middle Chinese ②$_{\mathrm{L}}$ had converted to ③$_{\mathrm{L}}^{24}$, while most other ②$_{\mathrm{L}}$ had merged with ②$_{\mathrm{H}}^{544}$. So the first syllable tones in the ②$_{\mathrm{L}}$ row are barely changed compared to the citation tones, just shortened in the case of $\mathrm{T}^{544}\rightarrow\mathrm{T}^{54}$. Similarly, the variation in the ②$_{\mathrm{L}}$ column might likewise be mostly from this shift in the tone categories, with T^{51} corresponding to isolation tone T^{544}. But there is no apparent explanation for the consistent split in the ①$_{\mathrm{H}}$ row, or any of the remaining variation. In addition, the sandhi is extensive and complicated. The row and column generalizations, which are simply the most common one or two contours, differ between row and column in five of the six tone categories. And the exceptions to these generalizations are scattered throughout the table, in both the first and second syllable. The row and column generalizations are also different than the isolation contours, except for ③$_{\mathrm{L}}$.

3.2.2 Cao Zhiyun

Cao (1996) produced a table of sandhi contours in the process of compiling a dictionary. The methodology was much like that of Qian, but much more in depth. Cao is a native of Tangxi, so even though his native phonology is somewhat different, he had more prior familiarity with Jinhua dialect than Qian and colleagues, who are from Northern Wu areas. The lexicon and pronunciation were gathered primarily from just one speaker born in 1927. Table 8 adapts the sandhi table found in the dictionary preface, which divides the syllables

Table 7: Jinhua sandhi of ①, ② and ③ according to Qian (1992). The first syllable tone categories are along the left margin and the second syllable tone categories are along the top margin. The tones are marked according to Middle Chinese (MC) categories. The body of the table contains the two-syllable tone contours, and the 'row/col trend' is a generalization of the first syllable contours in the rows and of the second syllable contours in the columns. The tone changes are highlighted. The row/column generalizations and the tone change highlights are not in the original source and are added here as a guide.

				2nd Syllable					
			MC class	①		②		③	
			MC reg.	**H**	**L**	**H**	**L**	**H**	**L**
	MC class	**MC reg.**	**row/col trend**	**24**	**31**	**51**	**51/24**	**24/45**	**24**
1st Syllable	①	**H**	**33/43**	33.55 43.24	33.55 55.31	33.51 43.24	33.51 43.24	33.55 43.24	33.55 43.24
		L	**32**	32.24 22.55 33.55	32.24	22.51 33.51	22.51 33.51	32.45	32.24
	②	**H**	**54**	54.24	55.31	54.51 54.44	54.51 54.44 54.24	54.45	54.24
		L	**54/24**	54.24 24.24	55.31 24.31	54.24 54.51 24.51	54.24 24.51 54.51 24.24	24.24 24.51 54.24	24.24 54.24
	③	**H**	**55/45**	55.31 54.24 33.55	55.31 45.24	33.51	33.51 45.24	55.31 45.45 33.55	45.24
		L	**24**	24.24	24.31 24.24	13.51	13.51 24.24	13.51	24.24 45.24

according to their modern tone categories (Table 9). We leave aside patterns labeled as verb+object and ④ tones again.

Compared to Table 7, this is a simpler cleaner table, but still there are extensive sandhi effects, and several cells with variation. The row and column generalizations are different for all six categories, and there are many exceptions to these generalizations. The column generalizations for class ① are also different than the isolation tones. There is much overlap between the columns for $①_H$ and $③_H$ and for $①_L$ and $③_L$, though the categories are not fully merged. And $②_H$ and $②_L$ are almost identical in the columns and rows, but that is not surprising since $②_L$ in isolation has merged with $②_H$. There are regularities, but a complete description in phonological rules, much less in a constraint system, would still be quite complicated and full of exceptions. The tone reduction pattern pointed out in Cao (2002) appears in this table as 55.3, and we can surmise that the 24.3 patterns are produced by $③_L$ syllables that historically came from $④_L$. Where this table and the previous one are comparable, in some places they agree well, such as the 33 in the top row and the 24 in the last column, and other places they don't, like the 51 in the Table 7 ② columns corresponding to 535 in the Table 8 ② columns.

3.2.3 You and Yang

The third study was part of a phonetic survey of Wu tone systems. You and Yang (2001) combine the traditional dialectology methods with the tonetics methods introduced to them by Zhu (1999). The sandhi contours are categorized and transcribed with the aid of automated analysis of recorded words, using three words per tone combination. The data was recorded from one speaker born in 1949. The resulting contours are displayed in Table 10.

In this data, there is considerably less unexplained variation, with just one cell showing two sandhi patterns, and there are more mergers than previously. However, the mergers that were hinted at in Table 8 are not the ones that are evident here. Column $①_H$ is quite different than $③_H$, as $①_L$ is from $③_L$. The ② rows and columns are again almost merged, but not quite, and the cells that distinguish the registers are not the same cells as in Table 8. If we look at contours, many of the differences are trivial, such as 323 vs 313 and 24 vs 14, but there are also again tones that do not match well, like the places 35 in this table corresponds with 55 in the previous two. This table is in agreement with the previous ones that there is extensive sandhi, but differs considerably on the details.

Table 8: Jinhua sandhi of ①, ② and ③ according to Cao (1996). The first syllable tone categories are along the left margin and the second syllable tone categories are along the top margin. The tones are marked according to and contemporary Jinhua (JH) categories. The body of the table contains the two-syllable tone contours, and the 'row/col trend' is a generalization of the first syllable contours in the rows and of the second syllable contours in the columns. The tone changes are highlighted. The row/column generalizations and the tone change highlights are not in the original source and are added here as a guide.

				2nd Syllable					
			JH class	①		②		③	
			JH reg.	**H**	**L**	**H**	**L**	**H**	**L**
	JH class	**JH reg.**	**row/col trend**	**55**	**24**	**535**	**535**	**55**	**24**
1st Syllable	①	**H**	**33**	33.55	33.55	33.55 33.535	33.55 33.535	33.55	33.55
		L	**31**	31.55	31.24	33.535	33.535	33.55 31.55	31.24
	②	**H**	**55/53**	55.33	55.313	53.535	53.535	53.55	55.24
		L	**55/53**	55.33 53.55	55.313	53.535	53.535	53.55	55.24
	③	**H**	**33**	33.55 55.3	33.24 33.55 55.3	33.535 53.535	33.535 53.535	53.55	55.24 33.55
		L	**53/24**	53.55 24.3	53.24 24.3	33.535 53.535	33.535 24.3	31.55 53.55	55.24 31.24 24.3

Table 9: Jinhua tone categories (Cao, 1996) compared to Middle Chinese

	MC ①	MC ②	MC ③	MC *-ap,*-at	other MC ④
MC $[-vc]$ onset	$①_H{}^{334}$	$②_H{}^{535}$	$③_H{}^{55}$	$③_H{}^{55}$	$④_H{}^{4}$
MC $[+vc, +son]$ onset	$①_L{}^{313}$	$②_H{}^{535}$	$③_L{}^{24}$	$③_L{}^{24}$	$④_L{}^{212}$
MC $[+vc, -son]$ onset	$①_L{}^{313}$	$③_L{}^{24}$	$③_L{}^{24}$	$③_L{}^{24}$	$④_L{}^{212}$

Table 10: Jinhua sandhi of ①, ② and ③ according to You and Yang (2001). The first syllable tone categories are along the left margin and the second syllable tone categories are along the top margin. The tones are marked according to contemporary Jinhua (JH) categories. The body of the table contains the two-syllable tone contours, and the 'row/col trend' is a generalization of the first syllable contours in the rows and of the second syllable contours in the columns. The tone changes are highlighted. The row/column generalizations and the tone change highlights are not in the original source and are added here as a guide.

				2nd Syllable					
			JH class	①		②		③	
			JH reg.	**H**	**L**	**H**	**L**	**H**	**L**
	JH class	**JH reg.**	**row/col trend**	**35/323**	**35/323**	**43/44**	**43/44**	**35**	**14**
1st Syllable	①	**H**	**33**	33.35	33.35	33.43	33.43	33.35	33.35
		L	**31**	31.35	31.35	33.43	33.43	31.35	31.14
	②	**H**	**44**	44.323	44.323	44.44	44.44	43.34	43.34
		L	**44**	44.323	44.323	44.44	44.44	43.34	45.14
	③	**H**	**22**	22.45	44.323 33.35	22.45	45.14	22.45	44.323
		L	**54**	54.45	45.14	54.55	54.55	31.35	45.14

3.3 The documentation problem

The major differences among the results of these three studies begs for an explanation. All three studies claim to document the disyllabic sandhi of urban Jinhua, but the tone contours and neutralizations within the sandhi tables are not compatible. For several reasons, the complexity and dimensions of variation in the sandhi of that area make it difficult to obtain reliable records.

Because words from the same combination of tone categories may represent two or more sandhi patterns, and it is currently unknown how to predict which sandhi patterns are associated with particular words, any set of stimuli might be incomplete. The resulting record of sandhi patterns would then be incomplete, and two such records would disagree. Similarly, if the researcher is not careful to distinguish the different sandhi domain types, the records would conflate sandhi subphonologies.

Because there is so much geographical variation in the region, it is possible that local variation even within the city might lead to conflicting records. Some neighborhoods have been settled by clans and work units from particular villages or neighboring cities. For example, the neighborhood around the old train station was settled by railroad workers, mostly from Hangzhou, about 100 years ago, and according to people from that neighborhood, the Wu spoken there is still a blend of Jinhua and Hangzhou features. And since Wu is not the major language of commerce in Jinhua, differences among neighborhoods can persist.

Generational differences are also known to exist but have not been documented well. Qian (1992) recorded generational differences for the isolation tones, but not for the sandhi contours. If it is true that Qian's sandhi data is from just one person, than none of these studies attempted to combine or compare sandhi data from multiple speakers. Since the disappearance of a distinct ②$_{\mathrm{L}}$ tone category and perhaps the loss of the obstruent voicing contrast are apparently fairly recent phenomena, it is possible that the sandhi system is undergoing a related reorganization. Some differences between the studies might reflect generational differences in a period of rapid language change.

Because none of these studies apparently compared data from different speakers, it is also possible that some elements of the recorded sandhi patterns reflect idiosyncrasies of the speaker's pronunciation.

Finally, some of the differences might arise from methodological differences. The earlier studies of Qian (1992) and Cao (1996) depended on traditional methods of word list elicitation and impressionistic transcription. The acoustic tonetic methodology used by You and Yang (2001) attempts to overcome the subjectivity in traditional methods of impressionistic transcription, but they still depend on subjective judgments to determine the contour categories and ultimately the Chao tone transcriptions. Impressionistic transcription of tones

might involve some unconscious abstracting away from coarticulation effects, whereas the transcriptions based on plotted pitch tracks would be either more phonetic or more consciously abstracting away from coarticulation effects.

To what extent do each of these factors account for the differences among the previous studies? To answer that question, we need to compare data from multiple speakers in a reproducible objective procedure.

4 Procedure and Results

Jinhua native speakers from two generations and two locations were recorded as they read disyllabic words in a carrier sentence. The extents of the tone domains in the target words were annotated by hand, and F0 contours were extracted algorithmically. A reparameterization of the F0 contours was then selected which reduced between-speaker variability and the dimensionality of the representation. The extracted pitch contours were used to measure the similarities among speakers' sandhi systems, and three speaker groups were formed based on the sandhi system similarities and the speakers' social variables. Finally, disyllabic sandhi contour categories were discovered through clustering the words by their F0 contours.

4.1 Elicitation

4.1.1 Sandhi domains

This study focuses on disyllabic sandhi, that is, sandhi phenomena that appear in two-syllable sandhi domains. It is therefore necessary to be clear what a sandhi domain is. Chen (2000) devotes much attention to this issue, concluding that sandhi phenomena are active on a wide variety of domains across Chinese languages. In each language, one or more levels of the prosodic hierarchy define the extents of the sandhi domains, and the syllables with higher prominences (promoted by headship within the prosodic domains) are less likely to undergo tone reduction or replacement and more likely to have their tone spread to neighboring syllables.

It is not clear from the literature what the domains of sandhi are in Jinhua, but it appears that several prosodic domains might be relevant, while prominence is right-headed. There are three reported kinds of sandhi domains in Jinhua (Cao, 1996, 2002). The typical sandhi patterns are found in compound words and domains composed of modifier+noun. But some of the Jinhua sandhi patterns are associated with verb+object and subject+predicate domains. And number+measure-word and content-word+function-word domains also have distinct sandhi patterns, frequently involving changes to the second syllable. The analysis of these patterns is beyond the scope of this discussion, but one possible analysis, along the lines of Chen's analysis of other dialects, is that the verb+object and subject+predicate domains are a phonological phrase (roughly, material between right edges of XPs), that the compound and modifier+noun domains are a phonological word, and that the content-word+function-word domain is a phonological word with an enclitic. The syllable prominence is somewhat difficult to diagnose, because Jinhua and surrounding dialects are unusual among Chinese languages in having many replacement tone changes on both syllables in disyllabic

sandhi. But tendencies towards tone change still suggest prominences. In compound and modifier+noun disyllabic domains, changes in the first syllable are reported to be a bit more than changes in the last syllable Cao (2002). And in three-syllable domains, the second and third syllables undergo the same complicated tonal replacement seen in two-syllable domains and the first syllable reduces to one of just three tone contours. The sandhi patterns that are unique to disyllabic verb+object domains have some change on the first syllable and no change on the last syllable, indicating right-headedness. The sandhi of disyllabic number+measure-word and content-word+function-word domains generally involves changes to both syllables or to just the final syllable, and since the final syllable is a function word, it is likely a clitic that is necessarily non-prominent. In order to reduce the complexity of the analysis, this study focuses on the compound and modifier+noun sandhi, avoiding the other two sandhi domains. And because sandhi domains are right prominent, the carrier sentence needs to have the target words at the right edge of the phonological phrase.

4.1.2 Stimuli

As with You and Yang (2001), the stimuli consisted of 3 disyllabic words from each of the possible combinations of the 8 traditional tone categories, a total of 192 words. They used slightly different word lists for each of the locations in their survey, because of the dialectal differences in vocabulary. They do not provide the word list used for Jinhua, but they do give lists for several of the other locations, and one of these (Linhai, a Northern Wu dialect) was used as the starting point for constructing the stimuli list for this study. A few Jinhua native speakers were asked to identify words in that list which are not used or are rarely used in Jinhua, and these words were replaced with words from the Jinhua dialect dictionary (Cao, 1996) and again checked for contemporary usage. The original word list is designed to all be of the modifier+noun and compound structures, especially excluding subject+verb and verb+object structures. The replacement words were also selected to match this compound/modifier+noun type. Because the words must be recognizable from Chinese characters, most of the words are cognate with Standard Mandarin words, though a couple words added from the dictionary are merely composed of common cognate morphemes. The complete word list is in Appendix A. At the time of word list preparation, it was thought that morphemes of MC ②$_{\mathrm{L}}$ moving to Jinhua ③$_{\mathrm{L}}$ (§ 2.1) were exceptional, and an attempt was made to reduce the number of these items. As a result, there are fewer obstruent-onset ②$_{\mathrm{L}}$ syllables in the word list than sonorant-onset ②$_{\mathrm{L}}$ syllables, which all merged with ②$_{\mathrm{H}}$ rather than ③$_{\mathrm{L}}$.

4.1.3 Speakers

A total of 15 speakers from Jinhua and the nearby village of Zhuma were recruited through personal contacts. From the city, there were 5 speakers born between 1931 and 1952 (2 M, 3 F) and 5 speakers born between 1984 and 1989 (2 M, 3 F). From the village, there were 2 speakers born in 1946 and 1949 (1 M, 1 F), and 3 speakers born in 1985 and 1986 (1 M, 2 F). The speakers understood that the study was about the pronunciation of words in the dialect, but except for speaker 04, who had helped in the preparation of the materials, none of the speakers, in interviews afterwards, indicated that they had noticed the importance of the tones.

