AGAINST STRICT CORRESPONDENCE BETWEEN PHONETIC MEASUREMENTS AND PHONOLOGICAL REPRESENTATIONS

By

Naiyan Du

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Linguistics—Doctor of Philosophy

2023

ABSTRACT

As observed in different subfields of psychology, the relationship between knowledge and performance is rather remote. Similarly, in linguistics, there is no guarantee about the correspondence between phonetic measurements and phonological representations. There are multiple interacting sources that affect speech production, e.g. lexical knowledge, phonological knowledge, memory and processing constraints, etc. Consequently, phonetic manifestations cannot automatically or solely be used as a diagnostic of any phonological representations. However, in previous literature, the strict correspondence between phonetic distributions and phonological representations is often explicitly or implicitly assumed. In this dissertation, I will argue using data from Huai'an Mandarin that non-linguistic performance factors can have significant influence on phonetic manifestations in unexpected directions. To be more specific, I will present production data from Huai'an Mandarin to make three points. First, I will show a case of phonologically complete neutralization that results in phonetically incompleteness with a large effect size. Second, I will show a case of phonetically complete neutralization that results in phonological behavioral differences. Third, I will show a case where optional phonological processes lead to more phonetic incompleteness.

For the first point, I will show in Huai'an Mandarin that derived Tone 3s from Tone 1 or Tone 4 sandhis are phonetically very different from underlying Tone 3 that did not undergo any phonological processes. The Tone 3 category of derived Tone 3 is established by the previous description that only Tone 3 can trigger Tone 3 sandhi in Huai'an Mandarin. Since both derived Tone 3 from Tone 1/Tone 4 sandhis and underlying Tone 3 can trigger Tone 3 sandhi, derived Tone 3 should be phonologically identical with underlying Tone 3. For the second point, I will use evidence also from Huai'an Mandarin arguing that two Tone 3s derived at the lexical and the post-

lexical levels have different phonological behaviors despite that they are arguably phonetically identical. The phonological behavior difference in this case is indicated by different rates of the two derived Tone 3s triggering another Tone 3 process. Despite being arguably phonologically different, the described derived Tone 3s are indistinguishable in all major phonetic cues including f0, duration and intensity. For the third point, I will show in Huai'an Mandarin that only optional phonological processes (Tone 1/Tone 4 sandhis) can have incomplete phonetic neutralization with a rather large effect size, while mandatory phonological process (Tone 3 sandhi) in the same language can only have incomplete neutralization with a very small effect size.

Overall, data from Huai'an Mandarin provide strong evidence for the gap between phonetic measurements and phonological representations. And phonetic measurements can only inform of phonological knowledge when accompanied by other evidence under certain circumstances.

Copyright by NAIYAN DU 2023 To all the people related to the City of Huai'an

ACKNOWLEDGMENTS

I would like sincerely thank my advisors: Karthik Durvasula and Yen-Hwei Lin for their help during this process. Without them, I cannot go this far in the field of linguistics. I would also like to thank other members of my committee: Silvina Bongiovanni and Suzanne Wagner for their support. I would also like to thank my cohort: Jason Smith, Komeil Ahari and Shannon Cousins for going through this process with me. Last but not least, I would like to thank my parents Du, Xiwei and Zhang, Huaying for everything they have provided for me.

TABLE OF CONTENTS

CHAPTER	1 INTRODUCTION AND OVERVIEW ¹	1
1.1	The gap between phonetic measurements and phonological representations	1
1.2	Phonologically identical forms can have different phonetic distributions	5
1.2.1	Incomplete neutralization.	5
1.2.2	Near merger	7
1.3	Phonologically different forms can have identical phonetic distribution	10
1.4	Organization of the dissertation	12
CHAPTER	2 THE PHENOMENON OF INCOMPLETE NEUTRALIZATION ⁸	14
2.1	Introduction	14
2.2	The issue of phonological neutralization versus phonetic implementation	18
2.3	Background of Huai'an Mandarin.	23
2.4	Experiment 1: Tone 1 sandhi	33
2.4.1	Participants	33
2.4.2	Stimuli	33
2.4.3	Procedure	36
2.4.4	Measurement	36
2.4.5	Results and statistical modelling	38
2.5	Experiment 2: Tone 4 sandhi	44
2.5	Participants	<u>4</u> 4
2.5.1	Stimuli	45
2.5.2	Procedure	46
2.5.5 2.5.4	Measurement	16
2.5.4	Results and statistical modelling	-τ0 //7
2.5.5	Phonological representation of Mandarin tone	58
CHAPTER	3 PHONETICALLY IDENTICAL FORM CAN HAVE DIFFERENT	
PH	IONOLOGICAL BEHAVIORS	65
3.1	More background	65
3.2	Experiment 3	73
3.2.1	Participants	73
3.2.2	Stimuli	73
3.2.3	Procedure	74
3.2.4	Measurement	74
3.2.5	Results and statistical modelling	75
3.3	Interim discussion	87
CHAPTER	4 THE EXPLANATION FOR THE OBSERVED GAP BETWEEN PHONOLO AND PHONETICS ³²	GY 88
4.1	The explanation for incomplete neutralization	88
4.1.1	Desiderata for any explanation for incomplete neutralization	88
4.1.2	Previous explanations for incomplete neutralization.	91
4	1.2.1 Explanations within phonology	92
Δ.	1.2.2 Explanations inside phonology-phonetics interface	93
1		

4.1.3 The current explanation on incomplete neutralization 97 4.2 The explanation for phonetically identical form can have different phonological behaviors 106 CHAPTER 5 EXPERIMENTAL EVIDENCE FOR SEPARATE EXPLANATIONS FOR INCOMPLETE NEUTRALIZATION WITH DIFFERENT EFFECT SIZES 110 5.1 Experiment 4. 110 5.2 Participants 110 5.3 Stimuli. 111 5.4 Procedure 113 5.5 Measurement. 113 5.6 Results and statistical modelling 114 5.7 Interim Discussion 134 6.1 Summary of the findings 134 6.2 Variation in effect size 136 6.3 The relationship between effect size of incomplete neutralization and application rate 137 6.4 Groud anough logic 141
4.2 The explanation for phoneucarly iterated form can have unrefer phonological behaviors
CHAPTER 5 EXPERIMENTAL EVIDENCE FOR SEPARATE EXPLANATIONS FOR INCOMPLETE NEUTRALIZATION WITH DIFFERENT EFFECT SIZES
CHAPTER 5 EXPERIMENTAL EVIDENCE FOR SEPARATE EXPLANATIONS FOR INCOMPLETE NEUTRALIZATION WITH DIFFERENT EFFECT SIZES
INCOMPLETE NEUTRALIZATION WITH DIFFERENT EFFECT SIZES
5.1Experiment 4
5.2 Participants
5.3 Stimuli
5.4 Procedure
5.5 Measurement. 113 5.6 Results and statistical modelling. 114 5.7 Interim Discussion. 133 CHAPTER 6 GENERAL DISCUSSION. 134 6.1 Summary of the findings. 134 6.2 Variation in effect size. 136 6.3 The relationship between effect size of incomplete neutralization and application rate. 137 6.4 Good anough logic 141
5.6 Results and statistical modelling. 114 5.7 Interim Discussion. 133 CHAPTER 6 GENERAL DISCUSSION. 134 6.1 Summary of the findings. 134 6.2 Variation in effect size. 136 6.3 The relationship between effect size of incomplete neutralization and application rate. 137 6.4 Good anough logic 141
5.7 Interim Discussion
CHAPTER 6 GENERAL DISCUSSION
6.1 Summary of the findings
 6.1 Summary of the findings
 6.3 The relationship between effect size of incomplete neutralization and application rate
rate
64 Good arough logic
141
6.5 Not an accident
CHAPTER 7 CONCLUSION
7.1 Summary
7.2 Lingering question
BIBLIOGRAPHY
APPENDIX A: STIMULI FOR EXPERIMENT 1 ON TONE 1 SANDHI 161
APPENDIX B: STIMULI FOR EXPERIMENT 2 ON TONE 4 SANDHI 163
APPENDIX C: STIMULI FOR EXPERIMENT 3 ON TONE 4 SANDHI AT THE LEXICAL
AND POST-LEXICAL LEVELS 165
APPENDIX D: STIMULI FOR EXPERIMENT 4 ON TONE I/TONE 4/TONE 3
SANDHIS 108
APPENDIX E. DISTRIBUTION OF UNDERLYING TONE 1. DERIVED TONE 3 AND
UNDERLYING TONE 3 IN EACH STEP IN EXPERIMENT 1 174
APPENDIX F: DISTRIBUTION OF UNDERLYING TONE 4, DERIVED TONE 3 AND
UNDERLYING TONE 3 IN EACH STEP IN EXPERIMENT 2 176

APPENDIX G: COMPARISON OF RAW DURATION OF THE TWO TONE 3S IN	
EXPERIMENT 3 BY SPEAKER	178
APPENDIX H: DISTRIBUTION OF UNDERLYING TONE 1, DERIVED TONE 3 AND	
UNDERLYING TONE 3 IN EACH STEP IN EXPERIMENT 4	. 179
APPENDIX I: DISTRIBUTION OF UNDERLYING TONE 4, DERIVED TONE 3 AND	
UNDERLYING TONE 3 IN EACH STEP IN EXPERIMENT 4	. 181

CHAPTER 1 INTRODUCTION AND OVERVIEW¹

1.1 The gap between phonetic measurements and phonological representations

As has long been recognized in discussions of linguistic competence as abstract knowledge, there are multiple potential interacting factors in performance (Chomsky, 1964, 1965; Schütze, 1996; Valian, 1982; inter alia). Similarly, there are also multiple interacting sources that affect speech production, e.g. lexical knowledge, phonological knowledge, memory and processing constraints, etc. (Warner et al., 2004; Whalen, 1991, 1992; Wright, 2004). For example, Whalen (1991, 1992) showed that less frequent words generally have longer duration by comparing highfrequency and low-frequency homophones. Overall, phonetic manifestations cannot automatically or solely be used as a diagnostic of any phonological representations. I will argue in this dissertation using data from Huai'an Mandarin that performance factors can have significant influence on phonetic manifestations in unexpected directions. I also argue in this dissertation that the ultimately reliable way to probe phonological knowledge is through phonological behavior. To be more accurate, here 'phonological behavior' should mean systematic behavior that depends on an abstract characterization of structural context, and such behavior can't be simply reduced to known sources of co-articulation or other performance factors (Durvasula, personal communication). The meaning of 'phonological behavior' will be consistent throughout this dissertation,

As an obvious and convenient way to probe phonological knowledge, phonetic measurements have been widely employed to test linguistic theories. By allowing such methodology to falsify or verify any parts of existing theories, theoretical frameworks often assume strict correspondence

¹ Part of this chapter comes from my collaborative work with Karthik Durvasula, see reference (Du & Durvasula, accepted).

between phonological representations and phonetic distributions. Such an assumption can be formalized as two statements in (1):

(1) Two statements about the strict correspondence between phonological representations and phonetic patterns

Statement 1: Phonologically identical surface forms necessarily correspond with identical phonetic distributions.

Statement 2: Phonologically different surface forms necessarily correspond with different phonetic distributions.

Although the two statements have been supported by a large amount of data in natural languages, they may not always be true given potential counterevidence. The evidence that can undermine Statement 1 is obviously that phonologically identical surface forms can correspond with different phonetic distributions. And the evidence that can undermine Statement 2 is obviously that phonologically different surface forms can correspond with identical phonetic distributions. Both kinds of evidence have been argued to exist.

Statement 1 has been repeatedly challenged by previous experimental studies trying to test traditional formal phonology (Labov et al., 1972; Pierrehumbert, 2002; Port & O'Dell, 1985; inter alia), which is a theoretical framework that assumes strict correspondence between phonological representations and phonetic distributions. Under such a framework, both categorical phonological representation and a certain version of a modular feed-forward model are assumed (Braver, 2019; Goldrick & Blumstein, 2006; McCollum, 2019; Port & Leary, 2005; Manaster Ramer, 1996; Roettger et al., 2014). Therefore, the phonological representations are discrete elements that do

not contain any gradient phonetic information, and phonetics only has access to the output of phonology (Kenstowicz, 1994; Pierrehumbert, 2002). This theoretical view is referred by Du and Durvasula (accepted) as the *Standard generative view of Phonology*. I will use this term to refer to this theoretical framework throughout the dissertation.² To give an example, according to *Standard* generative view of Phonology, phonological neutralization process will result in a complete categorical change. This means, under appropriate contexts, a phonological neutralization process will result in phonologically different underlying forms becoming phonologically identical in the surface. Also, under the Standard generative view of Phonology, the underlying phonological difference should not have any consequences on phonetic manifestations. However, in cases of Incomplete Neutralization (Port & O'Dell, 1985) and Near Merger (Labov et al., 1972), there are traces of the underlying representation in the phonetic manifestation, and putatively phonologically identical surface forms have been argued to correspond with different phonetic distributions. This situation obviously undermines Statement 1. And as commented by Pierrehumbert (2002), Standard generative view of Phonology is oversimplified and some performance factors are completely ignored. I will propose in Section 4.1.3 that Incomplete Neutralization is caused by a performance factor, namely phonological planning effect.

Statement 2, in contrast, is rarely challenged. One noticeable previous challenge comes from Hyman (1975). He presents a putative case from Sea Dayak (data from Scott 1957, 1964) where the existence or absence of /g/ in phonologically underlying representations does not lead to any differences in the phonetics. However, Hyman did not show any phonetic or phonological evidence

² While this conception of phonology is relatively standard according to Du and Durvasula (accepted), it is actually quite different from what Du and Durvasula refers to as *Classic generative view of Phonology*. Under *Classic generative view of Phonology*, although a certain version of modular feedforward model is still assumed, phonology is seen as *knowledge* and linguistic performance is a multi-factorial problem. I return to this issue at the end of this subsection and also in the Chapter 6, where I point out that the latter notion has no trouble in accounting for facts related to incomplete neutralization.

for phonetic identity or phonological inequality. Overall, to the best of my knowledge, there are no previous careful experimental studies that attempts to undermine Statement 2. Statement 2 leads to a deduction that phonetically identical forms necessarily correspond with identical surface forms in the phonology, all else being equal. Under any theoretical framework that assumes strict correspondence between phonetic measurements and phonological representations, a finding that can obviously undermine this deduction and Statement 2 will be that two phonetically identical entities can have different phonological behaviors. In this dissertation, I will present an empirical case from Huai'an Mandarin arguing that two described Tone 3s that are arguably phonetically identically identical behaviors with regard to the ability of triggering tone sandhi processes.

Overall I will argue that the two statements are both doubtworthy. Therefore theoretical frameworks that assume the strict correspondence between phonological representations and phonetic patterns are also similarly doubtworthy. In contrast, the gap between linguistic competence (which contains phonological knowledge) and performance (which includes phonetic measurements) is well recognized in what Du and Durvasula (accepted) call the *Classic generative view of Phonology*, where phonology is seen as knowledge (Chomsky, 1965; Chomsky & Halle, 1965, 1968; inter alia). As per this view, linguistic performance is a multi-factorial problem, and linguistic knowledge (i.e., competence) is only one of the many factors involved (Chomsky, 1965; Schütze, 1996; Valian, 1982; Warner et al., 2004; inter alia).³ Moreover, the *Classic generative view of Phonology* still assumes categorical phonological representation. And since *Classic generative view of Phonology*, like *Standard generative view of Phonology*, does not allow

³ I am not aware of any explicit argumentation that has ever been put forward in support of the *Standard generative view of Phonology* over the *Classic generative view of Phonology*. So, I am at a loss as to precisely when and, more importantly, why this change in viewpoints occurred. Here, I simply note the discrepancy.

gradience in phonology, it is much simpler than theories that attempt to introduce gradience into the phonology. The distinction between the *Standard generative view of Phonology* and the *Classic generative view of Phonology* are illustrated in Figure 1. In this dissertation, I will show that previous examined cases of Incomplete Neutralization and Near Merger and the Huai'an data I present can also find a simple and satisfying explanation under the *Classic generative view of Phonology*, and there is no need to appeal to more complicated theoretical models.