Table 11: Speakers and their social variables

Speaker	Location	Born	Gender	Note
18	Jinhua	1931	M	neighbor of speaker 17
16	Jinhua	1944	F	
21	Jinhua	1944	F	
17	Jinhua	1948	M	neighbor of speaker 18
20	Jinhua	1952	F	
22	Jinhua	1984	F	
04	Jinhua	1986	F	same-clan cousin of speaker 19
19	Jinhua	1987	M	same-clan cousin of speaker 04
24	Jinhua	1988	F	
23	Jinhua	1989	M	
12	Zhuma	1946	F	
06	Zhuma	1949	M	
07	Zhuma	1985	F	coworker of speakers 04 and 08
08	Zhuma	1986	F	coworker of speakers 04 and 07
13	Zhuma	1986	M	

4.1.4 Presentation

Each word was presented in a carrier phrase, written in Chinese characters and displayed on a notebook computer screen, in 40 pt. black type on a white background. An example word for practice was displayed first, and then the stimuli were each presented once, with each sentence preceded by the sequence index number. The speakers were asked to speak completely in Jinhua dialect, saying the index number once and the carrier phrase three times. Instructions were written in Mandarin, and additional oral explanation was given in Jinhua dialect, Standard Mandarin, or English, depending on the proficiency and preference of the speaker. Speakers took a 3 to 10 minute break half-way through, after 96 stimuli.

The carrier phrase differed from that of You and Yang (2001). Whereas they used the sentence $a_{②L}$ $kaŋ_{②H}$ _ $fəŋ_{①H}$ $noŋ_{②L}$ $t^hiŋ_{③H}$ ('I say _ for you to hear'), this study used the sentence $t^hiŋ_{③H}$ _ $sa_{①H}$ $pje_{③L}$ ('Listen to _ three times'). In either case, the target is a phonological word and thus a sandhi domain, and as the complete direct object, it forms a noun phrase and the right edge of a phonological phrase. Since the last two syllables of trisyllabic sandhi are described as being identical to disyllabic sandhi Cao (2002), if the verb and the target word form a sandhi domain together, this would not effect the sandhi tones on the target word. The target word is also contrast focused, so we can expect somewhat expanded pitch range and less coarticulation effects than in uncontrolled speech.

4.1.5 Recording

The speech was recorded in Audacity on the same computer used for presenting the stimuli, using a headset microphone and the built-in Realtek ALC272 HDA sound card, which had a frequency cut off at 6 kHz. Recording input levels were adjusted when the speakers practiced on the example stimulus, and they were sometimes readjusted during the break. Because speakers sometimes changed voice loudness or moved the microphone, the intensity varied, including frequent high-amplitude clipping for some speakers. The clipping may have made the annotation of tone domain extents less accurate, but it does not affect the measurement of fundamental frequency.

4.2 Annotation

The sound files were split apart and the extents of the tone domains were annotated in Praat with the assistance of scripts. The Audacity files were converted to WAV format, and each repetition of each stimulus, including carrier phrase, was segmented into its own file, by hand in Audacity, or in Praat with the assistance of Mietta Lennes' scripts for cutting and labeling sound files.[3] The extents of tone domains were annotated in the second repetition of each stimulus. The third repetition was also intended for annotation, but the large number of words and speakers meant that the annotation of the third repetition became a low priority. The creation of text tiers and the opening and closing of files were facilitated by Yi Xu's fundamental frequency (F0) extraction script[4], but extent boundaries were placed based only on visual inspection of Praat's waveform, spectrogram, intensity, and formant information. Pitch tracks and pulse marks were hidden from view. [5]

[3] http://www.helsinki.fi/~lennes/praat-scripts/#cutting

[4] http://www.phon.ucl.ac.uk/home/yi/_TimeNormalizeF0.praat version 2.7

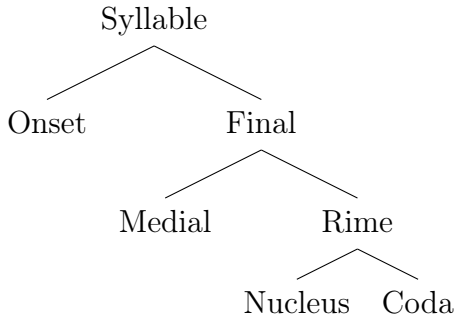
[5] A second annotator labeled 4.5% of the files. The mean inter-annotator difference on the locations of extent boundaries was 8.9 ms, less than 5% of the length of a typical tone domain.

4.2.1 Syllable structure

In order to compare tone contours across syllables that have different segmental structures, it is necessary to know what the general structure of the syllable is, and where the tone contour is within the syllable. There have been many different frameworks for the structure of syllables, but we focus here on one framework, the traditional Chinese model, which is derived from classical Chinese scholarship, formalized in Western terms by Cheng (1966), and recently revived in treatments of Mandarin (Y.-H. Lin, 1989; Weijer & Zhang, 2008) and Wu (Zhu, 1999; Zhang, 2006) as well as Chinese languages in general (Chen, 2000).

Looking at the segment sequence, the typical Chinese syllable fits the template CGVX, where C is one of a wide variety of consonantals, G is a glide (in Jinhua, one of /j/, /w/, or /ɥ/), V is a vowel, and X is one of a restricted set of glides, nasals, or stops. In Jinhua, X is one of /j/, /w/, /n/, /ŋ/ or /ʔ/. Considered in terms of syllable structure, these four segments are labeled as onset, medial, nucleus and coda. The nucleus and coda constitute a rime, as in conventional onset-rime theory (Blevins, 1995), but contrary to the onset-rime model, the final constitutes an additional layer of structure, combining the medial and rime before joining with the onset to form the syllable.

(11)



This additional layer of structure represents the belief that the medial is more associated with the rime than it is with the onset, yet not as associated with the nucleus as the coda is.

The nucleus is the only unambiguously obligatory element. But there has been an ongoing debate about whether the onset is also obligatory. Chao (1968) identified a 'zero onset' which he described as underlyingly unspecified, but surfacing in Beijing Mandarin as [ʔ], [ɦ], [ŋ], or [ɣ], depending on the context and the speaker. Some have followed Chao in considering the zero onset to be an obligatory but underspecified segment, while others, like Duanmu (2007) have taken the view that it is simply an effect of syllable constraints on a vocalic input. Duanmu notes that when there is a prenuclear glide or a high vowel nucleus, no 'zero onset' is possible, suggesting that the constraints that result in 'zero onset' phenomena are resolved by the glide when a glide or high vowel is available. In respect to tonal alignment,

Rose (1998) points out that in syllables that would have no consonantal onset but have a pre-nuclear glide or a nuclear high vowel, the onset is apparently realized as a lengthening of the glide. The pre-nuclear glide is extended in length to occupy the positions of both the pre-nuclear glide and the onset, and part of the glide is produced prior to the characteristic contour.

4.2.2 The tone domain

In some sense, the tone is associated with the whole syllable, and some feature geometries and articulation models (e.g., Cheng, 1966; Yip, 1989; Y. Xu & Wang, 2001) associate the tone with the syllable as a whole. Possible tone contours are constrained sometimes by the coda (dialects with stop codas restrict those syllables to class ④ tones) and by the onset (dialects with voicing contrasts often maintain the association between register and onset voicing). But onsets are not part of the characteristic tone domain, and pitch tracks of the medial and coda are apparently less reliable than that of the nucleus. As shown for Beijing Mandarin (Howie, 1976) and Zhenhai Wu (Rose, 1998), pitch tracks in sonorant consonantal onsets are not part of the characteristic tone contour, and similarly, glides in GV and GVX syllables (where the glide satisfies onset constraints) appear to be (mostly) outside the tone domain. In contrast, the glide of CGV syllables, at least in Zhenhai Wu (Rose, 1998), aligns (in overall timing and in contour position) with the beginning of the vowel in CV syllables, suggesting that prenuclear glides are within the tone domain when there is a consonantal onset. While a stop coda can restrict the tone to a short contour, it obviously has no pitch track. And a nasal or glide coda is known to carry part of the characteristic contour, though less reliably than the nucleus. M. Lin (1995) found that native speakers of Beijing Standard Mandarin could fairly reliably identify the tone when hearing a subsyllabic sample from the nuclear vowel portion, but could not identify the tone when hearing prenuclear or postnuclear segments. Several studies (e.g. Zhang & Hirose, 2004; Rose, 1987) have found that automated recognition of tones or coherence of tone categories improve when the initial and final sampling points in the contour are ignored. Some of this effect is due to tonal coarticulation or onset-pitch interactions, but the unreliability of the contour end pitch may be related to the coda segment. Zhu (1999) found very little benefit of excluding marginal sampling points in Shanghai, which does not have sonorant codas (Duanmu, 1993).

So though a tone may associate with a whole syllable, the characteristic contour is limited to the sonorant part of the final. And even within the final, the central pitch track carried by the nucleus might be more indicative of tonal identity than the marginal pitch track carried by prenuclear glides or sonorant codas.

4.2.3 Tone domain onsets

For the purpose of having a tone domain onset that is both reliably annotated and fairly comparable across syllable types, this study works with a definition of tone domain that includes prenuclear glides but excludes syllable onsets. After unaspirated stops, the tone domain begins at the first glottal pulse after the release. After aspirated stops, affricates and fricatives, the tone domain begins with the first pulse of modal voicing. In the case of aspiration, the onset of modal voicing is accompanied by increased intensity, but with affricates and fricatives, the transition from consonant to vocoid is more smooth, as the low frequency formants appear above the frication noise before the high frequency formants do. The onset of the tone domain was taken to be when voicing striations appeared in F2, or if F2 was never clearly visible (often the case for /u/ and /ɿ/)[6], striations in the first formant (F1). After nasals and laterals, the tone domain begins at the first pulse after the intensity jump of the release. Based on the alignment of Zhenhai Wu onsets (Rose, 1998) and the generalization from Standard Mandarin about the 'zero onset effect' (Duanmu, 2007, p. 72-79), this study assumes that even when the phonological specification of a syllable starts with a glide or vowel, some segment still occupies the onset position. If a glottal stop is found, or if initial breathy/creaky voice quickly gives way to modal voicing, the tone domain begins at the first pulse of modal voicing. Otherwise, if the intensity curve has a rise during a high vowel or prenuclear glide, the tone domain starts at the steepest point of the rise. If neither condition holds, the domain boundary is placed based on the formants, assuming an early steady state as the syllable onset, and a later dynamic portion as prenuclear glide.

4.2.4 Tone domain offsets

The tone domain was defined to include nasal and glide codas but exclude stop codas. When the syllable is followed by a voiceless segment or silence, the tone domain ends at the last voicing striation in F2 or F3. When the syllable is followed by a voiced segment, the tone domain ends when there is an intensity drop or change in formant structure. Nasal codas, when followed by another nasal in the onset of the following syllable, often fully assimilate to the following onset. When no formant transitions are apparent, half of that nasal portion is assigned to the tone domain, and the other half is assigned to the following syllable. In the case of underlying stop codas, however, the tone domain is assumed to end when the closure begins, even if the stop assimilated to a following sonorant, because the tone domains of different words of the same tone category have to be comparable. Finally, at the end of a syllable, highly creaky phonation with periodicity greater than 12 ms (i.e. 83 Hz) is assumed

[6] /ɿ/ is the conventional transcription of the Chinese vowel similar to IPA /z̩/

to represent a glottal stop, outside the tone domain.

4.3 Pitch contours

Pitch tracks were extracted from target regions of the sound files, and then reparameterized by normalization and summarization methods. The goal of contour parameterization is to convert the raw frequency samples to low-dimensional representations that can be more easily and accurately combined, compared and categorized.

4.3.1 Previous work

Speakers have large differences in larynx size and corresponding F0 ranges, as well as differences in habitual linguistic use of that range. Human recognition of linguistic pitch can abstract away from those differences (Moore & Jongman, 1997), even while speaker differences are also recognized and can be used for word recognition (Creel, Aslin, & Tanenhaus, 2008). Human speech processing apparently makes use of a rich memory of linguistic experience for speech perception (Johnson, 2005), but in the context of a limited quantity of lab speech from each speaker, automated approximation of pitch recognition must rely on a few general characteristics of each speaker's range. In order to combine and compare measurements among speakers, a normalization function must be established which maps the raw F0 measurements into a more speaker-independent space (Rose, 1987). In addition, since a pitch contour is composed of highly correlated sequential samples, a low-dimensional parameterization that summarizes a pitch contour may further improve category coherence as well as interpretability.

For a group of speakers who speak the same dialect, normalization should reduce between-speaker variability compared to phonemic distinctions, making categories more coherent. Machine learning studies (e.g., Zhang & Hirose, 2004; Thubthong, Kijsirikul, & Luksaneeyanawin, 2002) typically measure normalization success by the accuracy of automated classification, but this requires that the tone categories are known beforehand. Rose (1987) recommends reduction in Earle's dispersion coefficient (Earle, 1975) as the best measure of successful normalization. The dispersion coefficient is defined as the ratio of between-speaker variance to total sample variance. For N speakers and M pitch samples per speaker,

$$D = \frac{\sum_{i}^{N}\sum_{j}^{M}(f_{ij} - \mu_j)^2}{\sum_{i}^{N}\sum_{j}^{M}(f_{ij} - \bar{\mu})^2} \tag{12}$$

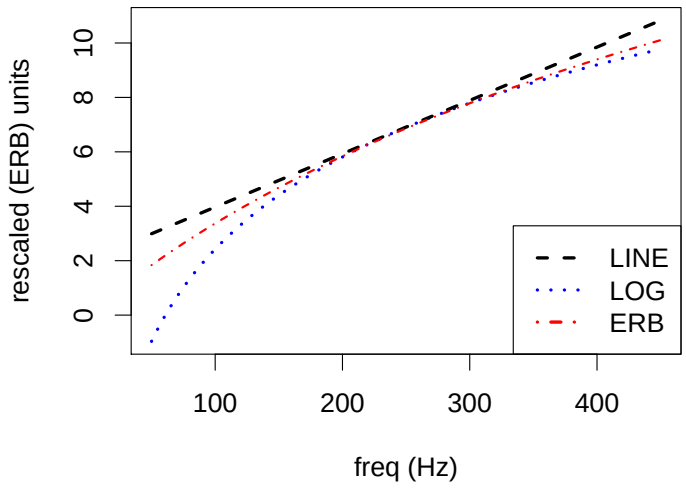


Figure 4: Comparison of linear, logarithmic and ERB scale curves, rescaled to overlap at f=250 Hz.

The f_{ij} denotes the frequency of speaker i at sample point j, which ranges over stimuli as well as time within the stimulus items. The μ_j is cross-speakers mean frequency at sample point j, that is, $\frac{1}{n}\sum_i^n f_{ij}$, and $\bar{\mu}$ is the grand mean. The ratio of pre-normalization D to post-normalization D is dubbed the normalization index. Rose (1987) and Zhu (1999) use the normalization index to choose a normalization function for isolation tones, while Rose (1991) uses it to choose a normalization function for running speech.

Among the various F0 normalization methods proposed in different fields (Zhu, 1999; Thubthong et al., 2002), there are three main issues in defining a normalization function: a frequency scale, a reference frequency, and a frequency range.

There are three kinds of frequency scales: linear, logarithmic and psychoacoustic. The linear scale is the most straightforward, measuring frequencies in cycles per second (Hz). Logarithmic scales (semitones or octaves) transform frequencies to a spectrum in which equivalent harmonics are equidistant. Psychoacoustic scales (mel, bark, or ERB) are derived from perceptual norming studies, judging subjective frequency distances or just-noticeable differences. In mathematical behavior, the psychoacoustic scales lie in between the linear and logarithmic scales, and in the range of F0, the ERB scale is closer to the linear scale than the logarithmic scale (Fig. 4). Zhu (1999) compared linear and logarithmic scales, and found

that the logarithmic scale consistently had a better normalization index on his Shanghai tone data. Thubthong et al. (2002) compared ERB scale with semitone and linear scales, in a machine learning tone recognition task on three Thai speech corpora, and found that ERB scale led to performance marginally better than the linear and semitone scales in two of the corpora, while semitone scale produced a slightly better result in the third corpus.

The frequency scales are speaker-independent, while incorporating a reference frequency and range in the normalization is an attempt to compensate for speaker differences, specifically the typical speaking pitch and pitch range. Rescaling from the raw frequency f_{ij} to the normalized frequency f'_{ij} is expressed generically as

$$f'_{ij} = \frac{f_{ij} - f_{ref}}{\Delta f} = \frac{f_{ij} - f_{ref}}{f_f - f_0} \tag{13}$$

where f_{ref} is the reference frequency and Δf is the frequency range factor, alternately expressible in terms of high (f_f) and low (f_0) reference points. Some normalizations make no re-centering adjustment to the frequencies, equivalent to setting $f_{ref} = 0$, while others try to identify a salient point that is hypothesized to characterize speakers' generic pitch, such as an extreme turning point or a level tone onset. Others use a statistical measure of central tendency derived from a large sample, typically the mean. Similarly, the range can be defined by two characteristic points, or by a statistical quantity such as the standard deviation. Rose (1987) considered two normalization functions for isolation tone data from Zhenhai: the fraction-of-range method, which uses individual salient points to define the reference and range, and the z-score method, which uses mean and standard deviation. Zhu (1999) considered the same two plus a third function, proportion-of-range, which uses a multiple of the standard deviation for the range, and uses one end of the range as the reference frequency. Both Rose and Zhu found that z-score has the best normalization index. Similarly, Thubthong et al. (2002) compared z-score with mean-normalization, which performs no re-centering ($f_{ref} = 0$) and takes the speaker's mean frequency as the range factor ($\Delta f = \mu_i$). Mean-normalization has been used by some researchers (e.g. Shriberg, Ferrer, Venkataraman, & Kajarekar, 2004; Stanford, 2008) and is approximately proportional to using standard deviation as range factor, because mean and standard deviation of speaker F0 distributions are highly correlated. But Thubthong et al. (2002) found z-score normalization led to higher accuracy tone recognition.

Additionally, a subsample of the measurement points or a summarized representation may make a better abstract model of a tone contour. Rose (1987) removed the first and last sampling points of each syllable's contour in the calculation of the z-score transform's mean and standard deviation, though he kept all sampling points in calculation of between-speaker

divergence. Likewise, Zhu (1999) removed the first sampling points in calculating the z-score parameters, but except for one tone type, no benefit was gained by removing the last sampling points in his data. Zhang and Hirose (2004) use a sophisticated algorithm to learn a 'tone nucleus', a central portion of the contour within the rhyme, on which their tone features are based. They re-represented the tone nucleus as a mixture of Gaussians over three frequency features and three energy features at five points within the tone nucleus. Thubthong et al. (2002) compared a third-order polynomial fit representation with a representation in terms of the slope at five equally-spaced points in the contour, plus F0 at either the endpoints or at all five points. They found that the highest accuracy was obtained with the five slopes plus five points.

4.3.2 Extraction of pitch tracks

Measurements of fundamental frequency at 10 points within each tone domain were obtained in Praat with an altered version of Yi Xu's F0 script. F0 measurements were obtained by Praat's native auto-correlation algorithm (Boersma, 1993) using the `To Pitch (ac)...` command, which permits more control over parameters than the `PointProcess` command used in the original script. In order to obtain an F0 reading even in breathy or creaky regions[7], the voicing threshold was reduced from the standard 0.45 to 0.25, as recommended in the Praat manual, and to reduce pitch excursions in creaky regions, the octave-jump cost was raised from the standard 0.35 to 2.0, and the voice/unvoiced cost was raised from the standard 0.14 to 1.0.

For accurate measurement of F0 in creaky regions, it is important to use an accurate estimate of the speaker's frequency range. The algorithm's frequency range for each speaker was determined by the distribution of F0 samples for that speaker. F0 measurements were initially acquired with the range set as 50 kHz to 450 kHz. Using the mean and standard deviation of the resulting F0 distribution, samples greater than 2.5σ from the mean were excluded, and then the mean and standard deviation were recalculated. Final measurements were acquired with the algorithm's frequency range set to $\pm 3\sigma$ from the mean.

In some cases the algorithm was still unable to find a smooth contour, and these contours with frequency jumps or long plateaus were excluded from the analysis. Frequency jumps, which result from creaky voice, were defined as increases or decreases of greater than 1/3 of an octave between sequential sample points, and the excluded plateaus were where more than half of a tone domain consisted of identical frequencies, which occurs when the auto-

[7]This study ignores the possibility of non-modal voicing being systematically related to tone identity, because Cao (2002) indicates that Jinhua does not have systematic non-modal voicing, while several other Southern Wu dialects do.

correlation algorithm finds no frequency in range, and the script interpolates beyond the last accepted measurement. Some items had no accepted measurement within a tone domain, and in these cases, the tone with no measurements is excluded, but the item's other tone domain was not excluded. The script's trimming of pitch excursions was disabled, to facilitate identification of problematic contours.

The stopped-tone (tone ④) stimuli included many words that some speakers pronounced with non-stopped tones or didn't recognize as Wu words, and because of the short duration and glottalization, pitch measurements were less reliable. For these reasons, tone ④ stimuli were removed from the analysis. Even after excluding all the ④ words, most speakers had additional words excluded for faulty pitch readings, and some speakers also skipped stimuli that they did not recognize or that they considered not used in their dialect. As a result, all speakers had missing items (Table 12), from a minimum of 2 to a maximum of 15.5, almost 15% of the 108 non-④ stimuli.

Table 12: Missing Stimulus Items by Speaker. Skipped items are stimuli that the speaker did not recognize as Wu words. Excluded items are stimuli that were too creaky or breathy to obtain a smooth pitch track for. The total number of stimuli is 108 (6 tones × 6 tones × 3 items)

Speaker	Skipped Items	Excluded Items	Total Gaps
16	0	2	2
17	0	4	4
18	0	15.5	15.5
20	0	6	6
21	0	2	2
04	0	10	10
19	3	3	6
22	0	8.5	8.5
23	1	2	3
24	1	2	3
06	0	5.5	5.5
12	2	0	2
07	0	2	2
08	5	0	5
13	0	3	3

4.3.3 Evaluation of parameterizations

A contour representation was chosen based on two measures of optimal parameterization. The ideal parameterization has minimal between-speaker dispersion or equivalently maximal normalization index (Rose, 1987), and it facilitates accurate recognition of tone cate-

gories by improving category coherence. Between-speaker dispersion coefficients and linear-discriminant (LDA) classification accuracies were calculated for each choice in defining the contour parameterization, and these guide the selection of a suitable parameterization.

The between-speaker dispersions were calculated according to equation 12, and standard errors were calculated by bootstrap sampling. Distributions of the same size and similar character to the actual frequency data were created via resampling the data by stimulus word with replacement, and the dispersion standard errors are calculated as the standard deviation of the dispersions of 100 such bootstrap samples.

Training a classifier to recognize the categories of disyllabic contours is not possible since it is not known a priori what the sandhi contour categories are, much less which words belong to each category. Instead, the contour of each stimulus word is taken to be a separate category, and a classifier is trained to predict the stimulus word from the frequency data, without access to speaker identities. Though optimizing stimulus word classification accuracy may lead us to maintain word differences that are not desired in a contour category abstraction, it will still lead us to minimize speaker differences. Because the stimuli of a single sandhi contour category have very similar contours and they are thus difficult to distinguish, the classification accuracies are numerically quite low. However, differences in accuracies for different parameterizations do reflect the relative merit of the parameterizations. The classification accuracy is the percentage of successes in leave-one-out cross-validation, so the standard error on the accuracy is calculated as the theoretical standard error of a binomial distribution's estimate of the percentage of successes.

4.3.4 Frequency scale and transform

First, combinations of three frequency scales and four scale transforms were considered. Linear, logarithmic and ERB scales were compared, and mean-norm, z-score and pointwise z-score transforms were compared with the baseline of no transform. The mean-norm and z-score transforms are the conventional normalization transforms described previously. Whereas the z-score treats the frequency samples as just a function of two variables, the speaker and a sample index, the pointwise z-score decomposes the sample index into a stimulus word index and a sampling time. With word index w and time points t, Equation 13 becomes:

$$f'_{iwt} = \frac{f_{iwt} - f_{ref}}{\Delta f} = \frac{f_{iwt} - f_{ref}}{f_f - f_0} \tag{14}$$

where i indexes the speakers as before. While the z-score transform takes reference frequency $f_{ref} = \frac{1}{WT}\sum_{wt}^{WT} f_{iwt}$, averaging over words and time points, the pointwise z-score takes $f_{ref} = \frac{1}{W}\sum_{w}^{W} f_{iwt}$, averaging over words only. The range factor (the standard deviation) is

Table 13: Between-speaker dispersions for four normalization transforms and three frequency scales.

	NONE	MEAN-NORM	Z-SCORE	POINT-Z
LINE	80.6 ± 1.2	45.7 ± 1.7	$\mathbf{42.6 \pm 2.1}$	42.7 ± 2.1
LOG	81.5 ± 1.2	47.4 ± 2.0	43.2 ± 2.1	43.3 ± 2.0
ERB	80.9 ± 1.1	46.4 ± 1.7	42.7 ± 2.0	42.8 ± 2.0

Table 14: LDA classification accuracy for three normalization transforms and three frequency scales.

	NONE	MEAN-NORM	Z-SCORE	POINT-Z
LINE	9.4 ± 0.7	12.6 ± 0.8	13.5 ± 0.9	14.5 ± 0.9
LOG	10.9 ± 0.8	11.1 ± 0.8	12.2 ± 0.8	13.8 ± 0.9
ERB	10.8 ± 0.8	12.2 ± 0.8	13.3 ± 0.9	$\mathbf{14.6 \pm 0.9}$

calculated in the same way, averaging over words only.

As seen in Table 13, the choice of frequency scale makes very little difference. The linear and ERB scales have nearly identical dispersions, regardless of frequency transform, and these dispersions are only slightly lower than the corresponding dispersions for the logarithmic scale. The differences among the scales are smaller than the standard error. All three transforms (mean-norm, z-score, and pointwise z-score) have large effects on the dispersion, reducing it by almost a factor of 2. The z-score transformed frequencies and pointwise z-score transformed frequencies have almost identical dispersion, and they have consistently lower dispersion than the mean-norm transformed frequencies. However, these differences are not much larger than the standard error.

The classification accuracies (Table 14) are qualitatively very similar. The scores suggest that there is very little difference between the frequency scales. The numerically highest accuracy is for the ERB scale, but considering the standard error, the linear and ERB scale accuracies are not different. Except in the case of untransformed frequencies, the logarithmic scale is about one standard error lower than the ERB scale. The accuracies furthermore indicate that all three transforms are improvements over the untransformed frequencies. The differences among the transforms are again of the same scale as the standard error, but they suggest that the pointwise z-score is best, followed by the z-score.

Compared to previous studies that published between-speaker dispersions, there is a large amount of residual variation among the speakers here. This variation includes some differences that might be attributable to unreliable automated F0 extraction in the speakers

Table 15: Dispersion of speakers' normalized contours w.r.t. across-speakers mean contours

Speaker	16	17	18	20	21
Variance	0.42	0.35	**0.77**	0.33	0.30
Speaker	04	19	22	23	24
Variance	0.36	0.36	0.43	**0.65**	0.41
Speaker	06	12	07	08	13
Variance	**0.54**	0.35	0.34	**0.57**	0.27

Table 16: Between-speaker dispersions for four normalization transforms and three frequency scales, after removing four high-dispersion speakers.

	NONE	MEAN-NORM	Z-SCORE	POINT-Z
LINE	70.8 ± 1.7	36.0 ± 2.4	34.3 ± 2.1	$\mathbf{33.8 \pm 2.1}$
LOG	71.2 ± 1.5	36.3 ± 1.9	34.5 ± 2.3	34.1 ± 2.5
ERB	70.8 ± 1.6	35.8 ± 2.2	34.2 ± 2.1	$\mathbf{33.8 \pm 2.2}$

with especially creaky voices, or attributable to lack of fluency among the younger speakers. For this reason, in choosing a parameterization, it was important to exclude speakers whose measured contours might be unreliable. The data was first normalized using the linear frequency scale and the pointwise z-score, because this combination is near optimal in each of the previous tables. Now that speakers' frequency range differences were normalized out, the speakers' average dispersions (Table 15) indicated which speakers' F0 tracks were particularly divergent from the mean. Speakers whose variances exceeded 0.5 were excluded: speakers 18, 23, 06 and 08.

After excluding those speakers, the dispersions (Table 16) and classification accuracies (Table 17) were recalculated for each combination of frequency scale and scale transform. The

Table 17: LDA classification accuracy for four normalization transforms and three frequency scales, after removing four high-dispersion speakers.

	NONE	MEAN-NORM	Z-SCORE	POINT-Z
LINE	10.7 ± 0.9	14.4 ± 1.0	15.0 ± 1.1	$\mathbf{17.2 \pm 1.1}$
LOG	10.4 ± 0.9	13.8 ± 1.0	13.9 ± 1.0	16.1 ± 1.1
ERB	11.3 ± 0.9	14.1 ± 1.0	14.5 ± 1.0	16.8 ± 1.1

resulting scores are qualitatively quite similar to before excluding the high variance speakers, just with lower dispersions and higher classification accuracies, as would be expected. The linear and ERB scales again have very comparable scores and are better than the logarithmic scale, and the pointwise z-score transform is better than the other two transforms. On the basis of the measured between-speaker dispersions and classification accuracies, the linear scale and the pointwise z-score transform were selected for normalizing speaker differences.

4.3.5 Summarized representations

The final step of parameterization is choosing a method of contour summarization.[8] Phonological descriptions of tone sandhi often reduce tone specifications to just one target with three levels (high, medium and low) per syllable, while machine learning and articulation models often use regression-based summaries of tone contours. Because the dimensionality of Jinhua tone contour categories was unknown, this step compared a variety of summary representations that had one, two or three scalar parameters per tone contour. Since summarization reduces the dimensionality of the representation, the between-speakers dispersion is not a useful measure of representation quality, and so this step depends on LDA classification accuracy for evaluating the representation.

Six candidate methods were considered. The first three methods are one (PTS1), two (PTS2) and three (PTS3) parameter point-target representations, comparable to Chao tone numbers, but on a continuous scale rather than limited to five levels. The sample points in each tone domain were divided into one, two or three approximately equal-size bins. With two tone domains per sandhi contour, this results in a total of two, four or six bins per sandhi contour. The mean normalized frequency of each bin become parameters in the model. The next two methods are two (POLY2) and three (POLY3) parameter polynomial regressions. A regression was performed on each tone domain separately, with the time units recentered so that the model intercept is the predicted frequency in the middle of the tone domain. Regressions were performed with two parameter (linear) and three parameter (quadratic) polynomials, for a total of four or six parameters per sandhi contour. The last method is a three parameter piecewise linear regression (PIECE). The model consists of an intercept confined to the center of the tone domain plus slopes for the first and second halves of the tone domain, for a total of six parameters per sandhi contour.

The sampling points were divided into the bins in Table 18 for the point-target representations, and they were fitted as two separate domains for the regression representations. The

[8] Exclusion of marginal sampling points was also considered, but no benefit was found for this data, taken together. When parameterizing the data of a low variance subset of speakers, like the subset of older urban speakers, exclusion of the highest dispersion sampling points improves the classification accuracy. But for more varied data, there is no benefit.

Table 18: Sample point bins in the point-target representations

model	1st syll	2nd syll
PTS1	1–10	11–20
PTS2	1–5; 6–10	11–15; 16–20
PTS3	1–3; 4–7; 8–10	11–13; 14–17; 18–20

Table 19: LDA accuracies for six methods of dimension reduction

PTS1	PTS2	PTS3	POLY2	POLY3	PIECE
7.1 ± 0.8	12.6 ± 1.0	16.0 ± 1.1	12.0 ± 1.0	$\mathbf{17.5 \pm 1.1}$	17.3 ± 1.1

classification accuracy of each representation is shown in Table 19. The differences among the accuracies is largely explainable by the number of parameters in the model. The single point representation (PTS1), with one parameter per tone domain, provides too little information for successful classification. The representations with two parameters per tone domain (PTS2 and POLY2) perform better, but still worse than the original 20 sampling point representation (17.2%, Table 17). The representations with three parameters per tone domain perform comparably to the 20 sampling point representation, with the regression models (POLY3 and PIECE) performing better than the point-target model (PTS3). On the basis of these accuracy scores, the quadratic regression representation (POLY3) was selected as the contour summarization model.