Figure 1: The distinction between Standard generative view of Phonology and Classic generative view of Phonology

1.2 Phonologically identical forms can have different phonetic distributions

1.2.1 Incomplete neutralization

The phenomenon of incomplete neutralization describes a situation where phonological identical surface forms can have different phonetic distributions, i.e., when underlying phonological contrast is collapsed due to phonological processes, the resulting phonological category can have a different phonetic distribution for each underlying representation.

The research on incomplete neutralization itself has a long history since at least the mid-1980s with empirical evidence from a variety of languages including Catalan (Dinnsen & Charles-Luce, 1984), Dutch (Ernestus & Baayen, 2006; Warner et al., 2004), Japanese (Braver & Kawahara, 2016), Polish (Slowiaczek & Dinnsen, 1985; Slowiaczek & Szymanska, 1989) and Russian (Dmitrieva, 2005; Kharlamov, 2012; Matsui, 2015). For example, in German it has been described that the phonological voicing contrast for obstruents is neutralized at the right edge of a prosodic word (Wagner, 2002). A rule-based mapping of the relevant phonological process is stated in (2). A real-world example can be that both underlying /alb/ (meaning: elf) and underlying /alp/ (meaning: mountain pasture) become /alp/ in the surface. However, careful phonetic and perceptual experimentation has shown that the neutralization is incomplete phonetically (Port & O'Dell, 1985; Roettger et al., 2014; inter alia). In other words, underlying voiceless stops, derived voiceless stops and *underlying* voiced stops all have different phonetic distributions. To be succinct, here I use '*underlying* voiceless stop' to mean surface voiceless stops that map from underlying voiceless stops, 'derived voiceless stop' to mean surface voiceless stops that map from underlying voiced stops, and 'underlying voiced stops' to mean surface voiced stops that map from underlying voiced stops. I will use the terms 'underlying' and 'derived' in the same fashion to describe stops and tones for the rest of the paper.

(2) [- sonorant] --> [-voice]/ __) ω

Note: ' ω ' means prosodic word.

The observed phenomenon of incomplete neutralization has been argued by many researchers to pose a challenge to the *Standard generative view of Phonology*. In such a phenomenon, the phonetic distributions seems not be decided wholly by the output of phonology. Therefore, strict correspondence between phonological representations and phonetic distributions that is assumed in *Standard generative view of Phonology* is doubtful. In contrast, the observed phenomenon of incomplete neutralization is perfectly compatible with *Classic generative view of Phonology*, where linguistic performance is viewed as a multi-factorial problem (Chomsky, 1964, 1965; Schütze, 1996; Valian, 1982; Warner et al., 2004; inter alia). Any performance factor other than phonological knowledge can result in incomplete neutralization in the phonetics.

1.2.2 Near merger

Similar to incomplete neutralization, the phenomenon of near merger describes an empirical situation where phonological identical surface forms can have different phonetic distributions. According to the definition of Yu (2007, 2011), near merger describes the situation where speakers consistently report that two classes of sounds are 'the same', yet consistently differentiate them in production at better than chance level. For example, some English speakers in New York produce <source>⁴ and <sauce> in a systematic different fashion but cannot distinguish them in perception (Labov et al., 1972). Therefore, in both cases of incomplete neutralization and near merger, speakers manage to maintain a systematic difference in the phonetics when they consistently fail to identify the distinction at the conscious or near-conscious level.

For Yu (2007, 2011), both incomplete neutralization and near merger results in a collapse in phonological contrast. However, Labov (1975) hypothesized that a near merger only brings two phonemes into a very close approximation. According to Labov, the phonemic difference is still maintained in the surface while semantic contrasts between them are suspended for native speakers

⁴ In this subsection, I present these forms in English spellings using angle brackets '<...>'.

of the dialect. The phonemic difference then naturally explains for the systematic difference in production. ⁵ For this dissertation, I assume Yu's (2007, 2011) view that both incomplete neutralization and near merger results in identical surface forms in the phonology. We don't need to rely on the surface phonemic difference in Labov's hypothesis to provide an explanation for the systematic different distributions in the phonetics. Similar with incomplete neutralization, near merger is also expected by *Classic generative view of Phonology*. Performance factors outside phonological knowledge can potentially explain the difference in the phonetics.

The main difference between incomplete neutralization and near merger lies in the fact that the same surface phonological category in incomplete neutralization comes from synchronic phonological processes while the same phonological category in near merger is a result of diachronic sound change. Another noticeable difference, as described by Yu (2007), is that near mergers are found among lexical items or morphemes and does not rely on any phonological contexts.⁶ In contrast, incomplete neutralization is a result of phonological processes and usually appears in certain phonological contexts.

Previous empirical evidence of near merger mainly comes from English. Famous examples include <fool> and <full> in Albuquerque (Di Paolo, 1988); <too> vs. <toe> and <beer> vs. <bear> in Norwich (Trudgill, 1974); <line> vs. <loin> in Essex (Labov, 1971; Nunberg, 1980); <meat> vs. <mate> in Belfast (Milroy & Harris, 1980; Harris, 1985). More recently, Yu (2007) presented experimental evidence that Cantonese Pinjam results in a near merger. In Cantonese, Pinjam

⁵ Under Labov's explanation, it is not clear how a phonemic difference can still be maintained when two phonemes are so close to each other. Labov also did not provide convincing phonological evidence that the phonemic difference still exists in cases of Near Merger.

⁶ It is worth noting that some synchronic phonological processes also do not rely on phonological contexts. These cases include low vowel harmony in Okpe (Pulleyblank, 1986) and neutralization of nasalized vowels in Kpelle (Hyman, 1975). The authors claimed full neutralization in the phonetics in the original papers but they did not offer any experimental evidence. Therefore, the possibility cannot be ruled out that these synchronic phonological processes that do not rely on phonological contexts can result in incomplete phonetic neutralization. Future experimental studies on these languages are needed to discern.

describes a diachronic tonal change which is correlated with semantic change (Downer, 1959; Kam, 1977). The only case of Pinjam in Modern Cantonese is a non-high-level or a non-mid-rising tone becoming a mid-rising tone (Yu, 2007). Such tonal change is correlated with nominalization of verbs as shown in examples in (3) (Examples from Yu, 2007). It is described in previous literature (Downer, 1959; Kam, 1977) that the tonal change results in a rising tone category that is indistinguishable from the corresponding lexical tone, the minimal pair examples are shown in (4), which are also used as stimuli in production experiment by Yu (2007). However, when measured, the derived tone and the corresponding lexical tone have slightly different phonetic distributions.

(3)	a. Level	a. Level tone			b. Rising tone		
	sou33	'to sweep'	\rightarrow	sou35	'a broom'		
	pəŋ22	'to weigh'	\rightarrow	pəŋ35	'a scale'		
	mə11	'to grind'	\rightarrow	m ə 35	'a grind'		
	tan22	'to pluck'	\rightarrow	tan35	'a missile'		
	wa22	'to listen'	\rightarrow	wa35	'an utterance'		
	jeu11	'to grease'	\rightarrow	jeu35	'oil'		
	liu11	'to provoke'	\rightarrow	liu35	'a stir'		
	ts ^h øų11	'to hammer'	\rightarrow	ts ^h 0y35	'a hammer'		
	ts ^h ə11	'to plough'	\rightarrow	ts ^h o35	'a plough'		

(4)	a. Lexical rising tone		b. Derived rising tone		
	søŋ55 <u>fan35</u>	'opposite'	kam55 <u>fan35</u>	'prisoner'	
	kvi55 <u>tan35</u>	'egg'	fei55 <u>tan35</u>	'a missile'	
	pow55 <u>pin35</u>	'to critique'	ts ^h oy21 pin35	'casual'	
	suŋ55 <u>pɔŋ35</u>	'to untie'	ts ^h uŋ23 <u>pɔŋ35</u>	'heavy'	
	ts ^h øŋ21 <u>kɛŋ35</u>	'long neck'	ŋan23 <u>kɛŋ35</u>	'glasses'	
	fa55 <u>fen35</u>	'pollen'	ku35 <u>fen35</u>	'(stock) share'	
	tshət 55 <u>pan 35</u>	'to publish'	siu35 <u>pan35</u>	'peddler'	
	tsi:22 <u>tin35</u>	'dictionary'	t ^h iu21 <u>kin35</u>	'terms, conditions'	

Overall, the phenomenon of near merger posts a challenge to the *Standard generative view of Phonology* in exactly the same fashion as the phenomenon of incomplete neutralization. However, again, the phenomenon does not pose a problem to the *Classic generative view of Phonology*.

1.3 Phonologically different forms can have identical phonetic distribution

As introduced in Section 1.1, little doubt is cast by previous literature on Statement 2 in (1), which states that phonologically different surface forms necessarily correspond with different phonetic distributions. Again Statement 2 leads to a deduction that phonetic identical forms necessarily correspond with identical surface forms in phonology, all things being equal. However, I will show a case in Huai'an Mandarin (Huai'an hereafter) where such a claim may not hold. The described Tone 3s that are derived at the lexical and post-lexical levels have identical phonetic distributions with regard to all important phonetic cues (f0, duration and intensity). Therefore, the two derived Tone 3s are arguably identical in the phonetics. However, the two derived Tone 3s

have different phonological behaviors with regard to the ability of triggering another tone sandhi process, which points to phonological inequality. If both phonological inequality and phonetic identity can be established as I will argue in this dissertation, Statement 2 will be undermined and Huai'an case will provide strong evidence for the gap between phonological representations and phonetic distributions from a different angle.

The difficulty of finding such cases in natural languages is obvious. It is extremely difficult to establish phonetic identity. There are logically an infinite number of phonetic cues to be observed for any phonological elements, and general phonetic identity between two phonological elements requires identity in all phonetic cues.⁷ Therefore, it is in fact logically impossible to assert phonetic identity. In this dissertation, I argue that if two phonological entities are identical in all important phonetic cues, phonetic identity can be established. In Mandarin languages, f0, duration and intensity are recognized as the main phonetic cues for lexical tone. There is evidence that native speakers of Mandarin languages rely solely on f0 contour to distinguish tones (Howie, 1976; Tupper et al., 2020), f0 contour identity can arguably indicate phonetic identity. Besides F0 contour, there is experimental evidence that native speakers can make use of only intensity or only duration to distinguish tones (Fu & Zeng, 2000; Whalen & Xu, 1992 for intensity; Blicher et al., 1990 for duration). If two tones are identical in these three dimensions, phonetic identity could be argued for, at least functionally speaking. I will show this is indeed the case in Huai'an in Chapter 3.

If the Huai'an case can in fact be established, this will again post a challenge to the *Standard generative view of Phonology*. It is not clear why phonologically different forms can be implemented in exactly the same fashion in the phonetics under such framework. And more

⁷ Thanks to Karthik Durvasula for many discussions about this issue.

importantly, it is not clear how native speakers can perceive identical phonetic signal as different phonological representations. However the Huai'an case is also compatible with *Classic generative view of Phonology*, where linguistic performance is viewed as a multi-factorial problem (Chomsky, 1964, 1965; Schütze, 1996; Valian, 1982; Warner et al., 2004; inter alia). Performance factors can logically bring two different phonological forms together in the phonetics.

1.4 Organization of the Dissertation

The dissertation will be organized as follows. In Chapter 2, I will explore the phenomenon of incomplete neutralization. And I will present relevant data from Tone 1 and Tone 4 sandhi processes in Huai'an, both of which crucially participate in feeding orders to trigger other tone sandhi processes, to argue that phonetically incomplete neutralization can still be phonologically complete. Therefore, phonologically identical form can have different phonetic distributions.

In Chapter 3, I will compare Tone 4 sandhi at different levels of prosodic hierarchy in Huai'an to show that derived Tone 3s from Tone 4 sandhi at lexical and post-lexical levels are phonetically indistinguishable. Despite the identity in phonetics, they have different phonological behaviors with regard to the ability to trigger another tone sandhi process. Therefore, phonetically identical form can have different phonological behaviors. And phonologically different forms does not necessarily correspond with different phonetic distributions.

In Chapter 4, I will provide my explanations for incomplete neutralization, and more generally why phonetically identical form can have different phonological behaviors. I will present a new experiment in Chapter 5 to support my proposal that separate explanations are needed for incomplete neutralization with small effect size and incomplete neutralization with large effect size.

Chapter 6 will be the general discussion and Chapter 7 will be the conclusion.

CHAPTER 2 THE PHENOMENON OF INCOMPLETE NEUTRALIZATION⁸

2.1 Introduction

The suspicion of the very existence of the incomplete neutralization has never ceased since such phenomenon is first discovered, and one trivial but widely-adopted solution has been to simply deny that such an phenomenon can be caused by grammatical knowledge (Dinnsen & Charles-Luce, 1984; Fourakis & Iverson, 1984; Manaster Ramer, 1996; Warner et al., 2004; inter alia). To support this claim, several criticisms have been raised against previous experimental designs as well as the interpretation of the results. One main criticism is whether the observed phenomenon of incomplete neutralization is due to task effects. Among these effects, the most discussed one is orthography. It has been noticed by many researchers (Fourakis & Iverson, 1984; Manaster Ramer, 1996; inter alia) that in the seminal work of Port and O'Dell (1985), participants were presented stimuli orthographically where minimal pairs were always in contrast. Native speakers of German may have hypercorrected and produced unnatural speech to match the forms of orthography. This suspicion becomes especially disturbing when Warner et al. (2004) showed a significant production difference in words that are identical in underlying representations but differ in orthography in Dutch.⁹ To circumvent the interference of orthography, two methods have been employed, namely changing the experimental paradigm and looking at languages where the relevant phonological contrast is not reflected orthographically.

In line with changing the experimental paradigm, Fourakis and Iverson (1984) employed a unique strategy aimed at concealing the morphological forms that native speakers of German are

⁸ This chapter largely comes my collaborative work with Karthik Durvasula, see reference (Du & Durvasula; accepted). ⁹ In support of my larger point in this paper, I would like to point out that their observation in fact shows how powerful performance (i.e., non-phonological) factors can be in accounting for phonetic manifestations.

supposed to produce, so the influence of orthography is expected to decrease. The participants were instead presented auditorily with the conjugated form where the contrast is maintained in both underlying and surface representations and asked to decompose and produce the bare form where the incomplete neutralization is expected to happen. Through this paradigm, Fourakis and Iverson camouflaged the task as a morphological exercise to distract the participants to elicit more natural pronunciations. Interestingly, the effect of incomplete neutralization was not observed, and Fourakis and Iverson concluded that the previously found incomplete neutralization was actually a task effect. Using a different strategy, Jassem and Richter (1989) asked participants to answer questions designed to elicit the target words in Polish, and observed no evidence of incomplete neutralization. However, by implementing the same strategy as Fourakis and Iverson (1984) and increasing the statistical power with more speakers and more test minimal pairs, Roettger et al. (2014) found an effect of incomplete neutralization. However, it is worth noting that, as Roettger et al. (2014) themselves pointed out, the strategy employed by Fourakis and Iverson can incur a potential artifact of phonetic accommodation. In the experimental paradigm, the participants hear the conjugated form where neutralization cannot happen and the voicing contrast is present, and have to produce the form where neutralization does happen. In such a paradigm, the observed effect of incomplete neutralization may be due to the participants mirroring vowel duration differences in the stimulus recordings they heard of the conjugated forms. When Roettger et al. (2014) controlled for this confound in one of their experiments, they only found a very small, nonsignificant effect (< 3 ms) in the right direction. This suggests that there might indeed be no clear evidence for incomplete neutralization even in their well-powered study. To sum up, this general strategy to solve the problem of orthography by changing the task performed by the participants leads to very weak evidence (if that) for the presence of incomplete neutralization.