4.4 Contour systems

4.4.1 Speaker groups

The full variation in speakers' tone contour systems was analyzed visually by representing the similarities of the contour systems in a multidimensional scaling (MDS) plot. For each speaker, the contour parameters for all stimulus words were together treated as dimensions in a euclidean space, with parameters z-score normalized to remove scale differences between the dimensions. So each speaker's sandhi system is represented by a 648 dimensional vector (3 parameters per tone domain $\times$ 2 tone domains per word $\times$ 108 words). Distances between speakers were computed, and the distances were reproduced in a reduced space.

The distances between speakers' contour systems are displayed in Fig. 5 in terms of the first three principle components of their multi-dimensional scaling (MDS) space. For each speaker, the contour parameters for all words are together treated as a location in a euclidean space, with parameters z-score normalized to remove scale differences between the

dimensions. Distances between speakers are computed, and the distances are reproduced in a reduced space. The three dimensions shown in the figure together represent 38% of the total variance.

The speakers' social variables are reflected in the similarities of their contour systems. The older speakers (black border) are separable from the younger speakers (gray border) in dimension 1. In addition, the speakers from Zhuma village (white interior) are nearly separable from the speakers from the urban area (gray interior) in dimension 3. But there is no obvious effect of gender (squares vs circles). Dimension 2 singles out speaker 23 as different from the others, for unclear reasons.

However, social networks among the speakers are not reflected in the similarities, with one exception. Speakers 04 and 19 are cousins of the same clan, and their contour systems are indeed similar. But speakers 04, 07 and 08 are coworkers, and their contour systems are not similar. Speakers 17 and 18 are lifelong neighbors, and their neighbor is the grandfather of speakers 04 and 19. These relationships are not reflected in the similarities.

The large degree of speaker variation in the data is difficult to interpret, especially when each speaker's sandhi system is already quite complicated. To abstract away from the full speaker variation, and to compensate for the large number of words that were missing contours from some speakers, the speakers were separated into three groups. Because speaker 23 is apparently quite different from the other speakers, his data was excluded from the rest of the analysis. The remaining speakers are taken to consist of three distinct groups: the older urban speakers (16, 17, 18, 20, and 21), the younger urban speakers (04, 19, 22, and 24), and the speakers from Zhuma village (06, 07, 08, 12, and 13). The speakers in the Zhuma group are not as similar to each other as the speakers in the urban groups are, but because the Zhuma speakers collectively represent one component of the variation in the data, they are treated as a single group.

4.4.2 Contour categories

A clustering algorithm was used to identify the categories of sandhi contours. For each speaker group, an average contour was calculated for every stimulus word, by taking the mean of each of the dimensions of the contour representation, among those speakers who were not missing that item. Each word was then represented as the vector of the contour parameters from all three groups. A mixture of Gaussians clustering routine[9] clustered the words, optimizing the number of clusters according to the Bayes Information Criterion (BIC). The software's default clustering initialization was used, except that cluster counts 1 through 15 were considered, rather than the default of 1 though 10.

[9]Mclust R library

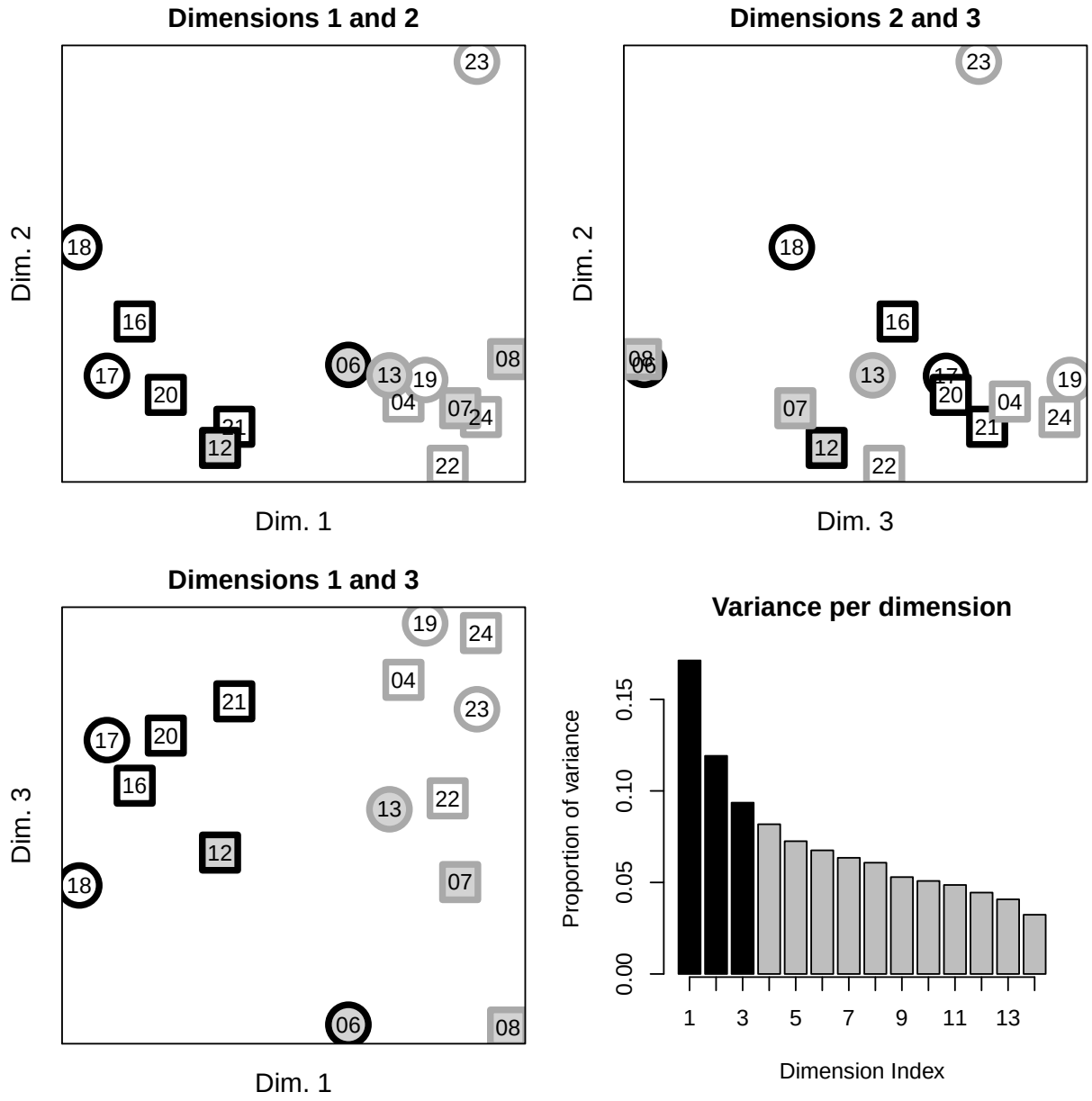


Figure 5: Similarities of speakers' contour systems, represented in an MDS space. Marker shapes indicates men (circles) and women (squares), marker borders indicate older (black) and younger (gray) generations, and marker fill indicates urban Jinhua (white) and Zhuma village (gray) locations. Dimension 1 is associated with speaker age, and Dimension 3 is associated with speaker location.

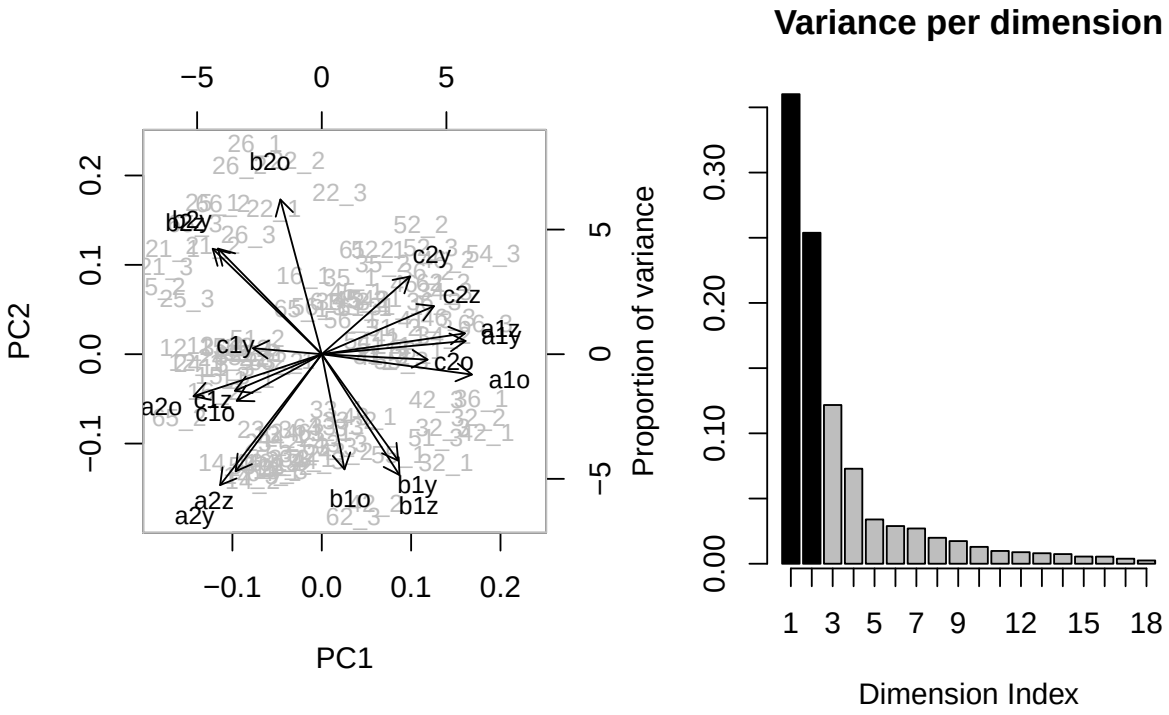


Figure 6: Principle components analysis of the speaker groups' representations of the words' mean contours. The 18 dimensions of the original representations are shown projected into the plane of first two principle components. The speaker groups ('o', 'y' and 'z' vectors) are correlated, with the original representation of the older speakers contours ('o' vectors) rotated clockwise compared to the younger ('y') and Zhuma ('z') speakers. The intercept heights ('a' vectors), slopes ('b' vectors) and curvatures ('c' vectors) are also related, such that the first two principle components contain 61% of the variation.

The clustering algorithm was supplied with vectors of 18 dimensions (3 degree polynomial $\times$ 2 tone domains $\times$ 3 speaker groups), but the 18 dimensions of the original representation are nearly reducible to two, because of high collinearity in the dimensions (Fig. 6). The first two dimensions contain 61% of the variance, but there is non-negligible variance in the third and fourth dimensions as well. The first four dimensions together contain 81% of the variance.

The mean contours of each of the groups — the older speakers ('o' vectors), the younger speakers ('y' vectors) and the Zhuma village speakers ('z' vectors) — are correlated, with the representation of the older speakers' contours rotated slightly clockwise compared to the other two groups. Additionally, the slope on the first tone domain (b1o, b1y, b1z) is

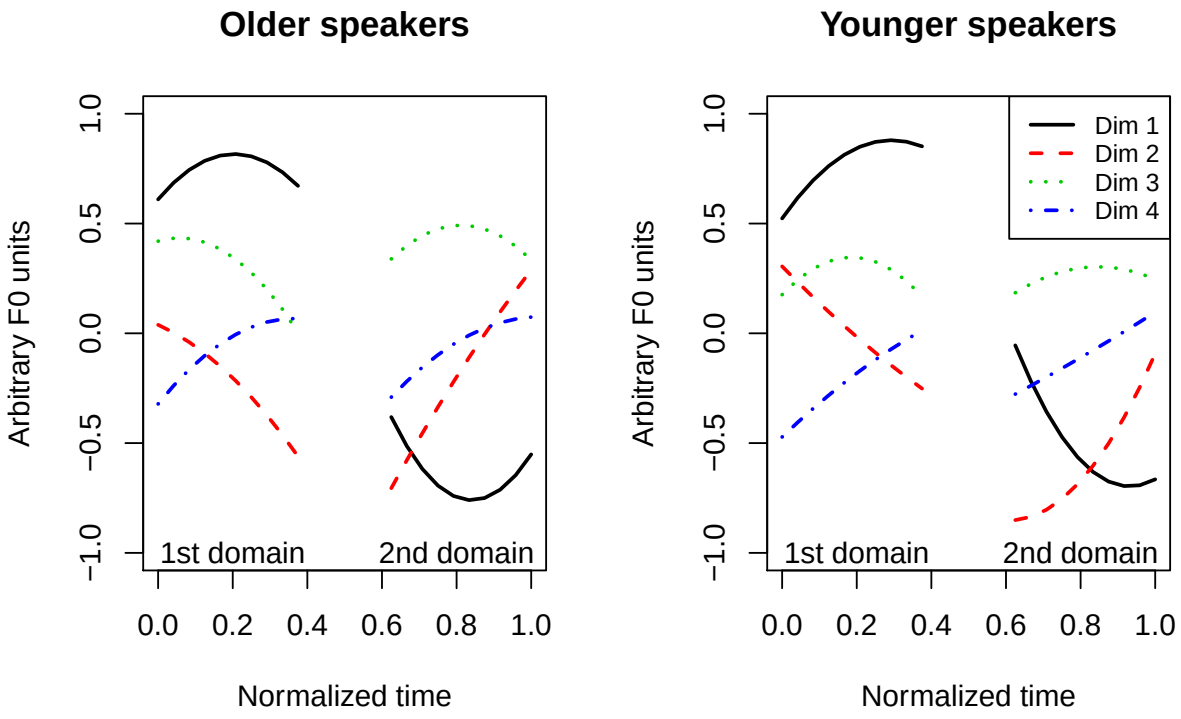


Figure 7: Archetype tone contours for each of the first four principle components. For the older speakers (left), dimension 1 approximately corresponds to the pitch height difference of the first and second tone domains, and dimension 2 corresponds to fall-rise vs rise-fall sandhi contours. Dimension 3 is roughly overall height of the two domains, and dimension 4 is the amount that the two domains rise equally (or fall equally). The meaning of the dimensions are slightly different for the younger speakers (right), and archetype contours of the Zhuma speakers are quite similar to the younger speakers, and are not shown here.

negatively correlated with the slope on the second tone domain (b2o, b2y, b2z), and intercepts on the first tone domain (a1o, a1y, a1z) are negatively correlated with the intercepts on the second tone domain (a2o, a2y, a2z). Finally, the curvatures (c1o, c1y, c1z, c2o, c2y, c2z) are negatively correlated with the corresponding intercepts (a1o, a1y, a1z, a2o, a2y, a2z), meaning that positive curvatures (concave curves) are associated with low intercepts on that tone domain (and with high intercepts on the neighboring tone domain). The archetypes of each of the four dimensions are shown in Fig. 7.

The clustering algorithm selected 11 clusters of word contours. The word clusters are represented in the first two principle component dimensions in Fig. 8. The colored letters indicate cluster membership, and the numbers are word labels, the first two numerals of which

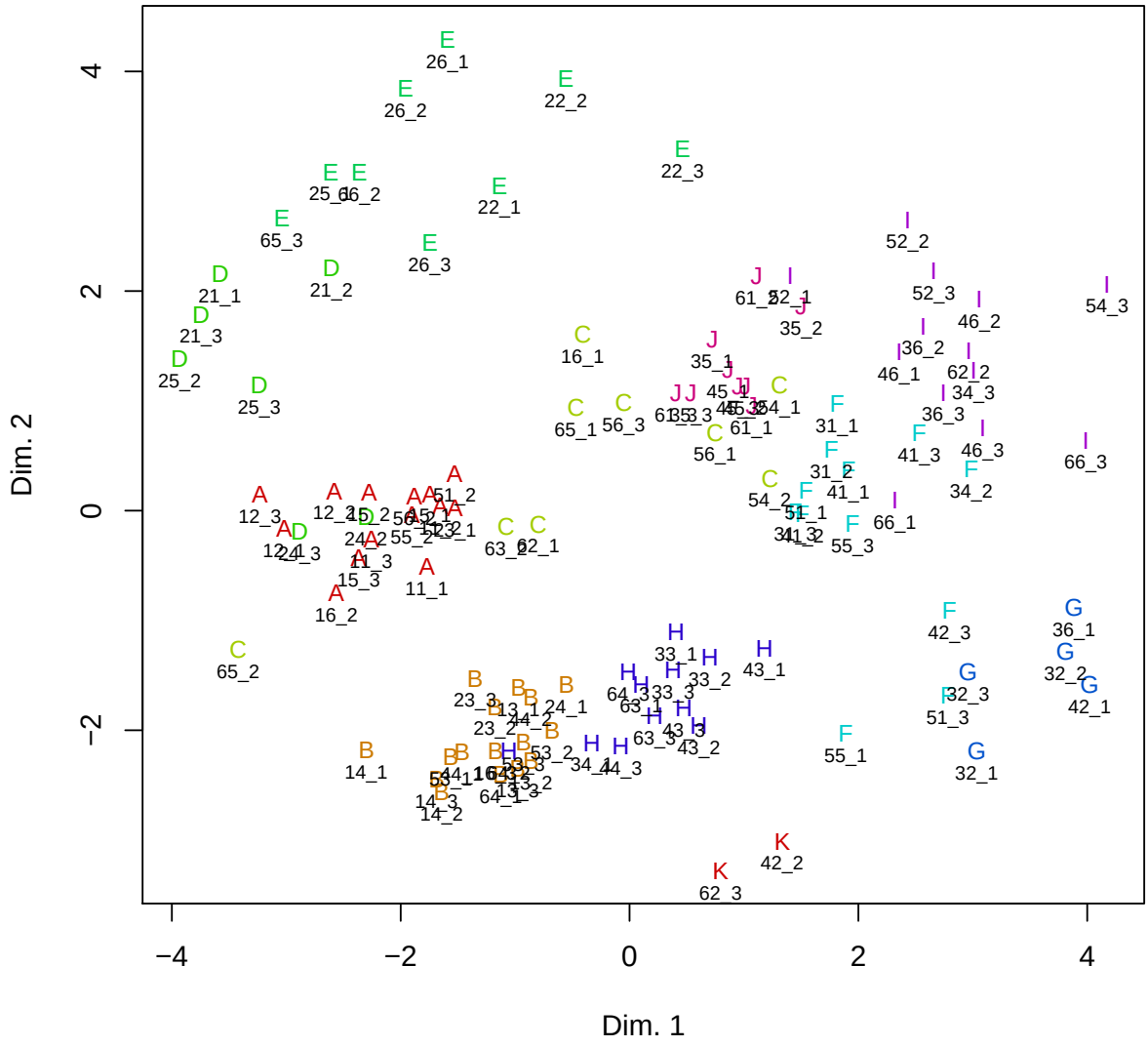


Figure 8: Plot of stimuli contour similarity, in the first two principle components. Word clusters are indicated by the colored letters. The first two digits of the word labels refer to the tone categories, numbered sequentially 1 through 6.

indicate the isolation tone categories. The tone categories are numbered sequentially, with odd numbers referring to H register tones, and the even numbers referring to L register tones. The unevenness of the word locations in this space suggest that there really are categories, and that 11 categories is not an unreasonable number. Words of the same tone combination frequently lie in close proximity and within the same cluster (e.g. 33_1, 33_2 and 33_3 in H). However, cluster C has very little coherence in these two dimensions, overlapping with clusters A and J, though cluster C may have more coherence in the other dimensions.

4.4.3 Group contour systems

In each speaker group, a mean contour was then calculated for all word clusters. Contours were averaged across speakers within the speaker group and across words within the cluster, using normalized but non-summarized contour frequencies, in order to visualize the finer details of contour shapes. In addition, standard deviations from the mean contour were calculated, to identify clusters that had large speaker or word variability.

For convenience of comparison with previous work, a two-numeral Chao tone representation was created for each tone domain of the contours. The 2-degree polynomial regression representation described in §4.3.1 was used to predict intercepts at the beginnings and ends of the tone domains. The frequency range of the predicted intercepts was divided into 5 equal segments, and intercepts within each segment were labeled with the corresponding numeral.

The following three figures (Figs. 9, 10, 11) show the mean contours of the 11 categories for each of the three speaker groups. Across all three groups, the mean contours are quite similar. The Chao tone numbers assigned to a contour sometimes differ by one unit between speaker groups, such as cluster B, assigned $T^{33}T^{54}$ for the older urban speakers, $T^{33}T^{55}$ for the younger urban speakers, and $T^{23}T^{55}$ for the Zhuma speakers. But even in these cases, the contour is usually still of the same general shape. There are, however, a few cases of differences that are larger. In cluster I, the older speakers have a low rising tone T^{13} on the second syllable, where the younger speakers have a low level tone T^{22}. In cluster J, the older speakers have a high clearly falling tone T^{54} where the younger speakers have a high level tone T^{55}. The Zhuma speakers have a low slightly falling tone T^{21} in cluster I, much like the urban young speakers, but have a high falling tone T^{54} in cluster J, like the urban older speakers. In addition, Zhuma speakers have high rising tones on the first syllables of cluster F (T^{45}) and cluster H (T^{35}) where the corresponding tones of both urban groups are quite level (T^{55} and T^{44}).

For all three speaker groups, cluster C has large standard deviations, indicating that there are large differences among the words in cluster C. Some of the other clusters have

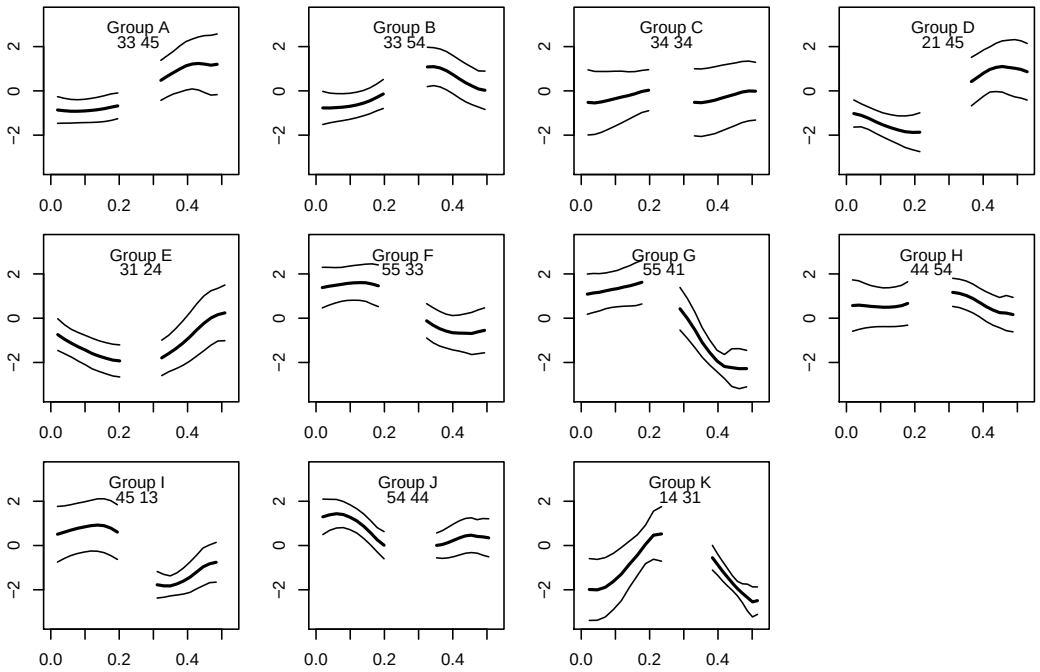


Figure 9: Older speakers' mean contours for each of the contour clusters. The thin lines show standard deviations about the mean. The numerals below the group label are the Chao tone numbers assigned to the contour.

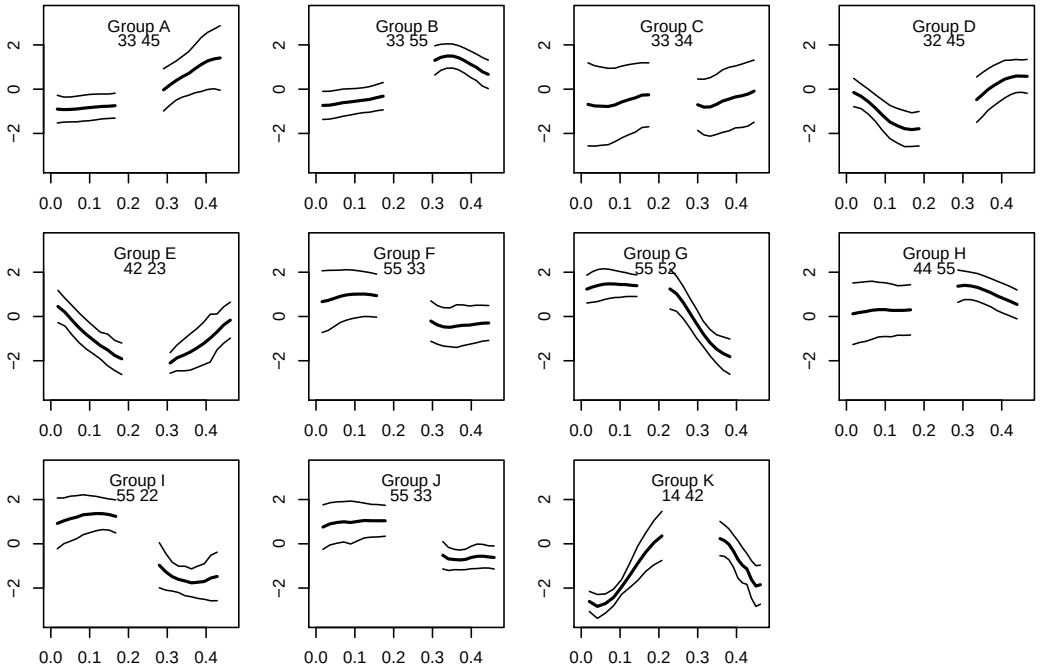


Figure 10: Younger speakers' mean contours for each of the contour clusters. The thin lines show standard deviations about the mean. The numerals below the group label are the Chao tone numbers assigned to the contour.

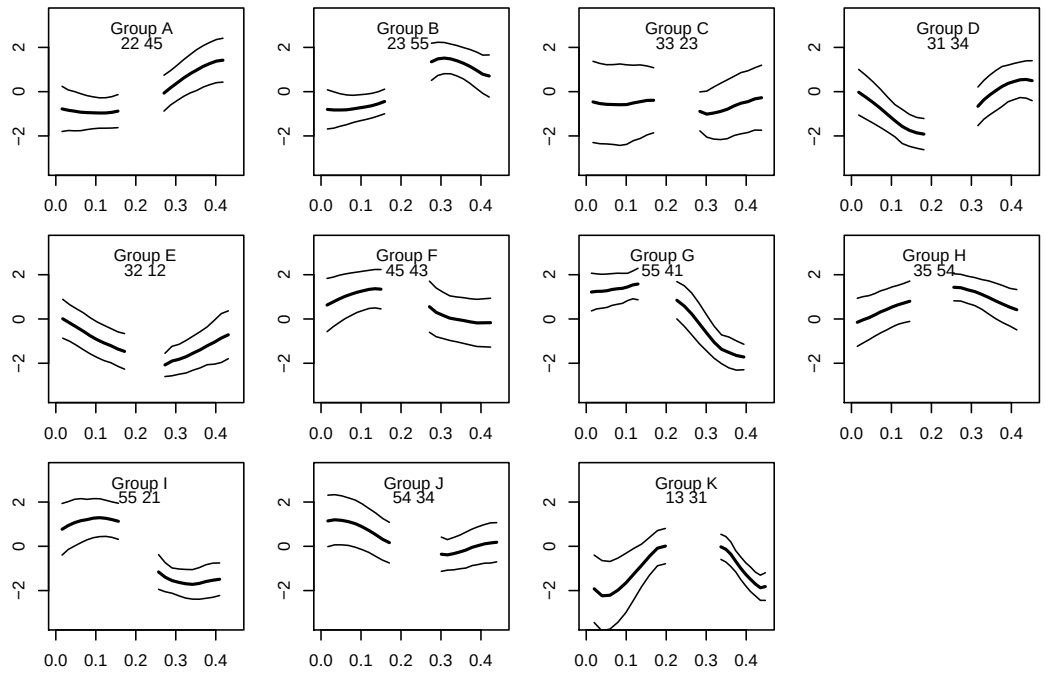


Figure 11: Zhuma speakers' mean contours for each of the contour clusters. The thin lines show standard deviations about the mean. The numerals below the group label are the Chao tone numbers assigned to the contour.

large standard deviations on one of the tone domains, but no cluster has such large standard deviations in all three speaker groups.

The mean contours are arranged in a lattice according to their tone combination, for each of the speaker groups separately (Figs. 12, 13, 14).

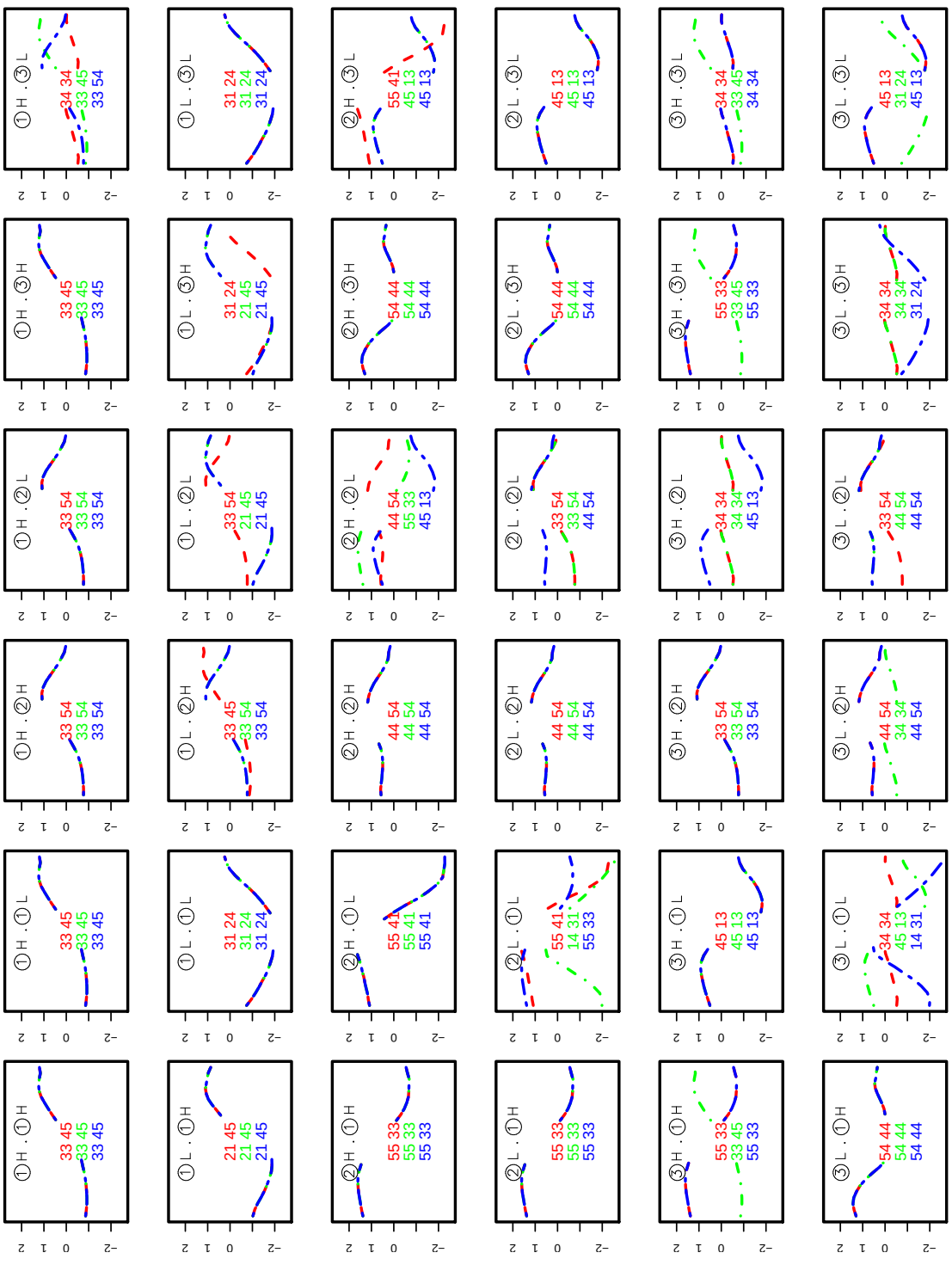


Figure 12: Older speakers' mean contours as a sandhi table. Rows have the same first syllable tone and columns the same second syllable tone. The three contours in each cell are the contours of the three stimuli for that combination of tones.