A second method employed to overcome task effects related to orthography has been to use a language where the crucial contrast is not marked in the orthography. For example, Catalan has been claimed to be a language that has a devoicing process but does not reflect an underlying voicing contrast orthographically under any phonological conditions, and Dinnsen and Charles-Luce (1984) did not observe any evidence of incomplete neutralization in the devoicing process of Catalan. However, Charles-Luce and Dinnsen (1987) later reanalyzed only a subset of their original data, and they found incomplete neutralization in the cue of voicing into closure. One purpose of this reanalysis was to avoid a potential unclarity with regard to underlying representations of some stimuli. The other purpose was to avoid the potential influence from phonological gemination, which was caused by interaction between some stimuli and carrier sentences. Finally, it is not morthography, so the real situation is more complicated and Catalan cannot simply be treated as a language that does not mark underlying voicing contrast in orthography, 1962; Manaster Ramer, 1996).

In another case, Braver and Kawahara (2016) observed a putative case of incomplete neutralization in Japanese. Most of their stimuli were presented in Chinese characters (Kanji), which is an orthographical system that is commonly used in Japanese but only has a very weak connection with pronunciation.¹⁰ Although most Chinese characters were originally created by combining a part that indicates pronunciation and a part that indicates meaning (Yang, 1995), the connection between characters and pronunciation is largely obscured by historical sound change

¹⁰ In Braver and Kawahara's experiment, 10 out of 13 sets of stimuli were presented only in Chinese characters while the other 3 sets partially contained Kana (2 in Katakana and 1 in Hiragana). In current usage, Kana refers to two syllabaries where each character represents a mora, which is an onset-vowel combination or a coda or the second half of a long vowel in Japanese phonology. It is also worth noting that geminates and some long vowels are marked with diacritics in Kana system. Braver and Kawahara observed incomplete phonetic neutralization for each set of stimuli.

and character change (B. Huang & Liao, 2017). In Japanese, most Chinese characters are used to represent both borrowed words from China (Sino-Japanese lexicon) and words that are originated in Japan (Yamato lexicon) (Itô & Mester, 1999; Japan Broadcasting Corporation, 1998), and the resulting multiple pronunciations (onyomi and kunyomi) of many Chinese characters can only further weaken the connection between Chinese characters and pronunciations. So, it is hard to imagine that Japanese speakers hypercorrected based on Chinese characters, and Braver and Kawahara still appeared to observe incomplete neutralization in monomoraic prosodic word lengthening process.

To sum up, although the case of Catalan is controversial, the case of Japanese provides good evidence that at least in some languages, the observed incomplete neutralization is not caused by orthographic knowledge.

Another source of criticism of incomplete neutralization is that the observed effect size is typically quite small. As a consequence, such a small effect size has been argued to likely not be functionally significant and therefore not in need of a grammatical explanation (Dinnsen & Charles-Luce, 1984; Mascaró, 1987; Warner et al., 2004).¹¹ For example, among the phonetic cues examined by Port and O'Dell (1985), preceding vowel duration before underlying voiced stops was only about 15 ms longer than that before underlying voiceless stops, voicing into closure of derived voiceless stops was only 5 ms longer than that of underlying voiceless stops, and duration of aspiration noise before underlying voiceless stops was only 15 ms longer than that of effect sizes were also found in Polish (Jassem & Richter, 1989; Slowiaczek & Dinnsen, 1985), Dutch (Warner et al., 2004), and two other studies on German (Piroth & Janker,

¹¹ To be clear, as pointed in Du and Durvasula (accepted), it is not claimed that the effect size must be large for incomplete neutralization that is caused by grammatical knowledge. It is only recognized here that it is a reasonable concern that incomplete neutralization with a small effect size may not even be captured by grammatical knowledge and therefore may not be able to pose any challenges to traditional formal phonology.

2004; Roettger et al., 2014). To summarize the discussion on the criticisms on incomplete neutralization, the debate on the existence of incomplete neutralization is still pretty much ongoing, especially with respect to the issue of effect size.

In this dissertation, as mentioned above, I will argue, using data from Huai'an, that incomplete phonetic neutralization can stem from phonologically complete neutralization. By using Huai'an, I avoid the orthographic confound discussed above as the stimuli can be presented in Chinese characters, an orthographical system that only has a weak connection with pronunciation (note, this is similar to the Japanese case I discussed above.). Furthermore, the language allows me to argue that effect sizes are tangential to the issue of phonological neutralization. Anticipating the results, I will show that despite there being a rather large phonetic difference with respect to incomplete phonetic neutralization, there is clear evidence that the relevant processes are phonologically categorically neutralizing as evidenced by the fact that their outputs feed other sandhi processes.

2.2 The issue of phonological neutralization versus phonetic implementation

As introduced in Section 1.2.1, the definition of incomplete neutralization is two-fold, which includes phonological neutralization and phonetically incomplete neutralization. An issue of many previous studies of incomplete neutralization is that researchers do not typically show evidence that the examined processes are truly phonological neutralization, as opposed to phonetic implementation (Cohn, 1993; Dunbar, 2013). Under the categorical view of phonological representations, phonological neutralization entails there is a change from one phonological category to another phonological category, while phonetic implementation does not result in any categorical changes. To give an example, it is assumed by Port and O'Dell (1985) and other

previous studies on German that the devoicing process results in a voiceless obstruent category in the phonological surface form.¹² However, there is no clear evidence, especially evidence from phonological behavior, that shows a derived voiceless obstruent is actually neutralized with the underlying voiceless obstruent in the phonology. If Dunbar's (2013) suspicion that the word-final devoicing in German is actually a phonetic implementational process turns out to be valid, then the so-called 'devoiced obstruent' at the right edge of prosodic word remains phonologically unchanged and still belong to the same 'voiced' category with voiced obstruents in other positions. As a result, it would not be surprising according to the *Standard generative view of Phonology* that the so-called 'devoiced obstruent' is phonetically different from an underlying voiceless obstruent since they are phonologically different, i.e., different in the surface representations.

To the best of my knowledge, the only careful previous study that attempted to establish phonological neutralization using evidence from phonological behavior is Braver and Kawahara's study on the lengthening of prosodic words in Japanese (Braver & Kawahara, 2016). Previous literature generally argue that all prosodic words (ω) in Japanese have to be at least bimoraic (Braver, 2019; Braver & Kawahara, 2016; Itô, 1990; Itô & Mester, 2003; Mester, 1990; Mori, 2002; Poser, 1990). The evidence mainly comes from processes where monosllabic prosodic words are avoided. These processes include word-formation patterns, including nickname formation, geisha-client name formation, loanword abbreviation, verbal root reduplication, scheduling compounds and telephone number recitation. To take nickname formation process as an example. A full name should be truncated to at least two morae long and then a suffix '-chan' should be added as shown in (5) (Data from Braver, 2019). The name Wasaburoo and Kotomi can

¹²As pointed in Du and Durvasula (accepted), to be fair to them, this is likely what they assumed since many phonologists have claimed as much in the extant literature.

be truncated to at least two moras long as in (5a) and (5b), which means a shortened form consisting of only one mora is ungrammatical.

(5) a.	Wasaburoo	(full name)	b. Kotomi	(full name)
	Wasa(-chan)	(2 moras)	Koto(-chan)	(2 moras)
	*Wa(-chan)	(1 mora)	Koc(-chan)	(2 moras)
			*Ko(-chan)	(1 mora)

As a result of the bimoraic constraint, an *underlying* monomoraic prosodic word has to lengthen to be bimoraic in Japanese. In relation to this, Braver and Kawahara (2016) showed that this neutralization process is incomplete phonetically, i.e. a derived/lengthened bimoraic prosodic word is still shorter than an underlying bimoraic prosodic word.

However, monosyllabic prosodic words being avoided in certain processes does not necessarily mean prosodic word of such length being generally prohibited. An analogy can be drawn in Standard Mandarin. In Standard Mandarin, monosyllabic prosodic unit is strictly prohibited in the process of adapting loanword of Japanese names as shown in (6). To conform to this constraint, prosodic reconstruction occurs wherever there are monosyllabic words in the original Japanese names as shown in (6a), (6b), (6d), (6e) and (6f). The parentheses indicate prosodic boundaries, which are boundaries of prosodic words in these cases. The prosodic structures on the left of arrows represent the original Japanese prosodic structures with segments being adapted into Standard Mandarin, and the prosodic structures on the right of arrows represent the prosodic structures employed in Standard Mandarin. All possible nonrecursive prosodic structures in four-character Japanese names have been explored as on the left of arrows.

- (6) Prosodic reconstruction of Japanese names in Standard Mandarin
 - a. $\omega(\text{teye})_{\omega} \omega(\text{uei ian nai})_{\omega} \rightarrow \omega(\text{teye uei})_{\omega} \omega(\text{ian nai})_{\omega}$

[堀 未央奈; ほりみおな; HORI, Miona; Singer Idol]

b. $\omega(t_{\xi^ha\eta} ku pu)_{\omega} \omega(t_{\xi^ha\eta})_{\omega} \rightarrow \omega(t_{\xi^ha\eta} ku)_{\omega} \omega(pu t_{\xi^ha\eta})_{\omega}$

[長谷部 誠; はせべ まこと; HASEBE, Makoto; Football Player]

c. $\omega(\mathfrak{si} \ i \varepsilon) \omega \ \omega(\mathfrak{ts}^{hi} \ lai) \omega \rightarrow \omega(\mathfrak{si} \ i \varepsilon) \omega \ \omega(\mathfrak{ts}^{hi} \ lai) \omega$

[西野 七瀬; にしの ななせ; NISHINO, Nanase; Singer Idol]

d. $\omega(ji\epsilon)\omega\omega(ia \, son)\omega \rightarrow \omega(t\epsilon i \, a \, son)\omega$

[堺 雅人; さかい まさと; SAKAI, Masato; Actor]

e. $\omega(\operatorname{fu} \operatorname{yen})_{\omega} \omega(\operatorname{iau})_{\omega} \rightarrow \omega(\operatorname{fu} \operatorname{yen} \operatorname{iau})_{\omega}$

[福原 遥; ふくはら はるか; FUKUHARA, Haruka, Actress and Voice Actress]

f. $\omega(s \ni n) \omega \omega(i) \omega \rightarrow \omega(s \ni n i) \omega$

[森 毅; もり つよし; MORI, Tsuyoshi; Mathematician]

However, monosyllabic prosodic unit can still appear in the surface in many cases. An example is shown in (7), for some trisyllabic words, the Tone 3 sandhi pattern clearly shows that a prosodic boundary that separates the grammatical word into two prosodic words.

(7) [mi [lau şu]]

'Mickey Mouse.'

UR 333

SR $\omega(3)\omega \omega(23)\omega$

It is worth distinguishing two concepts at this point, namely Nonpreferred unit size and Prohibited unit size. Nonpreferred unit size means this type of prosodic unit is phonologically not preferred by native speakers and should be avoided as much as possible, but this type of prosodic unit can still appear in the surface due to intervention of other constraints rooted in phonology or syntax-phonology interface. In contrast, Prohibited unit size should not appear in the surface under any circumstances, which of course means that such prosodic units are not preferred by native speakers. These two concepts are very clear under Optimality Theory framework (Prince & Smolensky, 1993). Back to the Japanese case, if monosyllabic prosodic word is a Nonpreferred unit, then the constraint preventing such unit size should not be ranked undominated, at least in some cases. While if monosyllabic prosodic word is a Prohibited unit, then the constraint preventing such unit size should always be ranked undominated. These two concepts are obviously confused in previous literature on Japanese. And there is no clear evidence that monosyllabic prosodic word being necessarily a Prohibited unit as argued in previous literature.

Overall, even the Japanese case is doubtworthy. Therefore there is still no convincing evidence that the examined processes are truly phonological neutralization, as opposed to phonetic implementation (Cohn, 1993; Dunbar, 2013). And if incomplete neutralization exists still remains an open question. The current paper utilizes a different strategy of examining rules in feeding orders to establish phonological neutralization. The fact that the output of a process can trigger another process provides evidence that the first process results in complete neutralization in the phonology. Yet despite there being complete neutralization in the phonology, I will show that there is incomplete neutralization in the phonetics for each of the feeding processes in Huai'an tone sandhi processes.¹³ I will elaborate the feeding orders in Huai'an in Section 2.3 with more background information. And Section 2.4 and Section 2.5 will present the two experiments I have run based on two feeding orders in Huai'an.

2.3 Background of Huai'an Mandarin

Huai'an belongs to the Jianghuai Guanhua Group (Lower Yangtze Mandarin) of the Mandarin language family. Native speakers are mainly from Huai'an city, which is located in the northern part of Jiangsu Province (Li, 1989). Huai'an has four phonemic tones, labelled as Tone 1, Tone 2, Tone 3 and Tone 4 (Jiao, 2004; Y. Wang & Kang, 2012). Following the tradition of tone description in Chinese languages, in Table *1*, the four tones are given in tone letters using a scale of 1 to 5 where 1 is the lowest f0 and 5 is the highest f0 and followed by a contour description in words (Chao, 1930).¹⁴ The tonal contours of phonemic tones in isolation are given in Figure 2. The speaker (male, age: 53) pronounced 4 repetitions of four monosyllabic morphemes that share the same segmental content ([so]) and only contrast in the tone on the vowel. f0 was extracted only

¹³ As pointed out by an anonymous reviewer of the paper (Du & Durvasula, accepted), Ernestus et al. (2006) showed that Dutch has an optional progressive voice assimilation process across word boundary for obstruents. A word-final devoicing process feeds into the assimilation process and causes the initial obstruent of the next word to become voiceless, which suggests that the devoiced obstruent in the preceding word belongs to same phonological category with its underlyingly voiceless counterpart. In relation to this, Ernestus and Baayen (2006) showed in a separate experiment that there is incomplete phonetic neutralization in word-final devoicing process in Dutch (see also Warner et al., 2004). Further study is needed to show that a devoiced obstruent that can trigger another devoicing process is actually incompletely neutralized in the phonetics. If this is indeed the case, Dutch will serve as another clear case of incomplete phonetic neutralization of complete phonological neutralization.

¹⁴ Note, Huai'an phonemic tones are different from those in Standard Mandarin. In Huai'an, Tone 1 is a high falling tone and Tone 4 is a high level tone. While in Standard Mandarin, Tone 1 is a high level tone (tone letter: 55) and Tone 4 is a falling tone (tone letter: 51). Tone 2s and Tone 3s in Huai'an and Standard Mandarin are largely similar.

from the vowel at 5% steps with a script in Praat. However, it is worth noting that the tonal contours in isolation for Mandarin tones have been noticed to be quite different when compared with their counterparts in context (Jongman et al., 2006; Shen, 1990; Xu, 1994, 1997; inter alia). So, I expected the same kind of differences in my experiments where tones are pronounced in sentences.