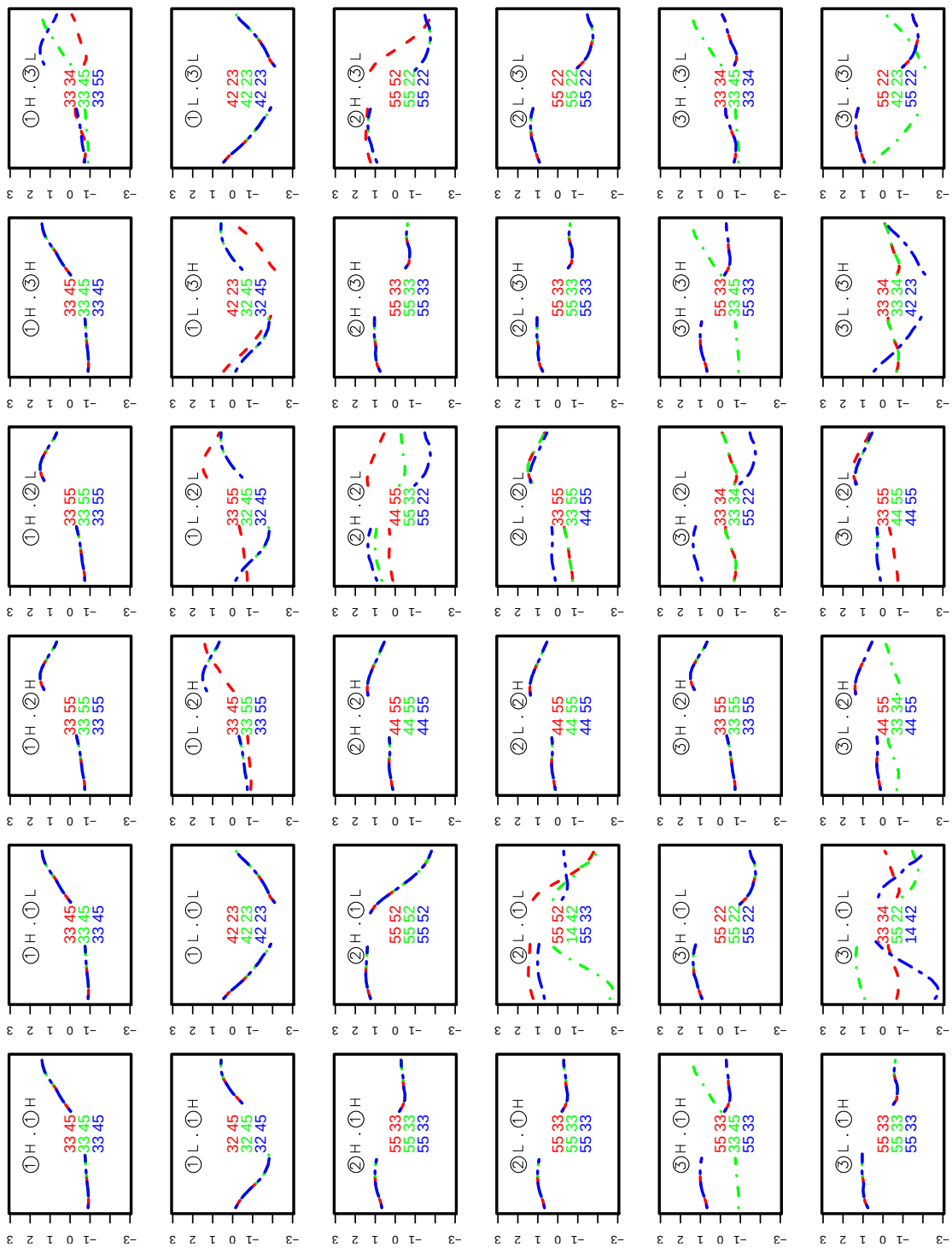


Figure 13: Younger speakers' mean contours as a sandhi table. Rows have the same first syllable tone and columns the same second syllable tone. The three contours in each cell are the contours of the three stimuli for that combination of tones.

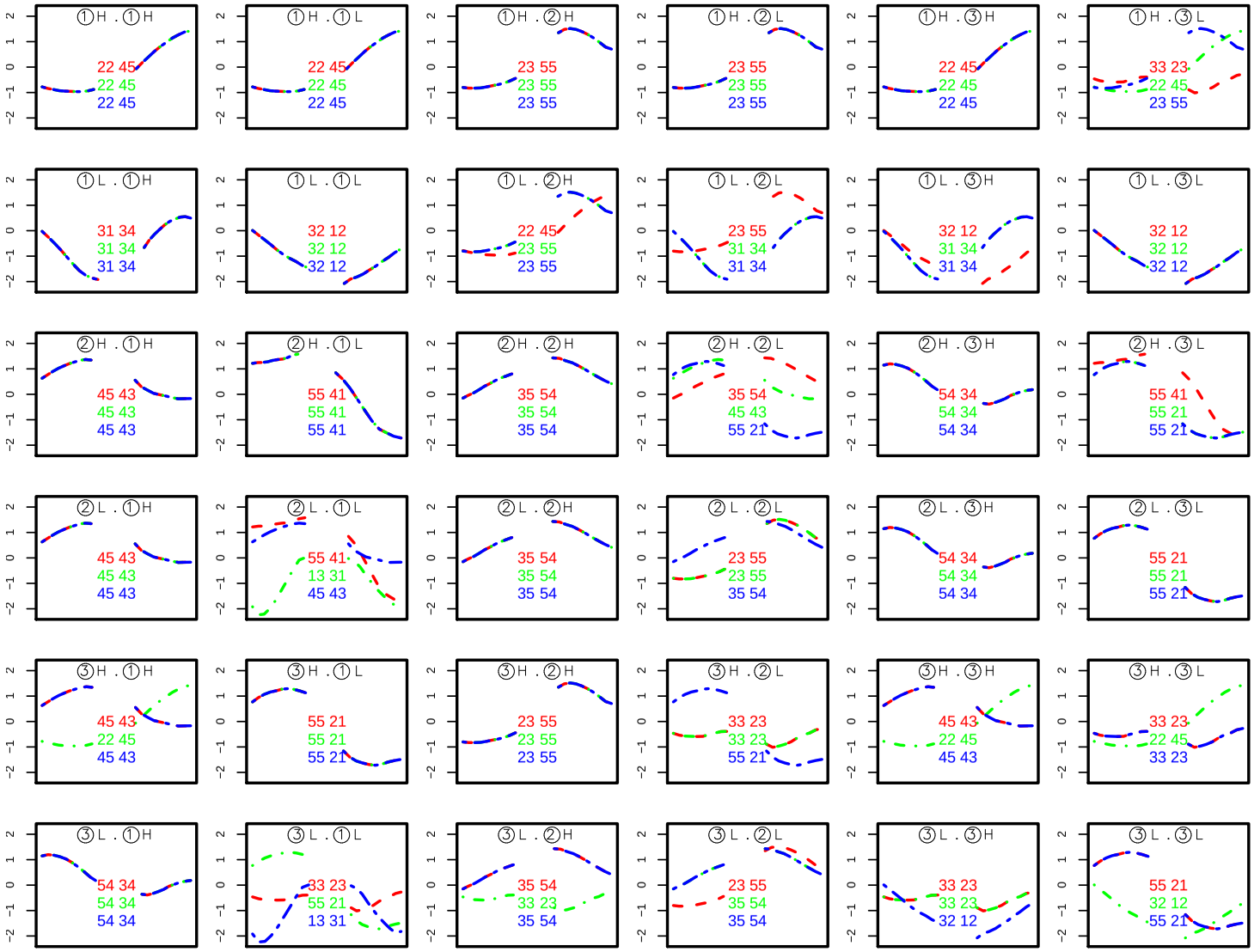


Figure 14: Zhuma speakers' mean contours as a sandhi table. Rows have the same first syllable tone and columns the same second syllable tone. The three contours in each cell are the contours of the three stimuli for that combination of tones.

5 Discussion

This study found large systematic differences between two generations of speakers in Jinhua, and less consistent differences between speakers from urban Jinhua and speakers from Zhuma village. These varieties of Jinhua dialect are distinguished here by small differences in the shapes of the tone contours associated with disyllabic words. But the observed tone contours raise a couple issues and leave a couple others unresolved.

5.1 Dimensionality of contours

Contour summarization needed three parameters per tone domain to obtain high classification accuracies, but the variation among contours is nearly reducible to just two parameters for both tone domains, because the tone domains are strongly coupled. This coupling highlights the effect of phonetic and phonological constraints. Tonal coarticulation is manifest in the curvatures. Convex curves tend to occur on high tones and they tend to occur next to low tones, while concave curves tend to occur on low tones and next to high tones. These are the effects of assimilating the pitch on the edges of the tone domain towards the pitch of the neighboring tone domain. Similarly, the opposition of neighboring slopes and intercepts is evidence of an obligatory contour principle (OCP) effect (S. Myers, 1997). The fact that slopes of the two tone domains are inversely correlated indicates that two rising tones together and two falling tones together are uncommon compared to the juxtaposition of rising and falling tones. The fact that the intercepts are also inversely correlated indicates that two high tones together and two low tones together are not as common as the juxtaposition of high and low tones. Note that neither of these correlations indicates anything particular about middle range values. Two level tones or two mid tones are not underrepresented. Also, if these tendencies are understood to reflect the preferences of an OCP constraint, the constraint is occasionally violated in this data. For the older speakers, the cluster I contour has two rising tones ($T^{45}T^{13}$). And for all speakers, the cluster H contour is high on both syllables (e.g. $T^{44}T^{55}$).

The disyllabic contours nearly fit within a two dimensional space and almost completely fit within a four dimensional space. The four dimensions are completely distinguishable with reference to just the intercepts and slopes, and yet the intercepts and slopes were not sufficient to obtain classification accuracies approaching the original 20 point F0 representation. So it appears that the curvature is informative primarily as a semi-redundant indicator of the pitch height within the tonal context. This suggests that a phonological representation that has just two parameters per tone domain should be sufficient for syllables that are not in isolation or phrase final. There is no evidence, however, for discrete levels of those parameters.

5.2 Differences among words

The sandhi lattice figures have many cells where the words which were chosen to be of the same tone combination have different contours. A good deal of the apparent variation in the lattice figures is reducible to failures of the clustering method or to the known isolation tone merger (obstruent-onset ②$_{\mathrm{L}}$ → ③$_{\mathrm{L}}$), but some variation simply cannot be accounted for.

One source of problems in the clustering method is that when there were categorical speaker differences for a particular item, the item is not located near similar items. Many of these words ended up in cluster C, the cluster whose words varied considerably. Inspecting individual speakers' contours for the words in cluster C helps clear up the situation. For example, with item 63_2 (tuɤ$_{③\mathrm{L}}$siɑu$_{②\mathrm{H}}$ 'size'), the urban speakers used a class H contour (e.g. T^{44}T^{54}) while the speakers from Zhuma used a class D contour (T^{31}T^{34}). With item 56_3 (siŋ$_{③\mathrm{H}}$miŋ$_{③\mathrm{L}}$ 'life'), the older speakers used a class A contour (e.g. T^{33}T^{45}) while the younger speakers used a class I contour (e.g. T^{55}T^{22}). Word 62_1 (ɑ$_{③\mathrm{L}}$pɤ$_{①\mathrm{L}}$ 'maternal grandma') has the same speaker split, with older speakers using class I (e.g. T^{45}T^{13}), and younger speakers using a contour T^{12}T^{45}. This contour T^{12}T^{45} was not identified by the clustering algorithm as a separate contour cluster, but it is used by all speakers for item 65_2 (uɑ$_{③\mathrm{L}}$fi$_{③\mathrm{H}}$ 'phone charge'). With item 54_2 (tɛ$_{③\mathrm{H}}$siɑŋ$_{②\mathrm{L}}$ 'partner'), about half the speakers used a class B contour (e.g. T^{33}T^{54}), and half used a class I contour (e.g. T^{45}T^{13}), with no apparent pattern to which speakers chose each contour. In the case of item 65_1 (tia$_{③\mathrm{L}}$pɑu$_{③\mathrm{H}}$ 'telegram'), the younger urban speakers used a class F contour (T^{55}T^{33}), and the other two groups had no consensus. Word 25_1 (toŋ$_{②\mathrm{L}}$pɛ$_{③\mathrm{H}}$ 'of the same generation') also has a speaker split, but it was assigned to class E (T^{31}T^{24}). Younger speakers do indeed use this contour, but older speakers use contour D (T^{21}T^{45}), like the rest of the ②$_{\mathrm{L}}$③$_{\mathrm{H}}$ stimuli. Item 42_3 (ie$_{②\mathrm{L}}$mɑu$_{①\mathrm{L}}$ 'wild cat') is pronounced as contour G (T^{55}T^{41}, like ②$_{\mathrm{H}}$①$_{\mathrm{L}}$) by speakers from Zhuma, but speakers from the urban area use contour F (T^{55}T^{33}, like ②①$_{\mathrm{H}}$). While the morpheme [mɑu] was historically tone ①$_{\mathrm{L}}$, and the Jinhua dialect dictionary records it as such, in Mandarin it has switched to ①$_{\mathrm{H}}$, and it may be that Mandarin has influenced Jinhua speakers to make the same adjustment in Wu. Straightening out these cases of categorical speaker differences eliminates some of the apparent word-level variation.

In many of the cells in the row or column for ②$_{\mathrm{L}}$, where there is variation among stimuli, one or more of the contours is associated with words that have obstruent onsets in the second syllable. In the ②$_{\mathrm{H}}$②$_{\mathrm{L}}$ cell, both 34_2 (hɑu$_{②\mathrm{H}}$siaŋ$_{②\mathrm{L}}$ 'appear') and 34_3 (kɛ$_{②\mathrm{H}}$tsau$_{②\mathrm{L}}$ 'reform') have obstruent onsets and thus could have changed to ②$_{\mathrm{H}}$③$_{\mathrm{L}}$. Inspecting the individual speakers' contours reveals that 34_2 has large speaker differences, whereas 34_3 is consistently contour I, like two of the stimuli from ②$_{\mathrm{H}}$③$_{\mathrm{L}}$. In the cell for ②$_{\mathrm{L}}$①$_{\mathrm{L}}$, item 42_2 (siɑŋ$_{②\mathrm{L}}$tɕi$_{①\mathrm{L}}$ 'chess') has an obstruent onset first syllable, so it makes sense for its contour

to be contour K like the third item of ③$_{L}$①$_{L}$. Finally, all three of the stimuli for ③$_{H}$②$_{L}$ have obstruent onsets in their second syllable. Moving these words to the ③$_{H}$③$_{L}$ cell creates more variability there, but of the six words now associated with ③$_{H}$③$_{L}$, four of them are of cluster C. Two of these words, 54_2 and 56_2, have already been discussed above. Word 54_1 (pɑu$_{③H}$sɤ$_{②L}$ 'newspaper office'), which was assigned to group C, belongs better in cluster J ($T^{54}T^{44}$). And word 56_1 (tsɑ$_{③H}$tan$_{③L}$ 'explosion') appears to belong to contour I, but was miscategorized because of several faulty contours that the creaky pitch jumps filter didn't catch.

Notably, not all ②$_{L}$ words have changed to the corresponding ③$_{L}$ contours. While the other contours in its cell look like ③$_{L}$①$_{L}$ contours, word 42_1 (tu$_{②L}$pi$_{①L}$ 'belly') is the same as ②$_{H}$①$_{L}$ contours, even though it does also start with an obstruent onset. Word 16_3 (kuɤ$_{①H}$ti$_{②L}$ 'brother') was chosen to represent ①$_{H}$③$_{L}$ because it was mistakenly thought the second syllable was historically ③$_{L}$, as it is currently in isolation. But it is in fact historically from the ②$_{L}$ tone class, and the contour it bears in this data (cluster B, e.g. $T^{33}T^{54}$) indicates that at least in this word it remains in ②. These two words are quite vernacular in use, whereas the words that are observed to show the ②$_{L}$ → ③$_{L}$ merger tend to be literary or governmental vocabulary.