Phonemic tone	Tone letter	Contour description
Tone 1	42	high falling
Tone 2	24	high rising
Tone 3	312	low/low rising
Tone 4	55	high level

Table 1: Descriptions of phonemic tones in Huai'an



Figure 2: Tonal Contour of Phonemic Tones in Huai'an

For the examples in the rest of the paper, to make it easier on the reader, I will only use the T plus tone number to refer to tones. For example, 'T3' refers to Tone 3. I will however continue to use the full form Tone 3 in the text.

The three tone sandhi rules that are related to the current paper are shown in (8). At post-lexical level, the low-register Tone 3 sandhi is mandatory only when the syllable that undergoes tone sandhi and the following syllable that triggers tone sandhi are in the same phonological phrase. Such a phonological formulation still allows Tone 3 sandhi to optionally apply when the two syllables are not in the same phonological phrase, therefore a Tone 3 sequence can still appear in the surface.¹⁵ In contrast, the high-register Tone 1 and Tone 4 sandhis are always optional and only applicable when the two syllables involved belong to the same phonological phrase. It has been noted in previous literature that tone sandhi patterns in many Mandarin languages are sensitive to prosodic structures (M. Y. Chen, 2000), and Huai'an is also in this group. One piece of evidence comes from variation of Tone 3 sandhi as shown in (9). Similar to Standard Mandarin (Duanmu, 2007), for trisyllabic right-branching sentences where all syllables are underlyingly Tone 3, two possible surface representations are possible, namely 'Tone 2 Tone 2 Tone 3' and 'Tone 3 Tone 2 Tone 3'. The variation is best accounted for through different prosodic structures. The analysis is further supported by perceived pause length between syllables. According to anecdotal reports from native speakers of Huai'an, for the surface representation of 'Tone 2 Tone 2 Tone 3', the pause length between each pair of immediately adjacent syllables is very short, suggesting all three syllables are in the same phonological phrase. In contrast, for the surface representation of 'Tone

¹⁵ The optionality of Tone 3 sandhi application across phonological phrase may be due to planning effects or other performance factors. Essentially, as per *Classic generative view of Phonology*, a categorical *competence* grammar can still manifest gradiently in *performance*.

3 Tone 2 Tone 3', there is a noticeable long pause between the first syllable and the other two syllables, suggesting a phonological phrase boundary in that position.

Furthermore, Tone 3 undergoes tone sandhi to become Tone 2, and this Tone 3 sandhi process can only happen when immediately preceding Tone 3 (underlying or derived).¹⁶ As dissimilation processes, the tone sandhis in Huai'an can be straightforwardly explained by the Obligatory Contour Principle (Leben, 1973; McCarthy, 1986; Yip, 2002; inter alia). However, some researchers reject the use of the Obligatory Contour Principle as the motivation of tone sandhi processes in Mandarin languages (Duanmu, 1994, 2007; inter alia). I will not address this debate since it is tangential to the main argument of this paper.¹⁷

(8) Relevant tone sandhi rules in Huai'an Mandarin (Y. Wang & Kang, 2012)

- a. Low-register tone sandhi¹⁸
 - i. Tone 3 sandhi: T3 + T3 \rightarrow T2 + T3
- b. High-register tone sandhi [optional processes]

i. Tone 1 sandhi: T1 + T1 \rightarrow T3 + T1

ii. Tone 4 sandhi: T4 + T4 \rightarrow T3 + T4

¹⁶ I reiterate the usage of 'underlying' and 'derived' in this paper: Consistent with other places in this paper, 'underlying Tone 3' means surface Tone 3 that maps from underlying Tone 3, and 'derived Tone 3' means surface Tone 3 that maps from underlying Tone 1 or Tone 4.

¹⁷ Depending on different proposed representations of the tones (M. Y. Chen, 2000; Duanmu, 2007; Yip, 2002; to name a few), the motivation of tone sandhis can be different. I will discuss the assumed tonal representation for this paper in Section 6.1 that a Mandarin tone is a single phonological unit.

¹⁸ Register is a tonal feature first proposed by Moira Yip (1980) and then widely adopted in the literature of Chinese tonal phonology. Here I simply use the feature to distinguish Tone 3 sandhi from other tone sandhi processes in Huai'an.

(9) Tone 3 sandhi variation in Huai'an Mandarin (Du, 2021)

a.	o 1sg	mε teiəu buy wine	b.	li tsa Mr. L	o ma i find horse
	ʻI buy	wine.'		'Mr.	Li finds horses.'
	UR	T3 T3 T3		UR	T3 T3 T3
	SR1	_φ (T2 T2 T3) _φ		SR1	_φ (T2 T2 T3) _φ
	SR2	φ(T3)φ φ(T2 T3)φ		SR2	

Note: '\phi' means phonological phrase.

Crucially, the Tone 3 output of the high-register tone sandhi processes feeds the low-register Tone 3 sandhi process as in (10). Since high-register tone sandhis are optional and Tone 3 sandhi is also optional given different possible prosodic structures for utterances in (10), multiple surface representations are possible for both examples.
- (10) Feeding Order in Huai'an Mandarin (boldface represents the locus of a potential tonal change due to the relevant tone sandhi process; the data is from the authors)
 - a. Tone 1 sandhi feeds Tone 3 sandhi

u Mr. Wu	ku estii	fən mate score		
'Mr. Wu estim	nates	scores.'		
UR		T3 T1 T1		
Tone 1 sandhi		T3 T3 T1		(or) T3 T1 T1
Tone 3 sandhi		T2 T3 T1	(or) T3 T3 T1	T3 T1 T1
SR		T2 T3 T1	(or) T3 T3 T1	(or) T3 T1 T1
Corresponding	g			
Prosodic Struc	cture	_φ (T2 T3 T1) _φ	_φ (T3) _φ _φ (T3 T1) _φ	_φ (T3 T1 T1) _φ
				$\phi(T3)_{\phi} \phi(T1 T1)_{\phi}$
Tone 4 sandhi feeds Tone 3 sandhi				

u	to	zəm
Mr. Wu	chop	meat

'Mr. Wu chops meat.'

b.

UR	<u>T3 T4 T4</u>		
Tone 4 sandhi	T3 T3 T4		(or) T3 T4 T4
Tone 3 sandhi	T2 T3 T4	(or) T3 T3 T4	T3 T4 T4
SR	T2 T3 T4	(or) T3 T3 T4	(or) T3 T4 T4
Corresponding			
Prosodic Structure	_φ (T2 T3 T4) _φ	_♦ (T3) _♦ _♦ (T3 T4) _♦	_φ (T3 T4 T4) _φ
	_φ (T2) _φ _φ (T3 T4) _φ		φ(T3)φ φ(T4 T4)φ

The feeding relationships between each of the high-register tone sandhis and Tone 3 sandhi suggest that the high-register tone sandhis result in a Tone 3 category that is phonologically the same as an underlying Tone 3. This interpretation of the data remains the same given a parallel approach to phonology such as Optimality Theory (Prince & Smolensky, 1993). A usually employed markedness constraint for low-register tone sandhi in Mandarin languages is '*Tone3Tone3', which is based on the Obligatory Contour Principle. This constraint prevents adjacent Tone 3 syllables (C.-Y. Wang & Lin, 2011; Zhang, 1997; see also M. Y. Chen, 2000 for using this constraint in an implicit fashion). For this constraint to trigger the structural change in the first syllable (namely, Tone 3 \rightarrow Tone 2), the second syllable in /Tone 3 Tone 4 Tone 4/ or /Tone 3 Tone 1 Tone 1/ must surface as Tone 3. Consequently, an Optimality Theory analysis would also maintain the crucial categorical aspects of the feeding order that are focal for the current paper.

With regard to this interpretation of phonological identity between derived and underlying Tone 3s, concerns may be raised about application rates, especially when an underlying Tone 3 mandatorily triggers Tone 3 sandhi while a derived Tone 3 can only optionally trigger Tone 3 sandhi when the two types of Tone 3 syllables are the middle syllable of a trisyllabic utterance. The comparison is shown in (10) and (11). And some researchers may want to ascribe the difference in application rates to a difference between derived Tone 3 and underlying Tone 3 in the phonology, either as different phonological representations or as the same representations that are indexed to different variable processes. However, intervening factors are not controlled for when compared in this way. For Tone 3 sandhi to mandatorily apply before an underlying Tone 3, only a disyllabic phonological phrase is needed, which includes the two Tone 3 syllables. In (11), the established phonological phrase always includes the first two syllables to avoid the dispreferred Tone 3 sequence (C.-Y. Wang & Lin, 2011; Zhang, 1997; M. Y. Chen, 2000) and may extend to include the last syllable.¹⁹ These possible prosodic structures ensure that Tone 3 sandhi is mandatory before underlying Tone 3. In contrast, for Tone 3 sandhi to mandatorily apply before a derived Tone 3, at least a trisyllabic phonological phrase is needed where both high-register Tone 1/Tone 4 sandhi and low-register Tone 3 sandhi occur. In (10), for Tone 3 sandhi to mandatorily apply before derived Tone 3, the phonological phrase must include all three syllables. It is well recognized in the previous literature that a larger planning window required by a longer prosodic unit has more planning difficulty (Ferreira & Swets, 2002; Kilbourn-Ceron & Goldrick, 2021; Wagner et al., 2010; inter alia). The reason is the increasing burden on working memory, which can lead to speech errors or delays. Huai'an turns out to not be an exception. Previous experimental study on Tone 3 sandhi in Huai'an does support the existence of the effect of planning difficulty on longer prosodic units. As a result, native speakers prefer disyllabic prosodic units over trisyllabic prosodic units (Du & Lin, 2021). Due to the planning difficulty, the trisyllabic phonological phrase may not be established in (10), and if the alternative disyllabic phonological phrase includes only the last two syllables, Tone 3 sandhi may not occur before derived Tone 3. Overall, the difference in application rates comes naturally from the planning difficulty effect and does not need to be accounted for in the phonology.

As pointed out by two anonymous reviewers of the paper (Du & Durvasula, accepted), proponents of gradient phonological representation may argue that although both underlying and derived Tone 3s can trigger Tone 3 sandhi, they may still have different phonological

¹⁹ Here, purely for the sake of expository convenience and consistency with the general consensus in the Mandarin phonology literature, I assume that phonological domains can be constructed based not only on morphosyntactic structures, but also on phonological restrictions (C.-Y. Wang & Lin, 2011; Zhang, 1997; M. Y. Chen, 2000). This is a view that is also consistent with recent discussions of Match Theory (Itô & Mester, 2013; Selkirk, 2011; inter alia). My argument does not depend on this assumption, and can in fact be made without reference to the prosodic domains at all.

representations. By this analysis, the difference in application rates would be explained by the difference in the phonological representations. First, I would like to point out that any analysis that predicts application rates based on gradient phonological representations or phonetic similarity would have to be precise in accounting for not only cases where the process is triggered but also cases where the process is *not* triggered; namely, it would have to explain why only the derived Tone 3 shows a variation in application rates and not the underlying Tone 3, and not the other way around. Furthermore, it would have to account for the fact that any other tones that are phonetically similar (along the relevant dimensions) do not trigger the process. While an evaluation of such an analysis is not possible without a concrete specification of the proposal, I suspect that, to explain the difference between derived Tone 3 and underlying Tone 3, one will have to make reference to performance factors anyway. Relatedly, I appeal to the need to prioritize relatively simple categorical phonological representations when they are sufficient to account for the observed patterns (Occam's razor/law of parsimony); in our case, the difference in application rates can be accounted for by independently needed performance factors, namely planning, and therefore I need not complicate our understanding of the relevant phonological (tonal) representations. For this reason, I see the feeding rule interaction as evidence of complete phonological neutralization of the derived Tone 3 from Tone 1 and Tone 4 sandhi processes. Furthermore, I use the processes to probe the phonetic (acoustic) consequences of the neutralizing processes in the case of the derived Tone 3 that in turn trigger Tone 3 sandhi.

- (11) Application of Tone 3 sandhi before underlying Tone 3 in trisyllabic utterances
 - a. Tone 1 sandhi feeds Tone 3 sandhi

u Mr. Wu	po protect	tei car
'Mr. Wu prote	cts cars.'	
UR	T3	<u>T3 T1</u>
Tone 3 sandhi	T2	<u>T3 T1</u>
SR	T2	T3 T1
Corresponding	5	
Prosodic Struc	ture ∳(T2	T3 T1)

(T2 T3)

(T1)

b. Tone 4 sandhi feeds Tone 3 sandhi

u	tε	ciæ
Mr. Wu	catch	elephant

'Mr. Wu catches elephants.'

UR	T3 T3 T4

Tone 3 sandhi	T2 T3 T4

SR T2 T3 T4

Corresponding

Prosodic Structure $\phi(T2 T3 T4)\phi$

To further ensure the phonological equality of derived Tone 3 and underlying Tone 3, I only analyze the derived Tone 3 tokens that actually trigger Tone 3 sandhi in this paper, which allows me to have perfect surface minimal pairs in each of our experiments. By doing so, I also exclude the possibility that any identified incomplete phonetic neutralization patterns arise as a result of averaging the outcomes of an optional phonological process, since I only look at the cases where I have reason to believe that the process applied. Despite the categorical phonological behavior of the derived Tone 3 in Huai'an, in the next two sections, I will show that there is substantial incomplete phonetic neutralization of derived Tone 3 and underlying Tone 3 for the feeding orders involving Tone 1 sandhi and Tone 4 sandhi.

2.4 Experiment 1: Tone 1 sandhi

2.4.1 Participants

I recruited 11 native speakers of Huai'an Mandarin via personal relationships in Huai'an City. The age range was from 37 to 55 years. Among them, 8 self-identified as female, and 3 as male. Due to the language standardization trend in mainland China (Ramsey, 1989), young speakers in Huai'an are generally bilingual and are native speakers of both Huai'an and Standard Mandarin. To minimize the influence of Standard Mandarin, I recruited older speakers who are only fluent in Huai'an in this study. All the participants were born and raised in Huai'an City. None of them had participated in any linguistic studies before or heard about the concept of incomplete neutralization.

2.4.2 Stimuli

The stimuli were composed of trisyllabic sentences with each syllable forming a separate word, therefore it is ensured that the tone sandhi processes are post-lexical and completely productive. Also, only right-branching utterances as in (10) are employed simply because not enough leftbranching utterances could be constructed by me given the paradigm to be introduced immediately. The stimuli were divided into four sets as shown in (12). Furthermore, the third syllable was always Tone 1. The second syllable was one of the following possibilities: a) an underlying Tone 1 that optionally underwent Tone 1 sandhi to become Tone 3, b) an underlying Tone 3 that did not undergo any tone sandhi in this context. The first syllable was underlyingly a Tone 3 or a Tone 2. As a consequence of the possibilities in the second syllable, there were a few different possibilities for the first syllable, including: a) an underlying Tone 3 that could undergo Tone 3 sandhi to become Tone 2 with reference to the second syllable, b) an underlying Tone 2 that did not undergo any tone sandhi in this context. The four sets were only different in tonal patterns but not in segmental content. Furthermore, the crucial second syllable was always a voiceless unaspirated stop plus vowel sequence. Voiceless unaspirated stops were chosen to make sure that there is a consistent way to annotate the acoustic onset of the vowel by referring to the burst of the stop. The full stimulus list is summarized in APPENDIX A. It is worth noting that one character [搭] may be pronounced with the only checked tone in Huai'an (Jiao, 2004; Y. Wang & Kang, 2012), which is an allophone of Tone 4 and appears only on monomoraic syllables ending with glottal stop. I excluded all checked tone productions when extracting f0 information.

- (12) Four sets of stimuli in Experiment 1 [the syllables crucial for the current comparison are underlined and boldface]
 - a. underlying T3 following underlying T2: $/T2 T3 T1/ \rightarrow [T2 T3 T1]$
 - b. underlying T3 following underlying T3: /T3 T3 T1/ \rightarrow [T2 T3 T1]
 - c. derived T3 following underlying T2: $/T2 T1 T1/ \rightarrow [T2 T3 T1]$ or [T2 T1 T1]
 - d. derived T3 following underlying T3: $/T3 T1 T1 / \rightarrow [T2 \underline{T3} T1]$ or [T3 T1 T1]

Out of the above set of possibilities, the most crucial comparison is between two tones in the second syllable, namely, an underlying Tone 3 as in (12) and a derived Tone 3 as in the first possibility in (12). This particular comparison controls for the preceding surface context (derived Tone 2) and the following surface context (underlying Tone 1) and is therefore a perfect minimal pair. Furthermore, the two cases also show evidence that both tones are in fact categorically Tone 3 as they trigger Tone 3 sandhi on the preceding tone. Finally, as mentioned previously, the comparison allows me to exclude the possibility that any identified incomplete phonetic neutralization pattern arises as a result of averaging the outcomes of an optional phonological process. This is the crucial pair I will focus on in this experiment.

The set of possibilities also allows me to visually compare the derived Tone 3 against an underlying Tone 1 in the same surface context, as in the second possibility in (c [albeit, the preceding syllable in this case is an underlying Tone 2 instead of a derived Tone 2].

Each participant produced 4 repetitions of 24 test and 27 filler sentences at a natural speech rate, which means each participant read a total of 204 sentences. All stimuli were randomized for each participant.

2.4.3 Procedure

The experiment was conducted entirely in Huai'an city. Each participant was recorded by a trained research assistant using Audacity (Audacity Team, 2019) and a Popu Line BK USB microphone on a Lenovo laptop in a quiet room that was either located in the participants' home or workplace. The participants were told that the purpose of the study was to collect some general information on Huai'an. None of the participants reported noticing the minimal pairs or the real purpose of the study being on tones in the post-experimental interview. The participants were instructed to read at a normal speech rate using their everyday voice, and the stimuli were presented in Chinese characters. The participants were also encouraged to read through the stimulus list to be familiar with the reading materials before producing them.

2.4.4 Measurement

Using Praat (Boersma & Weenink, 2021), the recordings were manually annotated by the first author, who is a native speaker of Huai'an. An example is shown in Figure 3. Only the second syllable was marked and the annotation file had 6 tiers in total. The first tier marked the vowel of the second syllable for phonetic analysis. The first zero crossing at the beginning of the voicing of the target vowel and after the burst of the unaspirated stop was identified as the vowel onset, and the zero-crossing immediately following the vowel's final glottal pulse was identified as the vowel offset. All other tiers marked the whole second syllable to index phonological information and recording quality. The onset of the second syllable was marked just before the release burst of the initial stop and the offset of the second syllable corresponded with the offset of the nuclear vowel. The second tier indicated the whole sentence in Pinyin, which is the official Romanization system for Chinese characters in China. The third tier was the tone sandhi condition where 'yes' meant the

second syllable had undergone tone sandhi and 'no' meant it had not, the fourth and fifth tiers had the underlying tones and surface tones information respectively and the last tier had the quality of the recording. I only used productions that were marked 'good' in the analysis. The reasons that productions were not marked as 'good' included background noise, speech errors, any long delay while producing the utterance, and checked tone pronunciation. f0 was extracted only from the vowel at 5% steps with a script in Praat.



Figure 3: Annotation scheme of Experiment 1 (Tone 1)

To compare across different speakers and different vowels, z-score transformation was performed for each vowel of each speaker based on Hz scale (Laplace, 1820; Lobanov, 1971).

2.4.5 Results and statistical modelling

All data analyses in this paper were performed in R (R Core Team, 2021) using the **tidyverse** suite of packages (Wickham et al., 2019). The statistical modelling was done using the **lme4** package (Bates et al., 2021).

The number of tokens for each possible combination of Underlying Representation and Surface Representation is summarized in Table 2. The application rate of Tone 3 sandhi before underlying Tone 3 is 97.2% while the application rate before derived Tone 3 is 74.0%.²⁰ 71 tokens were not marked as 'good' and excluded, which accounts for 6.7% of all test stimuli.

UR	SR	Number of tokens
T2T3T1	T2T3T1	259
T3T3T1	T3T3T1	7
T3T3T1	T2T3T1	242
T2T1T1	T2T1T1	74
T2T1T1	T2T3T1	167
T3T1T1	T3T1T1	59
T3T1T1	T3T3T1	46
T3T1T1	T2T3T1	131

Table 2: Number of Tokens for UR and SR combination

²⁰As mentioned before, the difference in application rates does not necessarily inform me of any differences in phonological representations since the difference is accountable through independently needed mechanisms, namely the difficulty in planning longer prosodic units.

The z-score transformed f0 contours on the crucial second syllable are shown in Figure 4. As a reminder, the crucial comparison is between a derived Tone 3 and an underlying Tone 3 after derived Tone 2s in the same surface context --- the context establishes that both the Tone 3s are categorically Tone 3 as they trigger Tone 3 sandhi. I also present the tone contour for an underlying Tone 1 in the same surface context for visual comparison with the two crucial Tone 3s.



Figure 4: Contours comparison of the second syllable in Experiment 1 (Tone 1)

(Error bars indicate standard error)

Based on the visual inspection of the data, the derived Tone 3 seems to start as an underlying Tone 3 and ends as an underlying Tone 1. And the contour shape of the derived Tone 3 is close to that of an underlying Tone 3. Furthermore, the comparison between underlying Tone 3 and derived Tone 3 clearly shows that the neutralization is incomplete.²¹

²¹ To further address the concern that incomplete neutralization patterns identified in Experiment 1 may arise as a result of averaging the outcomes of an optional phonological process, the distributions of underlying Tone 1, derived Tone 3, and underlying Tone 3 are shown for each time step in APPENDIX E. Crucially, the derived Tone 3 distribution is generally uni-modal, and distinct from the other two distributions, across the time-steps. Thus, there is no evidence of an averaging artifact over optional surface representations for the derived Tone 3 case.

For the purposes of statistical modelling, I used just the two-group factor (underlying Tone 3 vs. derived Tone 3), and ignored underlying Tone 1, in order to simplify the modelling and address only the crucial question of whether or not the underlying and derived Tone 3s have incompletely neutralized. The results turn out to support the observation that the neutralization is indeed incomplete phonetically.

In dealing with time course data, traditional techniques like t-tests and ANOVA have to divide continuous time into multiple time bins and therefore have to make multiple comparisons. This method has been argued by Mirman (2017) to be problematic for increasing the risk of 'false positives'. Since each time bin incurs the nominal 5% false positive rate implied by 'p < 0.05', overall, the false positive rate with multiple time bins and multiple comparisons will be much higher than a single comparison.

To solve this problem, multiple analysis methods have been developed, which includes Smooth Spline Analysis of Variance (SS-ANOVA) (Y. Wang, 1998), Generalized Additive Model (GAM) (Hastie & Tibshirani, 1990) and Growth Curve Analysis (GCA) (Mirman, 2017; Mirman et al., 2008). In this paper, I follow S. Chen et al. (2017), in modelling f0 contours using Growth Curve Analysis. Growth Curve Analysis uses multilevel linear regression to avoid multiple comparisons and has been argued to be a useful modelling technique in different fields (Baldwin & Hoffmann, 2002; McArdle & Nesselroade, 2003; inter alia). To apply Growth Curve Analysis in Huai'an tones, I started with a simple model as in (13) (Mirman et al., 2008).

(13)
$$Y_{ij} = (\gamma_{00} + \zeta_{0i}) + (\gamma_{10} + \zeta_{1i}) * Time_{ij} + \varepsilon_{ij}$$

Here i is the ith f0 (z-score transformed) contour and j is the jth time point, and Y_{ij} is the f0 (zscore transformed) value for ith contour at jth time point. you is the population average value for the intercept, ζ_{0i} is individual variation on the intercept, γ_{10} is the population average value for the fixed effect of time, ζ_{1i} is individual variation on the fixed effect of time and ε_{ij} is the error term.²² To optimize the model for the data, I employed higher-order polynomial functions, and allowed individuals to vary on each term only when those terms reached significance according to chisquare likelihood ratio tests (S. Chen et al., 2017; S. Chen & Li, 2021; inter alia). In Mandarin languages, a Tone Bearing Unit (TBU), which is assumed to be the syllable or the rhyme or the nucleus of the rhyme, has been widely argued to be associated with at most three tonal targets (Bao, 1990, 1992; Duanmu, 1994; inter alia). Therefore, the most complex tones can only have one change of direction, which will produce U-shaped contours, such as high-low-high and low-highlow. To conform to this observation, I only considered up to second-order functions to ensure that the final model is not more complex than a U-shape contour. Also, orthogonal polynomials were used to make sure that the linear and quadratic terms were not correlated (Mirman, 2017). After optimizing the model by including all significant terms, I first treated underlying Tone 3 and derived Tone 3 as the same and modelled them as one single contour to get Model 1. Then I built models that treat them as different, namely, models that include a tone sandhi condition (underlying Tone 3 vs. derived Tone 3) to do model comparison. Based on Model 1, tone sandhi condition is first allowed to affect only intercept to get Model 2. Then tone sandhi condition is allowed to affect both intercept and linear term to get Model 3. Finally, tone sandhi condition is

 $^{^{22}}$ Note, the individual variation terms, $\zeta 0i$ and $\zeta 1i$, are akin to the random intercept by participant, and random slope of time by participant.

allowed to affect all fixed effects, which include intercept, linear term and quadratic term, and the outcome is Model 4. Chi-square likelihood ratio test was used to determine whether two minimally different models differ significantly.

The result shows that the difference between underlying Tone 3 and derived Tone 3 is in fact supported by model comparisons. The addition of a tone sandhi condition improves the model on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=331.81$, p<0.01), on the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=118.34$, p<0.01) and on the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)=14.99$, p<0.01). Figure 5 shows how the full model (Model 4) fits the observed data. The parameter estimates for the full model are summarized in Table 3.



Figure 5: Observed data and Growth Curve Model fits for derived and underlying Tone 3

	Estimate	Std. Error	t	р
Intercept	0.07	0.06	1.13	0.28
Linear	-16.01	2.14	-7.47	< 0.01
Quadratic	9.57	1.59	6.04	< 0.01
Tone Sandhi: Intercept	-0.50	0.03	-19.68	< 0.01
Tone Sandhi: Linear	-13.64	1.24	-10.99	< 0.01
Tone Sandhi: Quadratic	4.80	1.23	3.90	< 0.01

 Table 3: Parameter estimates of the full model (Model 4) with the assumption of tone sandhi

 affecting every fixed effect (baseline: derived Tone 3)

Moreover, the effect size of incomplete neutralization is large in Tone 1 sandhi. The mean difference in f0 between underlying Tone 3 and derived Tone 3 across all steps is 18 Hz, which is more than 2 times the Just Noticeable Difference of f0 value (7 Hz) for Mandarin speakers (Jongman et al., 2017). Furthermore, across the last 10 steps (step 11 to step 20), the f0 difference is over 22 Hz, which is more than 3 times the Just Noticeable Difference. The f0 difference (f0 of derived Tone 3 - f0 of underlying Tone 3) of each step is summarized in Table 4. Recall that the underlying premise from those who criticize the small effect size of incomplete neutralization is that if the differences were robust and large in size, the existence of such an effect should be accepted as functionally relevant.²³ According to that standard, the case of Huai'an Tone 1 sandhi is clearly a case of phonetically incomplete neutralization.

²³ I acknowledge that it is not entirely clear to me what is intended by the use of phrases such as "functional relevance", since many aspects of linguistic behavior might be important to the speaker/listener while not stemming from the grammar per se; however, I retain the phrase here to reflect the terminology in the sub-field.

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	-5	11	22
1	-2	12	25
2	2	13	26
3	4	14	29
4	7	15	30
5	12	16	32
6	12	17	30
7	11	18	28
8	15	19	29
9	22	20	25
10	21		

Table 4: f0 difference of each step in Experiment 1 (Tone 1)

To show that the pattern is not unique to Tone 1 sandhi and to extend the scope of the current study, I ran a second experiment on Tone 4 sandhi process in Huai'an.

2.5 Experiment 2: Tone 4 sandhi

2.5.1 Participants

I recruited 20 native speakers of Huai'an Mandarin also via personal relationships in Huai'an City. The age range was from 33 to 57 years old. Again, to minimize the influence of Standard Mandarin, I avoided younger speakers in this study. Among them, 16 self-identified as female, and 4 as male. 5 speakers had also participated in Experiment 1. The interval between Experiment

1 and 2 was about 7 months; the 5 participants from Experiment 1 failed to guess and were not told the purpose of Experiment 2. Like Experiment 1, all the participants were born and raised in Huai'an City. Other speakers have not participated in any linguistic studies before or heard about the concept of incomplete neutralization.

2.5.2 Stimuli

The stimuli were organized in the same way as in Experiment 1. The four sets of trisyllabic sentences are shown in (14), and the full stimulus list is summarized in APPENDIX B.

- (14) Four sets of stimuli in Experiment 2 [the syllables crucial for the current comparison are underlined and boldface]
 - a. underlying T3 following underlying T2: /T2 T3 T4/ \rightarrow [T2 T3 T4]
 - b. underlying T3 following underlying T3: /T3 T3 T4/ \rightarrow [T2 T3 T4]
 - c. derived T3 following underlying T2: $/T2 T4 T4 / \rightarrow [T2 T3 T4]$ or [T2 T4 T4]
 - d. derived T3 following underlying T3: $/T3 T4 T4 / \rightarrow [T2 T3 T4]$ or [T3 T4 T4]

As with Experiment 1, the crucial comparison is between two tones in the second syllable, namely the underlying Tone 3 in (14b) and the derived Tone 3 as in the first possibility in (14d). This comparison allows me to control for the surface context, while also establishing that the two tones are indeed categorical Tone 3s since they trigger Tone 3 sandhi on the preceding tone. Furthermore, as mentioned previously, the comparison allows me to exclude the possibility that any identified incomplete phonetic neutralization pattern arises as a result of averaging the outcomes of an optional phonological process.

The set of possibilities also allows me to look at an underlying Tone 4 in roughly the same surface context, as in the second possibility in (14c), for visual comparison.

Each participant produced 4 repetitions of 20 test sentences at a natural speech rate with 20 fillers, which means that each participant read a total of 160 sentences.

2.5.3 Procedure

The procedure was identical to that of Experiment 1.