Taking all these cases into account, we can create cleaned up sandhi tables with less word-level variation (Tables 20, 21, 22). However, unexplained variation remains. For example, for the older speakers, cells ③$_{H}$①$_{H}$ and ③$_{H}$③$_{H}$ (with the same two contours) and four out of six cells in the ③$_{L}$ row have lexical variation that cannot be attributed to any known morphosyntactic, historical, or domain of use distinction.

5.3 Differences among speakers

The similarities of speakers' contour systems reflected their age and location but not their social networks. However, the seeming lack of influence of social networks can be explained, in each case uninterestingly. Speakers 07, 08 and 04 were coworkers, but they had not been coworkers for very long, and in the workplace they would speak Mandarin not Wu, so they are not exposed to each others' pronunciations of Jinhua tones. Speakers 17 and 18 had been neighbors for most of their lives, but Speaker 18 had unusually creaky voice. He had the highest number of words that the F0 extraction algorithm failed to find a smooth contour for. It is quite possible that a number of smooth contours for Speaker 18 still had pitch measurement errors. The grandfather of Speakers 04 and 19 was also their neighbor, so we might expect Speakers 04 and 19 to be close to Speaker 17, but considering that the generation difference is the most robust speaker difference in the contour systems, the

Table 20: Sandhi system of the older speakers. The first syllable tone categories are along the left margin and the second syllable tone categories are along the top margin. The tones are marked according to contemporary Jinhua categories. The body of the table contains the two-syllable tone contours, and the 'row/col trend' is a generalization of the first syllable contours in the rows and of the second syllable contours in the columns. The tone changes are highlighted.

				2nd Syllable					
			JH class	①		②		③	
			JH reg.	**H**	**L**	**H**	**L**	**H**	**L**
	JH class	**JH reg.**	**row/col trend**	**33/45**	**41/13**	**54**	**54**	**45**	**13**
1st Syllable	①	**H**	**33**	33.45	33.45	33.54	33.54	33.45	33.13 33.45
		L	**21**	21.45	31.24	33.45 33.54	21.45 33.54	21.45	31.24
	②	**H**	**55**	55.33	55.41	44.54	44.54	54.44	45.13
		L	**55**	55.33	55.41	44.54	44.54 33.54	54.44	45.13
	③	**H**	**33**	33.45 55.33	45.13	33.54	33.54	33.45 55.33	33.45 45.13
		L	**31/45**	54.44	14.31 45.13	44.54	44.54 33.54	31.24 12.45	31.24 45.13

Table 21: Sandhi system of the younger speakers. The first syllable tone categories are along the left margin and the second syllable tone categories are along the top margin. The tones are marked according to contemporary Jinhua categories. The body of the table contains the two-syllable tone contours, and the 'row/col trend' is a generalization of the first syllable contours in the rows and of the second syllable contours in the columns. The tone changes are highlighted.

				2nd Syllable					
			JH class	①		②		③	
			JH reg.	**H**	**L**	**H**	**L**	**H**	**L**
	JH class	**JH reg.**	**row/col trend**	**33/45**	**52/22**	**55**	**55**	**45/33**	**22**
1st Syllable	①	**H**	**33**	33.45	33.45	33.55	33.55	33.45	33.12 33.45
		L	**32**	32.45	42.23	32.45 33.55	32.45 33.55	32.45 42.23	42.23
	②	**H**	**55**	55.33	55.52	44.55	44.55	55.33	55.22
		L	**55**	55.33	55.52	44.55	44.55 33.55	55.33	55.22
	③	**H**	**33**	33.45 55.33	55.22	33.55	33.55	33.45 55.33	33.45 55.22
		L	**55**	55.33	55.22 14.42 12.45	44.55 12.43	44.55 33.55	55.33 42.23 12.45	55.22 42.23

Table 22: Sandhi system of the Zhuma speakers. The first syllable tone categories are along the left margin and the second syllable tone categories are along the top margin. The tones are marked according to contemporary Jinhua categories. The body of the table contains the two-syllable tone contours, and the 'row/col trend' is a generalization of the first syllable contours in the rows and of the second syllable contours in the columns. The tone changes are highlighted.

				2nd Syllable					
			JH class	①		②		③	
			JH reg.	**H**	**L**	**H**	**L**	**H**	**L**
	JH class	**JH reg.**	**row/col trend**	**43**	**41/21**	**55/54**	**55/54**	**45/34**	**21**
1st Syllable	①	**H**	**22**	22.45	22.45	23.55	23.55	22.45	33.13 22.45
		L	**32/31**	31.34	32.12	22.45 23.55	31.34 23.55	31.34 32.12	32.12
	②	**H**	**55/35**	45.43	55.41	35.54	35.54	54.34	55.21
		L	**55/35**	45.43	55.41	35.54	35.54 23.55	54.34	55.21
	③	**H**	**22**	22.45 45.43	55.21	23.55	23.55	22.45 45.43	22.45 55.21
		L	**55**	55.34	55.21 14.42 12.45	35.54 31.34	35.54 23.55	32.12 12.45	55.21 32.12

distance between them is understandable.

The generational difference represented by dimension 1 of the speaker similarities (Fig. 5) appears as the rotation of the contour representation vectors in the word contours PCA (Fig. 6). Inspection of the dimension archetype contours (Fig. 7) shows that dimensions 3 and 4 are pretty similar for the older and younger speakers, but that dimensions 1 and 2 are quite different. The source of this difference in the sandhi contours is a systematic pattern of tonal change from older to younger generations. The older generations's falling tones (like the first tone domain of the dimension 2 archetype) tend to be higher for the younger generation. The older generation's rising tones (like the second tone domain of the dimension 2 archetype) are lower for the younger generation. The older generation's high tones (dimension 1 first tone domain) become rising tones, and the older generation's low tones (dimension 1 second tone domain) become falling tones. These patterns can be observed in the individual contours of Fig. 9 and Fig. 10, and the generalization is simply more pronounced in the dimension archetypes. This pattern suggests that the younger generation perceives the contour as aligning with a slightly later portion of the syllable than the older generation does, and extrapolates the new initial portion of the contour on the basis of the rest of the contour and coarticulation with the previous syllable's tone. If this is a more widespread pattern of coarticulation reanalyzed as part of the sandhi, this could help explain how the sandhi of Jinhua became so complicated. However, there is no obvious explanation for why such active systematic tonal evolution would be observed here in Jinhua but has not been reported for better studied dialects.

The words that have categorical differences among speaker groups suggest some factors that are involved in word-level variation. As seen for the case of obstruent-onset MC ②$_{\mathrm{L}}$ merging either with ②$_{\mathrm{H}}$ or ③$_{\mathrm{L}}$, literary influence does cause variation among words of the same tone combination. It is possible that something of this kind is responsible for the variation observed elsewhere in the tables. Rare words sometimes have little consensus among speakers. For example, item 65_1 (tia$_{③\mathrm{L}}$pɑu$_{③\mathrm{H}}$ 'telegram') had very little agreement among speakers. However, recent common neologisms seem to follow usual patterns. For example, item 65_2 (uɑ$_{③\mathrm{L}}$fi$_{③\mathrm{H}}$ 'phone charge') had broad agreement, as did recent borrowings from Mandarin like item 11_3 (fitɕi 'airplane') and item 44_3 (mei ɳy 'pretty woman').

5.4 Differences among studies

Despite the fact that the methodology of this study was based on that of You and Yang (2001), the resulting sandhi tables more closely resemble that of Cao (1996) (Table 8) or Qian (1992) (Table 7) than that of You and Yang (Table 10). While You and Yang show

almost no variation within the cells of the table, many of the cells in the contour lattice plots, and even in Table 20, have multiple contour categories. In particular, the rows and columns for ③ tones and the column for tone ②$_{\text{L}}$ have frequent differences among stimulus words. These are also the areas where the most variation is found in the reports of Cao and Qian. In addition, the near mergers of certain rows and columns is not like You and Yang's. Like Cao's table, and unlike You and Yang's table, the sandhi figures of this study show similarities between the ①$_{\text{H}}$ column and the ③$_{\text{H}}$ column and between the ①$_{\text{L}}$ column and the ③$_{\text{L}}$ column. In Cao's table, the contours found in the ②$_{\text{L}}$ row are also found in the ③$_{\text{L}}$ row. In You and Yang's table, this is not the case, but in the figures here, it is. Finally, the Chao tone labels are often more similar to the data from Cao or even Qian than from You and Yang. The contour for ②$_{\text{H}}$①$_{\text{L}}$ is given by Qian as $\text{T}^{55}\text{T}^{31}$, by Cao as $\text{T}^{55}\text{T}^{313}$, and by You and Yang as $\text{T}^{44}\text{T}^{323}$. In my data, the label for older speakers is $\text{T}^{55}\text{T}^{41}$. In the cell just left of that one, ②$_{\text{H}}$①$_{\text{H}}$, Qian reports the contour $\text{T}^{54}\text{T}^{24}$ and Cao reports $\text{T}^{55}\text{T}^{33}$, while You and Yang report the contour $\text{T}^{44}\text{T}^{323}$ again. In my data, this contour is labeled $\text{T}^{55}\text{T}^{33}$ for both groups of urban speakers.

On the other hand, the Chao tone labels are sometimes more similar to the data from You and Yang. The four tone combinations of a ① tone and then a ② tone are labeled as $\text{T}^{33}\text{T}^{51}$ by Qian (among other contours), as $\text{T}^{33}\text{T}^{535}$ by Cao, and as $\text{T}^{33}\text{T}^{43}$ by You and Yang. In the data for the older speakers here, the corresponding contour is labeled $\text{T}^{33}\text{T}^{54}$. The contour for ①$_{\text{H}}$.①$_{\text{H}}$ words is reported by Qian and Cao as $\text{T}^{33}\text{T}^{55}$, but by You and Yang as $\text{T}^{33}\text{T}^{35}$. In my data, the contour was labeled $\text{T}^{33}\text{T}^{45}$, with a clearly rising contour, especially for the younger speakers.

These similarities between my labels and You and Yang's tone labels may reflect closer attention to measured pitch tracks, whereas the labels from Cao and from Qian may reflect abstracting away from assumed coarticulation effects. However, the differences between my data and You and Yang's data suggest that something was very different with either their stimuli selection or their speaker. It is likely that the word list they used for Jinhua was somewhat different than the word list they used for Linhai (the list on which mine was based), but it was probably not different enough to account for the very different patterns of merged categories. It is possible that the speaker they recorded was somehow not typical of urban Jinhua speakers. As shown by the quite similar contours of younger and older speakers, and of urban and village speakers, the typical differences among native speakers are not large enough to end up with a sandhi table like You and Yang's. But if I had recorded only speaker 23 (whose contours were so different from the other speakers) I might have come to contour labels quite different from what I have.

5.5 Conclusions

Though methodological differences may have had some effect, the differences among the results of previous studies of Jinhua sandhi is better attributed to a combination of the kinds of variation observed in urban Jinhua. The methodology of transcribing tones as Chao numbers by ear is inherently subjective. The linguist's own phonological categories are bound to influence the representation, and it is often unclear to what extent Chao tone transcriptions reflect coarticulation effects and efforts to remove coarticulation effects. Different methods of elicitation, especially the use of a carrier phrase, can also effect the data obtained. But the fact that You and Yang (2001), the previous study most similar to this one, is the one most different in results indicates that methodology is not the major factor. Methodological differences might account for small differences in contour shape, but not for differences in sandhi merger. The difference between the two generations of speakers in this study is both too small to account for the differences among the studies, and large enough to indicate that tone contours and categories can change quickly. The speakers recorded by Cao (1996) and by You and Yang (2001) were born 22 years apart and interviewed less than 10 years apart. The speakers interviewed by Qian (1992) were of approximately the same generation. The two generations I recorded are about 40 years apart, and yet they have far fewer differences than the previous studies do. The selection of stimuli clearly has an effect. Neither current isolation tone categories nor historical tone categories are sufficient to determine the sandhi contours. So when just a few words are selected to represent a particular tone combination, one such sample might miss a contour that was found in another, or two samples might have no contour in common. But the sandhi table from Cao (1996) is based on an extensive lexical inventory. If the selection of stimuli were all that mattered, Cao's sandhi table should include all the variants found in other tables, but it does not. The observed variation among studies must result from a combination of these factors.

One outstanding question is how such a complicated sandhi system is learnable. The system is at least semi-productive, since common neologisms appear integrated with the rest of the system. A good analogy, perhaps, is the English stress system, which shows a great deal of regularity, but because it is complicated by subregularities, historical change, and borrowings, the system can be quite opaque to learners, especially second language learners. But the system is nevertheless learned and maintained across generations.

This study confirms the claim made by Cao (2002) that sandhi patterns in Jinhua are both complicated and difficult to explain. There appear to be strata or subphonologies in Jinhua sandhi. It may be that some of these are more productive than others, and the more productive patterns might be susceptible to explanatory rules or constraint-based analysis. But considering the nature of active tonal evolution suggested here, it seems more likely that

the explanations for the observed patterns lies in history, not in the synchronic grammar. Fine phonetic detail matters considerably in tonal evolution, but the accumulated effects of tonal evolution create sandhi contours with almost arbitrary relation to the isolation contours. The most perspicuous synchronic representation of these phonological patterns may be nothing more than an exemplar model or arbitrary rules that essentially reproduce the sandhi table.

References

Ballard, W. L. (1988). The history and development of tonal systems and tone alternations in South China. *Study of Languages and Cultures of Asia and Africa: Monograph Series*(22).

Baxter, W. (n.d.). *An etymological dictionary of common Chinese characters.* (Draft available at http://www-personal.umich.edu/~wbaxter/etymdict.html, accessed October 2009.)

Baxter, W. (1992). *A handbook of Old Chinese phonology.* Mouton De Gruyter.

Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 206–44). Oxford/Cambridge, MA.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *IFA proceedings* (Vol. 17, pp. 97–110).

Cao, Z. (1996). *Jinhua fangyan cidian [Jinhua dialect dictionary]* (R. Li, Ed.). Nanjing: Jiangsu Jiaoyu Chuban She.

Cao, Z. (2002). *Nanbu Wuyu yuyin yanjiu [Southern Wu phonology research].* Beijing: Shangwu Yin Shuguan.

Chao, Y.-R. (1928). *Xiandai Wuyu de yanjiu [Studies of Modern Wu].* Tsing Hua College Research Institute.

Chao, Y.-R. (1930). A system of tone-letters. *Le Maitre Phonetique*, *30*, 24–27.

Chao, Y.-R. (1948). *Mandarin primer.* Harvard University Press.

Chao, Y.-R. (1968). *A grammar of spoken Chinese.* Berkeley: University of California Press.

Chen, M. Y. (1987). The syntax of Xiamen tone sandhi. *Phonology Yearbook*, *4*, 109–149.

Chen, M. Y. (1990). What must phonology know about syntax. In S. Inkelas & D. Zec (Eds.), *The phonology-syntax connection.* University of Chicago Press.

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

Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics*, *2*(2), 135–158.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*, 633–664.

Duanmu, S. (1993). Rime length, stress, and association domains. *Journal of East Asian Linguistics*, *2*(1), 1–44.

Duanmu, S. (1999). Metrical structure and tone: Evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics*, *8*, 1–38.

Duanmu, S. (2007). *The phonology of Standard Chinese* (2nd ed.). Oxford University Press.

Earle, M. A. (1975). *An acoustic phonetic study of Northern Vietnamese tones.* Santa Barbara, CA: Speech Communications Research Laboratory.

Gandour, J., & Potisuk, S. (1994). Tonal coarticulation in Thai. *Journal of Phonetics*, *22*, 477-492.

Gordon, M. (2001). A typology of contour tone restrictions. *Studies in Language*, *25*(3), 423–462.

Gu, W., & Lee, T. (2007). Effects of tonal context and focus on Cantonese f0. In *Proceedings of the 16th ICPhS* (pp. 1033–1036).

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones.* Cambridge University Press.

Hyman, L. M. (2007). Universals of tone rules: 30 years later. In T. Raid & C. Gussenhoven (Eds.), *Tones and tunes* (Vol. 1, pp. 1–34). Mouton De Gruyter.