2.5.4 Measurement

The recordings were manually annotated by the first author but with a somewhat different scheme. For this experiment, both the first and second syllables were marked. The first syllable was marked to confirm that derived Tone 3 can in fact trigger Tone 3 sandhi on this syllable. An example is shown in Figure 6. The annotation file had five tiers in total. The criteria for marking vowels and syllables remained the same. The first tier marked the vowel of the syllable. All other tiers marked the whole second syllable to index phonological information and recording quality. The second tier indicated the position of the syllable inside the sentences where a first syllable was marked '1' and a second syllable was marked '2', the third tier contained the Pinyin of the whole sentence followed by the underlying tone of the syllable. The fourth tier marked whether the syllable underwent tone sandhi. And the last tier indicated the quality of the recording. Similar to the previous experiment, I only used productions of recordings that were marked 'good'. The f0 extraction, normalization and visualization processes are identical to those in the previous experiment.



Figure 6: Annotation scheme of Experiment 2 (Tone 4)

2.5.5 Results and statistical modelling

The number of tokens for each possible combination of Underlying Representation and Surface Representation is summarized in Table *5*. The application rate of Tone 3 sandhi before underlying Tone 3 is 94.8% while the application rate before derived Tone 3 is 24.2%.²⁴ 79 tokens were not marked as 'good' and excluded, which accounts for 5.0% of all test stimuli.

²⁴ Again, I point out that the difference in application rates does not necessarily inform me of any differences in phonological representations. The difference can be accounted for by the difficulty in planning longer prosodic units. I also recognize that the application rate of derived Tone 3 from Tone 4 here is different from derived Tone 3 from Tone 1 in Experiment 1 (74.0%), and as one anonymous reviewer of the paper (Du & Durvasula, accepted) pointed out, such a difference may be used to argue for a phonological difference between the Tone 3s from different underlying sources. However, such a difference may simply be due to different groups of speakers in two independent experiments or even more simply that the observed difference in effect sizes is simply random variation, as would be expected between any two experiments measuring the same phenomenon. Future study is needed to compare Huai'an Tone 1 sandhi and Tone 4 sandhi on the same group of speakers in one single experiment. If the application rates are indeed replicable, an intriguing possibility that I note for the readers is that the phonetic difference between derived Tone 3 from Tone 4 may itself serve as a performance factor (related to the differential

UR	SR	Number of tokens
T2T3T4	T2T3T4	386
T3T3T4	T3T3T4	20
T3T3T4	T2T3T4	368
T2T4T4	T2T4T4	156
T2T4T4	T2T3T4	212
T3T4T4	T3T4T4	98
T3T4T4	T3T3T4	213
T3T4T4	T2T3T4	68

Table 5: Number of Tokens for UR and SR combination

The z-score transformed f0 contours on the crucial second syllable are shown in Figure 7. Again, the crucial comparison is between a derived Tone 3 and an underlying Tone 3 after a derived Tone 2 in the same surface context. We also present the tone contour for an underlying Tone 4 in the same surface context for visual comparison with the two crucial Tone 3s.

difficulty in implementing different tones that in turn effects planning) that can account for the difference of application rate outside phonology. At the moment, the explanations based on planning difficulty are simply speculative since it is not clear how phonological planning can affect tone sandhi application rate, and future study is needed to quantify the size of variation caused by planning difficulty.



Figure 7: Contours comparison of the second syllable in Experiment 2 (Tone 4)

(Error bars indicate standard error)

Based on visual inspection of the data, the pattern seems to be different from the case of Tone 1 sandhi. The derived Tone 3 seems to start as an underlying Tone 4,²⁵ instead of as an underlying Tone 3 as in Experiment 1. Furthermore, the derived Tone 3 gradually deviates from underlying Tone 4 through the whole contour; note, this is in contrast to Experiment 1, where the derived Tone 3 ended up at a value almost identical to the underlying Tone 1. However, the contour shape of the derived Tone 3 is again close to that of an underlying Tone 3 as in Experiment 1. Despite

²⁵ As observed by one anonymous reviewer of the paper (Du & Durvasula, accepted), more accurately, derived Tone 3 from Tone 4 actually starts higher than underlying Tone 4 as shown in Figure 7, although the difference in raw pitch between derived Tone 3 and underlying Tone 4 is very small (less than 5 Hz for the first 4 steps). As suggested by this anonymous reviewer, this pattern may be due to some effort to maintain contrast between derived Tone 3 and underlying Tone 4 is relatively high and close to underlying Tone 4.

the difference, incomplete phonetic neutralization is again clearly observed in the comparison between underlying Tone 3 and derived Tone $3.^{26}$ ²⁷

The modelling method remains the same as in Experiment 1, and four models are generated. The observation of incomplete phonetic neutralization is again supported by model comparisons. The addition of a tone sandhi condition improves the model on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=1429.23$, p<0.01), the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=66.22$, p<0.01) and the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)=32.67$, p<0.01). Figure 8 shows how the full model (Model 4) with the assumption of tone sandhi affecting every fixed effect fits the observed data. And the parameter estimates for full model are summarized in Table 6.

²⁶ I recognize that the underlying Tone 3 in Experiment 2 appears to be slightly phonetically different from its counterpart in Experiment 1, especially at the tonal offset position. Tone 3, which is usually used to represent low tone in Mandarin languages, generally involves creakiness. The small difference between underlying Tone 3s in Experiments 1 and 2 may be caused by inconsistency related to the Praat f0 estimation algorithms for creaky sounds. Whether it is due to this issue or due to simple random variation is something that I leave for further inquiry.

²⁷ Again, to further address the concern that incomplete neutralization patterns identified in Experiment 2 may arise as a result of averaging the outcomes of an optional phonological process, the distributions of underlying Tone 4, derived Tone 3, and derived Tone 3 are shown for each time step in APPENDIX F. Again, crucially, the derived Tone 3 distribution is generally uni-modal, and distinct from the other two distributions, across the time-steps. Thus, there is no evidence of an averaging artifact over optional surface representations for the derived Tone 3 case.



Figure 8: Observed data and Growth Curve Model fits for derived and underlying Tone 3

	Estimate	Std. Error	t	р
Intercept	0.50	0.06	8.26	< 0.01
Linear	-22.09	2.23	-9.89	< 0.01
Quadratic	4.38	1.33	3.29	< 0.01
Tone Sandhi: Intercept	-1.01	0.02	-43.58	< 0.01
Tone Sandhi: Linear	-10.43	1.26	-8.30	< 0.01
Tone Sandhi: Quadratic	7.17	1.25	5.75	< 0.01

 Table 6: Parameter estimates of the full model (Model 4) with the assumption of tone sandhi

 affecting every fixed effect (baseline: derived Tone 3)

Again, the effect size of incomplete neutralization is also large in Tone 4 sandhi. The mean difference in f0 between underlying Tone 3 and derived Tone 3 across all steps is 17 Hz, which is more than 2 times the Just Noticeable Difference of f0 value (7 Hz) for Mandarin speakers

(Jongman et al., 2017). Also, across the last 11 steps (step 9 to step 20), the f0 difference is over 21 Hz, which is more than 3 times the Just Noticeable Difference. The f0 difference (f0 of derived Tone 3 - f0 of underlying Tone 3) of each step is summarized in Table 7. Therefore, the case of Huai'an Tone 4 sandhi should also be safely defined as phonetically incomplete neutralization, and not susceptible to the criticism of a small effect size.

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	-4	11	24
1	-1	12	23
2	1	13	24
3	3	14	23
4	5	15	23
5	11	16	23
6	14	17	24
7	16	18	26
8	19	19	27
9	22	20	27
10	22		

 Table 7: f0 Difference of each step in Experiment 2 (Tone 4)

The coding in Experiment 2 also allowed me to answer another question that we did not answer for Experiment 1. In Experiment 1, we impressionistically coded whether or not the first syllable was in fact subject to Tone 3 sandhi. One could have argued that this impressionistic coding could have been inaccurate, and was based on a perceptual bias of the annotator (the author). To address this concern, it would have been optimal if I could have shown through phonological behavior that the derived Tone 2 is indeed phonologically identical to underlying Tone 2. Although historically Tone 2 sandhi (Tone 2 + Tone 2 \rightarrow Tone 3 + Tone 2) existed in Huai'an (Y. Wang & Kang, 2012), this tone sandhi rule was not observed in my fieldwork in early 2020 probably due to the influence of the standard language, as is generally observed in other languages (Labov, 1963; Milroy, 2001; inter alia). And no researchers before have tested if derived Tone 2 can trigger another tone sandhi process. Therefore, I cannot verify if the derived Tone 2 can trigger Tone 2 sandhi like an underlying Tone 2. Furthermore, I am not aware of any other phonological processes in the language that are triggered by Tone 2. As a result, it is not possible to establish Tone 2 category by phonological behavior in Huai'an and I turn to provide phonetic evidence for the Tone 2 identity of the derived rising tone.

To make some inroads into the question of the phonological nature of the (putatively) derived Tone 2 in initial position, in Experiment 2, I also annotated the first syllable, and are therefore able to observe the f0 contours for derived Tone 2 (from underlying Tone 3) and compare it to an underlying Tone 2 to see if the impressionistic coding was appropriate. The tone contours of the z-score transformed f0 for the relevant first syllables are shown in Figure 9. For the benefit of the reader, I also present the tone contour for an underlying Tone 3 in the first syllable that comes from a derived Tone 3 failing to trigger Tone 3 sandhi on the preceding syllable. By doing so, a three-way visual comparison is possible at the position of the first syllable under the same phonological environment, i.e. before derived Tone 3.



Figure 9: Contours comparison of the first syllable in Experiment 2 (Tone 4) (Error bars indicate standard error)

Based on the visual inspection of the data, the derived Tone 2 that undergoes Tone 3 sandhi with reference to the following derived Tone 3 is phonetically highly similar to an underlying Tone 2 with regard to the f0 contour. Both derived Tone 2 and underlying Tone 2 f0 contours are phonetically very different from underlying Tone 3. Furthermore, as with the other tone sandhi processes discussed in this paper, there is incomplete phonetic neutralization of the derived Tone 2 (from an underlying Tone 3) and the underlying Tone 2 in the first syllable. With the modelling method introduced in Section 2.4.5, the addition of a Tone Sandhi condition improves the model on the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)=4.96$, p=0.03), but not on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=2.16$, p=0.14) or the linear term as shown by comparing Model 3 ($x^2(1)=1.10$, p=0.29). Figure 10 shows how the full model (Model 4) with the assumption of tone sandhi affecting every fixed effect fits the observed data. And the parameter estimates for the full model are summarized in Table 8.



Figure 10: Observed data and Growth Curve Model fits for derived and underlying Tone 2

	Estimate	Std. Error	t	р
Intercept	-0.26	0.06	-4.07	< 0.01
Linear	10.00	2.69	3.72	< 0.01
Quadratic	9.63	1.89	5.11	< 0.01
Tone Sandhi: Intercept	-0.04	0.03	-1.51	0.13
Tone Sandhi: Linear	-1.32	1.30	-1.02	0.31
Tone Sandhi: Quadratic	-2.88	1.29	-2.23	0.02

before derived Tone 3 (Error bars indicate standard error)

 Table 8: Parameter estimates of the full model (Model 4) with the assumption of tone sandhi

 affecting every fixed effect (baseline: derived Tone 2)

However, consistent with my larger claim, this should not be interpreted as incomplete phonological neutralization. The mean difference in f0 between underlying Tone 2 and derived Tone 2 across all steps is only 1 Hz, which is *much* lower than the Just Noticeable Difference of

f0 value (7 Hz) for Mandarin speakers (Jongman et al., 2017). The f0 difference (f0 of underlying Tone 2 - f0 of derived Tone 2) of each step is summarized in Table 9. This indicates that native speakers of Huai'an may not be able to distinguish underlying vs derived Tone 2s and therefore are likely to analyze them to be in the same category in phonology. It is worth noting that an assumption has been made here that a phonetic difference that is *much* smaller than or around Just Noticeable Difference means phonologically complete neutralization, and a phonetic difference that is *much* bigger than the Just Noticeable Difference is compatible with both phonologically complete neutralization (as in Huai'an Tone 1 and Tone 4 sandhis) and phonologically incomplete neutralization. I acknowledge that some previous studies on incomplete neutralization have shown that phonetic differences that are smaller than the relevant Just Noticeable Difference are still perceptually distinguishable (Port & O'Dell, 1985; Warner et al., 2004; inter alia). However, the substantial phonetic difference between derived Tone 2 and underlying Tone 3 and the phonetic similarity between derived Tone 2 and underlying Tone 2 are difficult to account for by any mechanism known to me other than Tone 3 sandhi – it cannot simply be random variation or a coarticulatory change. Therefore, the impressionistic coding was in my opinion appropriate.

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	-8	11	2
1	-8	12	2
2	0	13	1
3	2	14	1
4	2	15	1
5	3	16	1
6	3	17	1
7	3	18	2
8	3	19	3
9	2	20	3
10	2		

Table 9: f0 Difference of each step for first syllable in Experiment 2 (Tone 2)

To summarize the results of Experiment 2, we showed using the feeding interaction between Tone 4 sandhi and Tone 3 sandhi, that the Tone 4 sandhi results in a phonological completely *derived* Tone 3. Despite this phonologically complete neutralization, we observed (a rather large) incomplete neutralization between the derived Tone 3 and underlying Tone 3 in the same surface tonal context. The experiment therefore replicates the results of Experiment 1.

2.6 Phonological representation of Mandarin tone

It is worth noting that the interpretation of Huai'an tone sandhi cases as incomplete neutralization relies on the general consensus that a Mandarin tone is a single phonological unit despite appearing as a phonetic tonal contour (Bao, 1990, 1992; Yip, 1989; inter alia). To form a Mandarin tone, tonal targets like High (h) and Low (l) should be grouped together as one unit by an intermediate node to be linked to Tone Bearing Unit (TBU) as in (15). The TBU is the syllable in Mandarin languages like Huai'an, and the intermediate node is called Tonal Root node by Yip (1989:150) and Contour node by Bao (1990:2). Therefore, a Mandarin tone is inseparable and should be examined as a whole. And a phonetic difference on any part of the contour between a *derived* Mandarin tone and its *underlying* counterpart indicates phonetically incomplete neutralization. In the case of Huai'an, there is a clear phonetic difference at the later part of tonal contour in Experiment 1 and 2.

(15) Phonological Representation of Mandarin Tone



Note: TBU means tone bearing unit, in Mandarin languages like Huai'an, a TBU is a syllable or a rhyme.

Based on this view, phonologically, it is not possible for part of a Mandarin contour tone to neutralize while part of the tone remains unchanged. Perhaps, the most convincing evidence for this single-phonological-unit representation in Mandarin languages comes from Contour Tone spreading. And the most discussed case is undoubtedly Danyang (Chan, 1991; M. Y. Chen, 1991; Yip, 1989; data from Lü, 1980). The pattern of interest is given in (16):

- (16) a. 2-syllable: hl. lh
 - b. 3-syllable: hl. hl. lh
 - c. 4-syllable: hl. hl. hl. lh

Note: 'h' means high tone, 'l' means low tone, and '.' indicates syllable boundary.

According to the analysis by Yip (1989), in these cases, a falling tone is associated with the first syllable and a rising tone is associated with the last syllable. Then the falling tone spreads rightwards over the domain as one single unit. If the falling tone is not a unit in phonology, one would expect only the low tone but not the whole contour to spread. A similar phenomenon of tone spreading is also found in Changzhi (Hou, 1983). It is worth noting that Duanmu (1994) challenges the above evidence by pointing out that Contour Tone spreading examples are only found in two languages and restricted to certain morpho-syntactic structures. However, since Changzhi City and Danyang City are geographically far away from each other (roughly 456 miles or 734 kilometers apart), tone spreading may be discovered in more languages and potentially more morpho-syntactic structures. To summarize, despite dispute, the tone spreading pattern itself offers strong support for phonological contour tone. It is also worth pointing out that despite disagreement with the single-unit analysis of contour tones, Duanmu (1994) claims that tone

sandhis results in a categorical change, which is argued in this paper to support the interpretation of incomplete neutralization in Huai'an.

Another piece of evidence supporting single-phonological-unit representation in Mandarin languages and specifically in Huai'an comes from Huai'an itself. By analyzing tone sandhi processes in Huai'an as being applied at the Tonal Root/ Tonal Contour tier, an elegant explanation can be provided for the tone sandhi patterns in Huai'an (Du & Lin, 2019). Therefore the Tonal Root/Tonal Contour node that can group tonal targets as one unit is probably real.

Crucially, if Tone 3 (low tone) is analyzed as low-register tone while other tones are analyzed as high-register tones as in Standard Mandarin (Bao, 1990; Duanmu, 2007; Yip, 1989), then tone sandhi is only applicable within the same register category. In Huai'an, Tone 3 only undergoes tone sandhi before Tone 3 while Tone1/Tone 2/Tone 4 only undergoes tone sandhi before Tone 1/Tone 2/Tone 4. Therefore Huai'an tone sandhi has a clear pattern of the application of Obligatory Contour Principle (Leben, 1973; McCarthy, 1986) on the Tonal Root/Contour node tier. There are phonetic basis for treating Tone 3 differently from other tones in phonology. Only Tone 3 in Standard Mandarin and Huai'an involves creaky voice (Duanmu, 2007). However, as argued in this dissertation, the relationship between phonetics and phonology is rather remote, so observation in speech production itself may not indicate phonological representation. I argue here it is unlikely to be just an accident that such tonal representation can provide an elegant explanation of tonal patterns in Huai'an. Tone sandhi patterns that are not predicted by such tonal representation are never found in Huai'an, now or before. All tone sandhi processes in Huai'an are listed in (17), including historically existed ones.

- (17) All tone sandhi rules in Huai'an Mandarin (Y. Wang & Kang, 2012)
 - a. Low-register tone sandhi²⁸
 - i. Tone 3 sandhi: T3 + T3 \rightarrow T2 + T3
 - b. High-register tone sandhi [optional processes]

i. Tone 1 sandhi:	$T1 + T1 \rightarrow T3 + T1$
ii. Tone 1 sandhi before Tone 3:	$T1 + T2 \rightarrow T3 + T2$
iii. Tone 1 sandhi before Tone 4:	$T1 + T4 \rightarrow T3 + T4$
iv. Tone 2 sandhi before Tone 1:	T2 + T1 \rightarrow T3 + T1 (historically existed)
v. Tone 2 sandhi:	T2 + T2 \rightarrow T3 + T2 (historically existed)
vi. Tone 2 sandhi before Tone 4:	T2 + T4 \rightarrow T3 + T4 (historically existed)
vii. Tone 4 sandhi before Tone 2:	$T4 + T2 \rightarrow T3 + T4$
viii.Tone 4 sandhi:	$T4 + T4 \rightarrow T3 + T4$

I assume the tonal representation in (18). And for this paper, I use uppercase vs. lowercase to distinguish features on the Tonal Root/Contour tier and the Tonal Target tier. The whole pitch rage is then divided as follows:

²⁸ Register is a tonal feature first proposed by Moira Yip (1980) and then widely adopted in the literature of Chinese tonal phonology. Here I simply use the feature to distinguish Tone 3 sandhi from other tone sandhi processes in Huai'an.

(18) Phonological Representation of Mandarin Tone

	high tone (h)	
High Register (H)	low tone (l)	
Low Pagistar (L)	high tone (h)	
Low Register (L)		

low tone (l)

With all these, the four Huai'an tones can be represented as in (19). Again, the tone sandhi has a clear pattern of the application of Obligatory Contour Principle (Leben, 1973; McCarthy, 1986) on the Tonal Root/Contour node tier. One example of high-register tone sandhi and one example of low-register tone sandhi are provided in (20). Therefore, tone sandhi will not be triggered on two adjacent tones that belong to different registers.

(19) Phonological Representation of Huai'an Tone



(20) Phonological Representation of Mandarin Tone



With the above single-phonological-unit viewpoint of tonal representations as backdrop, in Huai'an, the fact that both *derived* Tone 3 and *underlying* Tone 3 can trigger Tone 3 sandhi suggests that a derived Tone 3 is phonologically identical to underlying tone 3. In fact, to the best of my knowledge, I am not aware of any Mandarin languages where only underlying Tone 3 triggers Tone 3 sandhi, and not derived Tone 3 - this would be a correlation that is accounted for by analyzing it as phonological neutralization. However, a phonetic difference on any part of the contour between a derived contour tone and its underlying counterpart indicates phonetic incomplete neutralization of the whole contour tone unit. In the case of Tone 1 and Tone 4 sandhis in Huai'an, there is a clear phonetic difference at the tonal offset position as shown in Experiment 1 and 2.

Based on the above, I would like to explicitly acknowledge that my claims in the dissertation about incomplete phonetic neutralization in the face of complete phonological neutralization are contingent on the phonological representations we have assumed. As I see it, it cannot be any other way. Note, the argument for incomplete neutralization in any language depends on a certain set of assumed phonological representations. For example, in German, the interpretation of incomplete
neutralization depends on the devoicing rule actually resulting in a [-voice] feature (or equivalent). If the devoicing process results in some other phonological representation with similar phonetics, then the whole issue of incomplete neutralization vanishes, and there is no need to entertain any more gradience in the phonological system to explain the observed phonetic patterns. In fact, a version of such a featural account is implied by Hale et al. (2007), who argue that language-specific phonetics can in fact be accounted for by different phonological feature combinations. Similarly, in Huai'an, it is possible to explain what is observed in the phonetics by changing or adding new phonological representations, but then of course independent evidence of the same representations in the language or in other related languages generally needs to be provided, otherwise it becomes an ad hoc, and therefore unjustified, claim. More generally, any set of representations or computations cannot simply be post-hoc accounts of the data/patterns but need to be independently justified claims.

CHAPTER 3 PHONETICALLY IDENTICAL FORM CAN HAVE DIFFERENT PHONOLOGICAL BEHAVIORS

3.1 More background

For this chapter, I will present experimental evidence again from Huai'an tone sandhi to argue that phonetically identical forms can have different phonological behaviors. Therefore, phonologically different surface forms do not necessarily correspond with different phonetic distributions, which undermines Statement 2 in (1). If this argument can be established, it would also mean that some distinctions in phonological representations do not necessarily have consequences in the relevant phonetics. I will elaborate on the analysis assuming the distinction in phonological representations in Chapter 4 under Section 4.2. I will also state another possibility in Section 4.2 where prosodic structures, instead of the phonological representation difference, can explain the effect observed in Chapter 3.

As mentioned in Chapter 1, it is logically impossible to assert phonetic identity between two phonological elements due to the possibility of an infinite number of phonetic cues. However, in this dissertation, I assume that if two phonological entities are identical in all functionally important phonetic cues, phonetic identity can be argued for. If a native speaker make use a single phonetic cue to distinguish among relevant similar phonemes, this phonetic cue should be defined an important cue.

In Mandarin languages, f0, duration and intensity should be recognized as the important phonetic cues for lexical tone. There is evidence that native speakers of Mandarin languages rely solely on f0 contour to distinguish tones (Howie, 1976; Tupper et al., 2020), f0 contour identity can arguably indicate phonetic identity. Besides the F0 contour, there is experimental evidence

that native speakers can make use of only intensity or only duration to distinguish tones (Fu & Zeng, 2000; Whalen & Xu, 1992 for intensity; Blicher et al., 1990 for duration). If two tones are identical in these three dimensions, it is perhaps reasonable to argue for phonetic identity. In the Huai'an case, the crucial comparison is between derived Tone 3 from Tone 4 at the lexical level and derived Tone 3 from Tone 4 at the post-lexical level. I will show that the two derived Tone 3s are indeed indistinguishable with regard to f0, duration and intensity in Experiment 3.

Despite being arguably phonetically identical, I will show that the two derived Tone 3s have different phonological behaviors, which is indicated by different rates of triggering another Tone 3 sandhi process. The comparison is made between derived Tone 3 from Tone 4 at the lexical level and derived Tone 3 from Tone 4 at the post-lexical level. It is worth noting that I argued in Chapter 2 that triggering rates of tone sandhi processes need not be used to infer phonology. However, in this specific case where the comparison is made between the same phonological processes and under the same experimental paradigm. It may be reasonable to analyze that the triggering rate of Tone 3 sandhi can function as indicator of phonological behaviors, therefore different triggering rates of Tone 3 sandhi indicates phonological difference between two described derived Tone 3s. I will introduce the experimental paradigm in detail in Section 3.2.

The two tone sandhi rules that are related to this chapter are shown in (21). Importantly, the high-register Tone 4 sandhi optionally applies at both the lexical level and the post-lexical level. To be more specific, Tone 4 sandhi applies at the right edge of noncompositional words and the right edge of an utterance as shown in (22a) and (22b) (Du & Lin, 2019). All syllables used in (22) except (22d) are underlyingly Tone 4. In (22a), [o-ta-li-ia] is an noncompositional word, Tone 4 sandhi applies at the penultimate syllable with reference to the last syllable of this noncompositional word. This syllable that undergoes Tone 4 sandhi does not change when [o-ta-

li-ia] forms a compound word with another monosyllabic word to the right. In (22b), each syllable forms a separate word by itself, so the Tone 4 sandhi clearly applies at the post-lexical level. Tone 4 sandhi applies at the penultimate syllable with reference to the last syllable of the utterance. Moreover, when the noncompositional word in (22a) is the penultimate word of an utterance and the last word is monosyllabic as in (22c), only lexical Tone 4 sandhi applies. Post-lexical Tone 4 sandhi ceases to apply probably to avoid adjacent Tone 3 syllables. Here the noncompositional word is marked by square brackets and is not located at the right edge of the utterance. Lexical Tone 4 sandhi applies at the right edge of the noncompositional word to make the fifth syllable a described Tone 3. Although the last two syllables of the sentences are both Tone 4, post-lexical Tone 4 sandhi fails to apply to avoid the marked form of Tone 3 sequence at any positions of an utterance. Under Optimality Theory (Prince & Smolensky, 1993), this can again be explained by the employed '*Tone3Tone3' constraint. which is based on the Obligatory Contour Principle. This constraint prevents adjacent Tone 3 syllables (C.-Y. Wang & Lin, 2011; Zhang, 1997; see also M. Y. Chen, 2000 for using this constraint in an implicit fashion). When the non-compositional word with all syllable being Tone 4 in (22c) is replaced with another non-compositional word with the last two syllables being Tone 2 and Tone 4 as in (22d), Tone 4 sandhi applies at the end of the utterance. This further proves that post-lexical Tone 4 sandhi fails to apply in (22c) to avoid Tone 3 sequence.

Overall, it is clear that there are two mechanisms of applying Tone 4 sandhi, namely at the lexical and the post-lexical levels. A single mechanism of Tone 4 sandhi that is only applicable at the right edge of an utterance cannot explain the patterns in (22a) and (22c).

- (21) Relevant tone sandhi rules in Huai'an Mandarin (Y. Wang & Kang, 2012)
 - a. Low-register tone sandhi

Tone 3 sandhi: T3 + T3 \rightarrow T2 + T3

b. High-register tone sandhi [optional processes]

Tone 4 sandhi: T4 + T4 \rightarrow T3 + T4

- (22) The application of Tone 4 sandhi at different levels
 - a. The application of Tone 4 sandhi at the lexical level

	o-ta-li-ia 'Australia'	ə-ta-li-ia lu 'Australia Avenue'
Morphological structure	[ɔ-ta-li-ia]	[[ɔ-ta-li-ia] [lu]]
UR	T4 T4 T4 T4	T4 T4 T4 T4 T4
SR	T4 T4 <u>T3</u> T4	T4 T4 <u>T3</u> T4 T4

b. The application of Tone 4 sandhi at the post-lexical level

lu	iə	tso	zəm		
Mr. Lu	want	cook	meat		
'Mr. Lu wants to cook meat'					

- UR T4 T4 T4 T4
- SR T4 T4 <u>T3</u> T4

c. The boundary effect of noncompositional word

	lu Mr. Lu	t¢ ^h y go	[ɔ-ta-li-ia] Australia	ko live
	'Mr. Lu goe	s to Aus	tralia to live.'	
UR	T4 T4 [T4]	[4 T 4 T 4	-] T4	
SR	T4 T4 [T4]	Γ4 <u>T3</u> T4	4] T4	

d. Application of Tone 4 sandhi across the boundary of noncompositional word

	lu Mr. Lu	t¢ ^h y go	[t¢ia-li-fɔʔ-ni-ia] California	ko live
	'Mr. Lu goes	s to Cali	fornia to live.'	
UR	T4 T4 [T1 T3 T1 T2 T4] T4			
SR	T4 T4 [T1 T	3 T1 T2	2 T3] T4	

Also, as partially stated in Chapter 2, Tone 4 sandhi is only applicable when the two syllables involved belong to the same phonological phrase at the post-lexical level or prosodic word at the lexical level. At the post-lexical level, as stated in Chapter 2, the low-register Tone 3 sandhi is mandatory only when the syllable that undergoes tone sandhi and the following syllable that triggers tone sandhi are in the same phonological phrase. In Huai'an, like many other Mandarin languages (C.-Y. Wang & Lin, 2011), the boundary between subject and predicate may block the formation of a phonological phrase. Therefore Tone 3 sandhi is optional across the boundary between subject and predicate in Huai'an.

Crucially, the described Tone 3 outputs of the Tone 4 sandhi process at both levels feed the low-register Tone 3 sandhi process as in (23). Since Tone 4 sandhis at both levels are optional and

Tone 3 sandhi is also optional given different possible prosodic structures for utterances in (23), multiple surface representations are possible for both examples. In (23a), the last two syllables form a single word and they combine with the monosyllabic subject to become a sentence. Therefore the Tone 3 is derived at the lexical level. While in (23b), each syllable forms a separate word by itself, so the Tone 3 is clearly derived at the post-lexical level. The wordhood of the last two syllables in (23a) is confirmed by the test of Conjunction Reduction²⁹, which is a phrasal-level rule according to C.-T. Huang (1984). Conjunction Reduction is argued by C.-T. Huang to apply to coordinated phrases but not to coordinated words in both English and Chinese languages. To give an example in English, it is grammatical to say 'used and new books' to mean used books and new books; but it is ungrammatical to say 'New York and Orleans' to mean the city New York and the city of New Orleans. To apply this test in Huai'an, for the two examples in (23), it is possible to apply Conjunction Reduction to the last two syllables in (23b), as shown in (24b). But it is not possible to apply this rule to the last syllables in (23a), as shown in (24a). Although both [pa-ts \tilde{a}] and [pa-lin] are words, it is not possible to extract their common part and connect the left. This suggests [to $z_{\overline{2}}u_{\overline{2}}$] in (23b) is at the phrasal level while [pa-ts \tilde{a}] in (23a) is a single word and not at the phrasal level.

²⁹ This is a term created by C.-T. James Huang (1984), and I follow his terminology here. I do not use 'reduction' as a computation term here.