Johnson, K. (2005). The handbook of speech perception. In D. B. Pisoni & R. E. Remez (Eds.), (chap. Chapter 15: Speaker Normalization in Speech Perception). Blackwell.

Kennedy, G. A. (1953). Two tone patterns in Tangsic. *Language*, *29*(3), 367–373.

Ladd, D. R. (1984). Declination: A review and some hypotheses. *Phonology Yearbook*, *1*, 53–74.

Ladd, D. R. (1988). Declination "reset" and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America*, *84*, 530.

Lin, M. (1995). A perceptual study on the domain of tones in Standard Chinese. *Chinese Journal of Acoustics*, *14*(4), 350–357.

Lin, Y.-H. (1989). *Autosegmental treatment of segmental processes in Chinese phonology.* Unpublished doctoral dissertation, University of Texas at Austin.

Matisoff, J. (1973). Tonogenesis in Southeast Asia. In L. Hyman (Ed.), *Consonant types and tone* (Vol. 1, pp. 71–95). Los Angeles: UCLA.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese. *Journal of the Acoustic Society of America*, *102*(3), 1864–1877.

Myers, J., & Tsay, J. (2003). A formal functional model of tone. *Language and Linguistics*, *4*(1), 105-138.

Myers, S. (1997). OCP effects in optimality theory. *Natural Language and Linguistic Theory*, *15*(4), 847-892.

Peng, S.-H. (1997). Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics*, *25*, 371–400.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (Vol. 7, pp. 101–139). Mouton de Gruyter.

Port, R. F., & Leary, A. P. (2005). Against formal phonology. *Language*, *81*(4), 927–964.

Qian, N. (1992). *Dangdai Wuyu yanjiu [Contemporary Wu research]*. Shanghai: Shanghai Jiaoyu Chuban She.

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication*, *6*, 343–351.

Rose, P. (1991). How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency? *Speech Communication*, *10*, 229–247.

Rose, P. (1998). The differential status of semivowels in the acoustic phonetic realisation of tone. In *Proceedings of the 5th International Conference on Spoken Language Processing.* (Paper 298)

Shen, X. S. (1992). On tone sandhi and tonal coarticulation. *Acta Linguistica Hafniensia*, *25*, 83-94.

Shih, C. (1988). Tone and intonation in Mandarin. In N. Clements (Ed.), *Working papers of the Cornell Phonetics Laboratory* (p. 83-109). CLC Publications.

Shih, C. (1997). Declination in Mandarin. In *Proceedings of INT-1997 intonation: Theory, models, and applications* (pp. 293–296).

Shih, C., Kochanski, G., & Yoon, S.-Y. (2007). The missing link between articulatory gestures and sentence planning. In *Proceedings of ICPhS2007.*

Shriberg, E., Ferrer, L., Venkataraman, A., & Kajarekar, S. (2004). SVM modeling of "SNERF-grams" for speaker recognition. In *Proceedings of ICSLP 2004.*

Stanford, J. N. (2008). A sociotonetic analysis of Sui dialect contact. *Language Variation and Change*, *20*, 409–450.

Tang, C. (2009). *Mutual intelligibility of Chinese dialects: an experimental approach.* Unpublished doctoral dissertation, Leiden University.

Thubthong, N., Kijsirikul, B., & Luksaneeyanawin, S. (2002). An empirical study for constructing Thai tone models. In *Joint international conference of SNLP-Oriental COCOSDA 2002.*

Ting, P.-H. (1996). Tonal evolution and tonal reconstruction in Chinese. In C. teh James Huang & Y.-H. A. Li (Eds.), *New horizons in Chinese linguistics.* Springer.

Weijer, J. van de, & Zhang, J. (2008). An X-bar approach to the syllable structure of Mandarin. *Lingua*, *118*, 1416–1428.

Wong, M. K.-S. (1982). *Tone change in Cantonese.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Xu, B. (1999). *Hanyu fangyan da cidian [chinese dialect dictionary]*. Beijing: Zhonghua Shuju.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*, 55–105.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from mandarin chinese. *Speech Communication*, *33*, 319–337.

Yip, M. (1989). Contour tones. *Phonology*, *6*(1), 149–174.

You, R., & Yang, J. (2001). Jinhua fangyan shengdiao shiyan yanjiu [Jinhua dialect tone experimental research]. In *Wuyu shengdiao de shiyan yanjiu [Wu tone experimental research].* Shanghai: Fudan Daxue Chuban She.

Yue-Hashimoto, A. O. (1987). Tone sandhi across Chinese dialects. In Chinese Language Society of Hong Kong (Ed.), *Chinese language* (pp. 445–474). Hong Kong: Joint Publishing Co.

Zhang, J. (1999). Duration in the tonal phonology of Pingyao Chinese. In M. K. Gordon (Ed.), *Papers in phonology* (Vol. 3, pp. 147–206).

Zhang, J. (2002). *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis.* Routledge, New York.

Zhang, J. (2006). *The phonology of Shaoxing Chinese.* LOT.

Zhang, J. (2007). A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*.

Zhang, J., & Hirose, K. (2004). Tone nucleus modeling for Chinese lexical tone recognition. *Speech Communication*, *42*, 447–466.

Zhu, X. S. (1999). *Shanghai tonetics.* Lincom Europa.

A Stimuli

Item	Written	IPA	gloss
11_1	医生	i.sɑŋ	doctor
11_2	西瓜	sie.kua	watermelon
11_3	飞机	fi.tɕi	airplane
12_1	工人	koŋ.ɳiŋ	laborer
12_2	天雷	t^{h}ia.lɛ	lightning
12_3	花瓶	hua.piŋ	vase
13_1	青草	ts^{h}iŋ.ts^{h}ɑu	green grass
13_2	风水	foŋ.ɕy	feng shui
13_3	资本	tsɿ.pəŋ	capital
14_1	沙眼	sua.ɑ	trachoma
14_2	经理	tɕiŋ.li	manager
14_3	婚礼	huəŋ.li	wedding
15_1	青菜	ts^{h}iŋ.ts^{h}ɛ	bok choy
15_2	车票	tɕ^{h}ia.p^{h}iau	train ticket
15_3	相信	siɑŋ.siŋ	believe
16_1	山洞	sa.toŋ	cave
16_2	医院	i.yɤ	hospital
16_3	哥弟	kuɤ.ti	brother
17_1	中国	tɕioŋ.kuəʔ	China
17_2	钢笔	kaŋ.piəʔ	fountain pen
17_3	资格	tsɿ.kəʔ	seniority
18_1	中学	tɕioŋ.oʔ	secondary school
18_2	樟木	tɕiɑŋ.moʔ	camphor wood
18_3	阴历	iŋ.liəʔ	lunar calendar
21_1	田鸡	tia.tɕie	frog
21_2	梅花	mɛ.huɑ	plum blossom
21_3	洋葱	iɑŋ.ts^{h}oŋ	onion
22_1	羊皮	iɑŋ.pi	sheep skin
22_2	皮球	pi.tɕiu	ball
22_3	明年	miŋ.ɳia	next year
23_1	红枣	oŋ.tsao	red date
23_2	牙齿	uɑ.ts^{h}ɿ	teeth
23_3	锣鼓	luɤ.ku	gong and drum

Item	Written	IPA	gloss
24_1	牛奶	ȵiu.nɑ	milk
24_2	城里	ɕiŋ.li	downtown
24_3	完美	uan.mɛ	perfect
25_1	同辈	toŋ.pɛ	same generation
25_2	迷信	mi.ɕiŋ	be superstitious
25_3	棉裤	mie.k^{h}u	padded trousers
26_1	蚕豆	sɤ.tiu	fava bean
26_2	强盗	tɕiɑŋ.tao	robber
26_3	茶叶	tsua.ie	tea leaf
27_1	毛笔	mau.piəʔ	brush pen
27_2	南北	nɤ.pəʔ	North and South
27_3	颜色	ɑ.sɤʔ	color
28_1	农历	noŋ.liəʔ	lunar calendar
28_2	咸肉	ɑ.ȵioʔ	bacon
28_3	红木	oŋ.moʔ	mahogany
31_1	祖宗	tsu.tsoŋ	ancestors
31_2	火车	huɤ.tɕ^{h}ia	train
31_3	点心	tia.siŋ	snack
32_1	好人	hɑu.ȵiŋ	good person
32_2	酒瓶	tɕiu.piŋ	wine bottle
32_3	草鞋	ts^{h}au.ɑ	straw sandals
33_1	火腿	huɤ.tɛ	ham
33_2	水果	ɕy.kuɤ	fruit
33_3	厂长	tɕ^{h}iɑŋ.tɕiɑŋ	factory manager
34_1	表演	piɑu.ian	perform
34_2	好象	hɑu.siaŋ	resemble
34_3	改造	kɛ.tsau	transform
35_1	广告	kuɑŋ.kau	advertisement
35_2	比赛	pi.sɛ	competition
35_3	考试	k^{h}au.sɻ	exam
36_1	酒酿	tsiu.ȵiɑŋ	fermented rice
36_2	草帽	ts^{h}ɑu.mɑu	straw hat
36_3	本事	pəŋ.sɻ	skill

Item	Written	IPA	gloss
37_1	粉笔	fəŋ.piəʔ	chalk
37_2	补贴	pu.tʰia	subsidy
37_3	指甲	tsɻ.kua	finger nail
38_1	小学	siau.oʔ	primary school
38_2	体育	ti.ioʔ	sport
38_3	坦白	tʰan.bəʔ	candid
41_1	已经	i.tɕiŋ	already
41_2	老公	lɑu.koŋ	husband
41_3	米缸	mie.kɑŋ	rice jar
42_1	肚皮	tu.pi	belly
42_2	象棋	siɑŋ.tɕi	chess
42_3	野猫	ie.mɑu	wild cat
43_1	耳朵	ŋ.tuɤ	ear
43_2	冷水	lɑŋ.ɕy	cold water
43_3	户口	u.kʰiu	home registration
44_1	道理	tɑu.li	reason
44_2	马桶	mɤa.toŋ	toilet bowl
44_3	美女	mei.nʯy	pretty woman
45_1	老货	lɑu.huɤ	old folks
45_2	冷气	lɑŋ.tɕʰi	cold air
45_3	野菜	ie.tsʰɑu	wild greens
46_1	马路	mɤa.lu	road
46_2	冷汗	lɑŋ.ɤ	cold sweat
46_3	眼泪	ɑ.lɛ	tears
47_1	负责	fo.tsəʔ	complicated
47_2	冷粥	lɑŋ.tɕioʔ	cold congee
47_3	道德	tɑu.təʔ	morals
48_1	厚薄	kiu.poʔ	thickness
48_2	老实	lɑu.ɕiəʔ	honest
48_3	动物	toŋ.fəʔ	animal
51_1	信心	siŋ.siŋ	faith
51_2	汽车	tɕʰi.tɕʰia	automobile
51_3	衬衫	tsʰəŋ.sɑ	shirt

Item	Written	IPA	gloss
52_1	算盘	sɤ.pɤ	abacus
52_2	汽油	tɕʰi.iu	gasoline
52_3	桂圆	kui.yan	longan
53_1	报纸	pɑu.tsɻ	newspaper
53_2	汽水	tɕʰi.ɕy	soda water
53_3	对比	tɛ.pi	compare
54_1	报社	pɑu.sɤ	news agency
54_2	对象	tɛ.siɑŋ	partner
54_3	创造	tɕiɑŋ.tsɑu	create
55_1	兴趣	ɕiŋ.tɕy	interest
55_2	唱片	tɕʰiɑŋ.pʰie	vinyl record
55_3	世界	sɻ.tɕie	world
56_1	炸弹	tsɑ.tan	explosion
56_2	笑话	siɑu.ua	joke
56_3	性命	siŋ.miŋ	life
57_1	信壳	ɕiŋ.kʰoʔ	envelope
57_2	印刷	iŋ.ɕyɤ	print
57_3	庆祝	tɕiŋ.tsoʔ	celebrate
58_1	性别	siŋ.piəʔ	sex/gender
58_2	汉族	han.tsoʔ	Han ethnicity
58_3	泡沫	pʰɑu.moʔ	foam
61_1	外甥	ɑ.sɑŋ	nephew
61_2	汗衫	ɤ.sɑ	undershirt
61_3	电灯	tia.təŋ	electric light
62_1	外婆	ɑ.pɤ	maternal grandma
62_2	地球	ti.tɕiu	Earth
62_3	暂时	tsan.sɻ	temporary
63_1	字典	sɻ.tia	dictionary
63_2	大小	tuɤ.siɑu	size
63_3	胃口	ui.kʰiu	appetite
64_1	号码	ɑu.mɤa	number
64_2	大雨	tuɤ.y	heavy rain
64_3	饭碗	fan.uan	rice bowl

Item	Written	IPA	gloss
65_1	电报	tia.pɑu	telegram
65_2	话费	uɑ.fi	phone charge
65_3	饭店	fan.tia	restaurant
66_1	外地	a.ti	distant lands
66_2	豆腐	tiu.fu	toufu
66_3	部队	pu.tɛ	troops
67_1	外国	ɑ.kuəʔ	foreign countries
67_2	字帖	sɿ.tʰia	model calligraphy
67_3	办法	pʏ.fɤa	method
68_1	蛋白	tan.pəʔ	eggwhite
68_2	大学	tɑ.oʔ	college
68_3	婺剧	fu.tɕiəʔ	local (Jinhua) opera
71_1	浙江	tsiəʔ.kɑŋ	Zhejiang
71_2	菊花	tɕioʔ.huɑ	chrysanthemum
71_3	北风	pəʔ.foŋ	North wind
72_1	足球	tsoʔ.tɕiu	soccer
72_2	发条	fɤa.tiau	spring
72_3	色盲	səʔ.mɑŋ	colorblind
73_1	脚底	tɕiəʔ.tie	foot sole
73_2	出版	tɕʰyəʔ.pan	publish
73_3	屋顶	oʔ.tiŋ	roof
74_1	黑市	həʔ.sɿ	black market
74_2	粟米	soʔ.mie	corn
74_3	尺码	tɕʰiəʔ.mʏa	size
75_1	百货	pəʔ.huʏ	general goods
75_2	客气	kʰəʔ.tɕʰi	polite
75_3	尺寸	tɕʰiəʔ.tsəŋ	size
76_1	革命	kəʔ.miŋ	revolutionize
76_2	铁路	tʰia.lu	railroad
76_3	赤豆	tɕʰiəʔ.tiu	red bean
77_1	法国	fɤa.kuəʔ	France
77_2	隔壁	kəʔ.piəʔ	neighbor
77_3	发作	fɤa.tsoʔ	flare up

Item	Written	IPA	gloss
78_1	锡箔	ɕiəʔ.poʔ	tin foil
78_2	雪白	sie.pəʔ	snow white
78_3	笔直	piəʔ.tɕiəʔ	straight
81_1	蜜蜂	miəʔ.foŋ	bee
81_2	辣椒	lua.huũ	hot pepper
81_3	肉丝	ɳioʔ.sɿ	shredded pork
82_1	药房	iəʔ.fɑŋ	drugstore
82_2	毒蛇	toʔ.ɕia	viper
82_3	木头	moʔ.tiu	log
83_1	药水	iəʔ.ɕy	lotion
83_2	局长	tɕioʔ.tɕiɑŋ	postmaster
83_3	木板	moʔ.pan	board
84_1	物理	fəʔ.li	physics
84_2	落后	loʔ.eu	fall back
84_3	白马	pəʔ.mɤa	white horse
85_1	热气	ɳiəʔ.tɕʰi	heat
85_2	力气	liəʔ.tɕʰi	strength
85_3	学费	oʔ.fi	tuition
86_1	服务	foʔ.u	service
86_2	绿豆	loʔ.tiu	greenbean
86_3	木匠	moʔ.ɕiɑŋ	carpenter
87_1	蜡烛	luɑ.tɕioʔ	candle
87_2	合作	hɤ.tsuo	cooperate
87_3	及格	tsiəʔ.kɤ	pass
88_1	腊肉	lua.ɳioʔ	cured pork
88_2	毒药	toʔ.iəʔ	poison
88_3	独立	toʔ.liəʔ	independent