- (23) Feeding Order in Huai'an Mandarin (boldface represents the locus of a potential tonal change due to the relevant tone sandhi process; the data is from the author)
 - a. Tone 4 sandhi feeds Tone 3 sandhi at the lexical level

u pa Mr. Wu fo	a-tsæ orcibly occupy		
'Mr. Wu forcibly	occupy (something	y).'	
UR	<u>T3 T4 T4</u>		
Tone 4 sandhi	T3 T3 T4		(or) T3 T4 T4
Tone 3 sandhi	T2 T3 T4	(or) T3 T3 T4	T3 T4 T4
SR	T2 T3 T4	(or) T3 T3 T4	(or) T3 T4 T4
Corresponding			
Prosodic Structu	re (T2 T3 T4)∳	$\phi(T3)_{\phi} \phi(T3 T4)_{\phi}$	_φ (T3 T4 T4) _φ
	_♦ (T2) _♦ _♦ (T3 T4) _♦		_♦ (T3) _♦ _♦ (T4 T4) _♦

b. Tone 4 sandhi feeds Tone 3 sandhi at the post-lexical level

u Mr. Wu	to chop	zəu meat		
'Mr. Wu chops	s meat.	.'		
UR		<u>T3 T4 T4</u>		
Tone 4 sandhi		T3 T3 T4		(or) T3 T4 T4
Tone 3 sandhi		T2 T3 T4	(or) T3 T3 T4	T3 T4 T4
SR		T2 T3 T4	(or) T3 T3 T4	(or) T3 T4 T4
Corresponding				
Prosodic Struc	cture	¢(T2 T3 T4)∳	_φ (T3) _φ _φ (T3 T4) _φ	_φ (T3 T4 T4) _φ
	φ	(T2) _{∳ ∲} (T3 T4) _∮		_φ (T3) _φ _φ (T4 T4) _φ

(24)a. *o ра tsæ hu līn forcibly occupy and 1sg bully Intended: "I forcibly occupy and bully." tshã b. o zəm hu to 1sg chop meat and vegetables

'I chop meat and vegetables.'

In the next section, I will present that, for trisyllabic utterances as in (23), the two described derived Tone 3s at different levels trigger Tone 3 sandhi at different rates. Again, different my argument in Chapter 2, I here analyze the triggering rate of Tone 3 sandhi in this specific case functions as indicator of phonological behaviors. In this case, Tone 3 sandhi applies across the boundary between subject and predicate at both levels. Also, one performance factor identified in Huai'an is also controlled. The planning difficulty effect caused by long utterance, as a performance factor and discussed in Chapter 2, should not have an influence here. For Tone 3 sandhi to apply before derived Tone 3 at both levels, a trisyllabic phonological phrase is needed where both high-register Tone 4 sandhi and low-register Tone 3 sandhi occur. Overall, it is difficult to assign the difference in triggering rate to prosodic boundary strength or utterance length, which are usually used to explain variation in phonological processes in previous literature (Kilbourn-Ceron & Goldrick, 2021; Tanner et al., 2017; Wagner, 2012). Therefore application rate differences, which I will show in the next section, will provide good evidence for my argument that described derived Tone 3s at both levels are phonologically different. In the next section, I will also present experimental data for the described derived Tone 3s at both levels being phonetically identical with regard to f0 contour, duration and intensity contour.

3.2 Experiment 3

3.2.1 Participants

I recruited 5 native speakers of Huai'an Mandarin also via personal relationships in Huai'an City. The age range was from 53 to 58 years old. Again, to minimize the influence of Standard Mandarin, I avoided younger speakers in this study. Among them, 2 self-identified as female, and 3 as male. 3 speakers had participated in the pilot experiment for Experiment 1 and the other 2 speakers had participated in Experiment 2. The interval between pilot experiment and Experiment 3 was about 18 months, and the interval between Experiment 2 and Experiment 3 was about 10 months; all 5 participants were not told and failed to guess the purpose of Experiment 3. Like Experiments 1 and 2 (in Chapter 2), all the participants were born and raised in Huai'an City. Other speakers have not participated in any linguistic studies before or heard about the concept of phonology or phonetics.

3.2.2 Stimuli

The stimuli are first divided into two groups. Group 1 is composed of trisyllabic sentences with a monosyllabic subject and a disyllabic verb like that in (23a) and Group 2 is composed of trisyllabic sentences with each syllable forming a separate word like that in (23b). Each group is further divided into four sets, which were organized in the same way as in Experiments 1 and 2. The four sets of trisyllabic sentences are shown in (25), and the full stimulus list is summarized in APPENDIX C.

- (25) Four sets of stimuli in each group in Experiment 3 [the syllables crucial for the current comparison in each group are underlined and boldface]
 - a. underlying T3 following underlying T2: $/T2 T3 T4/ \rightarrow [T2 T3 T4]$
 - b. underlying T3 following underlying T3: /T3 T3 T4/ \rightarrow [T2 T3 T4]
 - c. derived T3 following underlying T2: $/T2 T4 T4 \rightarrow [T2 T3 T4] \text{ or } [T2 T4 T4]$
 - d. derived T3 following underlying T3: $/T3 T4 T4 \rightarrow [T2 T3 T4]$ or [T3 T4 T4]

The crucial comparison is between derived Tone 3s in both groups, namely the derived Tone 3s exist in (25c) and (25d) in both groups. Since triggering rate of Tone 3 sandhi is employed to indicate phonological behaviors, I will not exclude derived Tone 3 tokens that do not trigger Tone 3 sandhi as in Experiment 1 and 2.

Each participant produced 4 repetitions of 40 test sentences at natural speech rate with 8 fillers, which means the total number of utterances a participant has read is 192.

3.2.3 Procedure

The procedure was identical to that of Experiment 1 and 2.

3.2.4 Measurement

The recordings were also manually annotated by the author and with the same scheme as in Experiment 2. An example is shown in Figure *11*.



Figure 11: Annotation of Experiment 3 (Tone 4 sandhi at the lexical and the post-lexical levels)

3.2.5 Results and statistical modelling

The number of tokens for each possible combination of Underlying Representation and Surface Representation at both the lexical and the post-lexical levels is summarized in Table *10* and Table *11*. The application rate of Tone 3 sandhi before derived Tone 3 at the lexical level is 26.7% while the application rate before derived Tone 3 at the post-lexical level is 47.8%. At the lexical level, 7 tokens were not marked as 'good' and excluded, which accounts for 1.2% of all test stimuli. At the post-lexical level, 2 tokens were not marked as 'good' and excluded, which accounts for 0.5% of all test stimuli. It is obvious that the application rate of Tone 3 sandhi before derived Tone 3 at the lexical level is lower than that at the post-lexical level.

In this experiment, triggering rates of Tone 3 sandhi on the first syllable are assumed by me to be phonologically meaningful and I trusted my annotation completely. I treated all annotated derived Tone 3s at the position of the second syllable as Tone 3 when calculating triggering rates of Tone 3 sandhi and doing phonetic analysis. This means I did not treat derived Tone 3s that do not trigger Tone 3 sandhi as surface Tone 4. In Table *10*, for the annotation surface forms of 'Tone 3 Tone 3 Tone 4', the annotated derived Tone 3 on the second syllable is treated as Tone 3. Similar practice is also employed in Table *11*. I have shown in Experiment 2 that my annotation is reliable, so it is reasonable to continue to trust my annotation in Experiment 3.

UR	SR	Number of tokens
T2T3T4	T2T3T4	77
T3T3T4	T3T3T4	6
T3T3T4	T2T3T4	110
T2T4T4	T2T4T4	50
T2T4T4	T2T3T4	48
T3T4T4	T3T4T4	16
T3T4T4	T3T3T4	63
T3T4T4	T2T3T4	23

Table 10: Number of Tokens for UR and SR combination at the lexical level

UR	SR	Number of tokens
T2T3T4	T2T3T4	99
T3T3T4	T3T3T4	2
T3T3T4	T2T3T4	97
T2T4T4	T2T4T4	48
T2T4T4	T2T3T4	52
T3T4T4	T3T4T4	10
T3T4T4	T3T3T4	47
T3T4T4	T2T3T4	43

Table 11: Number of Tokens for UR and SR combination at the post-lexical level

The number of tokens for derived Tone 3 triggering and not triggering Tone 3 sandhi at both levels for each speaker is summarized in Table *12*. And the triggering rate for each speaker is summarized in Table *13*. For all participants except Speaker 1, derived Tone 3 at the post-lexical level triggers Tone 3 sandhi at a higher rate than derived Tone 3 at the lexical level. For speaker 1, the two derived Tone 3s trigger Tone 3 sandhi at the same rate. Therefore, it is safe to assert that derived Tone 3 at the post-lexical level indeed triggers Tone 3 sandhi at the higher rate than derived Tone 3 at the post-lexical level indeed triggers Tone 3 sandhi at the higher rate than derived Tone 3 at the lexical level, Again, it is difficult to assign the difference in triggering rate to prosodic boundary strength or utterance length, which are usually used to explain variation in phonological processes. Therefore such difference is more likely to be analyzed to be meaningful in the phonology.

Lexical L	evel	Post-Lexi	cal Level
Yes	No	Yes	No
3	15	3	15
3	15	11	9
5	10	7	9
2	17	7	12
10	6	15	2
	Lexical Lo Yes 3 3 5 2 10	Lexical Level Yes No 3 15 3 15 5 10 2 17 10 6	Lexical Level Post-Lexi Yes No Yes 3 15 3 3 15 11 5 10 7 2 17 7 10 6 15

 Table 12: Number of tokens for derived Tone 3 triggering and not triggering Tone 3 sandhi at the

 lexical and the post-lexical levels (by speaker)

Speaker	Lexically derived Tone 3	Post-Lexically derived Tone 3
	triggering Tone 3 sandhi	triggering Tone 3 sandhi
1	16.7%	16.7%
2	16.7%	55.0%
3	33.3%	43.8%
4	10.5%	36.8%
5	62.5%	88.2%

Table 13: Rate of derived Tone 3 triggering Tone 3 sandhi at the lexical and the post-lexical levels (by speaker)

I will then show that the two derived Tone 3s, which arguably have different phonological behaviors, are indistinguishable in all important phonetic cues. These cues include f0, duration and intensity.

With regard to f0, the z-score transformed f0 contours on the crucial second syllable are shown in Figure *12*. Again, the crucial comparison is between a derived Tone 3 at the lexical level and a derived Tone 3 at the post-lexical level. Based on visual inspection of the data, the derived Tone 3s at both levels are indistinguishable.



Figure 12: Contours comparison of the second syllable in Experiment 3 (Error bars indicate standard error)

The modelling method remains the same as in Experiment 1 and 2, and four models are generated. I first treated the two derived Tone 3s as the same and modelled them as one single contour to get Model 1. Then I built models that treat them as different, namely, models that include a level condition (derived Tone 3 at the lexical level vs. derived Tone 3 at the post-lexical level) to do model comparison. Based on Model 1, level condition is first allowed to affect only intercept to get Model 2. Then level condition is allowed to affect both intercept and linear term to get Model 3. Finally, level condition is allowed to affect all fixed effects, which include intercept,

linear term and quadratic term, and the outcome is Model 4. Chi-square likelihood ratio test was used to determine whether two minimally different models differ significantly.

The observation of identical phonetic distribution is supported by model comparisons. The addition of a level condition does not improve the model on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=0.15$, p=0.70), or the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=0.16$, p=0.69) or the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)<0.01$, p=0.98). Figure 13 shows how the base model (Model 1) with the assumption of level not affecting every fixed effect fits the observed data. And the parameter estimates for the base model, namely Model 1, are summarized in Table 14.



Figure 13: Observed data and Growth Curve Model fits for derived Tone 3s at the lexical and the post-lexical levels (Error bars indicate standard error)

	Estimate	Std. Error	t	р
Intercept	-0.01	0.02	-0.38	0.71
Linear	-15.65	3.74	-4.18	0.01
Quadratic	0.80	0.80	1.01	0.36

Table 14: Parameter estimates of the base model (Model 1) with the assumption of level not affecting every fixed effect³⁰

The phonetic identity is also supported by comparing raw f0. The mean difference in f0 between underlying Tone 3 and derived Tone 3 across all steps is only 5 Hz, which is less than the Just Noticeable Difference of f0 value (7 Hz) for Mandarin speakers (Jongman et al., 2017). Also, across all steps except the first one (step 1 to step 20), the f0 difference is less than 7 Hz. The f0 difference (f0 of derived Tone 3 at the post-lexical level - f0 of derived Tone 3 at the lexical level) of each step is summarized in Table *15*. Again, I acknowledge that some previous studies on incomplete neutralization have shown that phonetic differences that are smaller than the relevant Just Noticeable Difference are still perceptually distinguishable (Port & O'Dell, 1985; Warner et al., 2004; inter alia). Perceptual studies in the future are needed in Huai'an to settle this issue.

³⁰ It is worth noting here the quadratic term is not significant, therefore statistically, intercept and linear term are enough to model the z-score transformed f0 contours here. However, as shown in Experiment 2, derived Tone 3 from Tone 4 in Huai'an needs quadratic term to model. to be consistent, I also employed quadratic term here. Adding more terms should not change the results of model comparison.

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	8	11	4
1	4	12	4
2	4	13	3
3	5	14	6
4	5	15	5
5	4	16	5
6	4	17	3
7	4	18	6
8	5	19	5
9	6	20	4
10	5		

Table 15: f0 Difference of each step in Experiment 2 (Tone 4)

Since all critical syllables measured were always a voiceless unaspirated stop plus vowel sequence. It is not unreasonable to use vowel duration to represent tone duration. The distributions of z-score transformed duration are shown in Figure 14. Based on visual inspection of the data, the two derived Tone 3s seem to be different with regard to vowel duration. The vowel duration at the lexical level is slightly longer than that at the post-lexical level. However, Paired t-test shows that such difference is not statistically significant different: t(4)=0.59, p=0.59. Moreover, the distributions of raw durations of the two derived Tone 3s almost perfectly match as shown in Figure 15. And the difference of means of raw duration is only 0.2 milliseconds. It is hard to

imagine that native speakers can make use of such small difference to distinguish derived Tone 3s at the lexical and the post-lexical levels.



Figure 14: Comparison of z-score transformed duration of the two Tone 3s in Experiment 3



Figure 15: Comparison of raw duration of the two Tone 3s in Experiment 3³¹

³¹ It is worth noting that here durations at both lexical and post-lexical levels are bimodal. Part of the source comes from interspeaker variation. The raw duration by speaker is shown in APPENDIX G. For individual, the distributions of duration are generally unimodal.

With regard to intensity, the z-score transformed intensity on the crucial second syllable are shown in Figure *16*. Based on visual inspection of the data, the derived Tone 3s at both levels are indistinguishable with regard to intensity.



Figure 16: Intensity contours comparison of the second syllable in Experiment 3 (Error bars indicate standard error)

For statistical analysis, it turns out that Paired t-test suggests no statistically significant difference at all steps. The results are summarized in Table *16*. Recall in Chapter 2, I stated the problem of dividing continuous time into multiple time bins to do statistical analysis. Such method has to make multiple comparisons, which is problematic for increasing the risk of 'false positives'. Since each time bin incurs the nominal 5% false positive rate implied by 'p < 0.05', overall, the false positive rate with multiple time bins and multiple comparisons will be much higher than a single comparison. Therefore, with this method, derived Tone 3s are more likely to analyzed to be different. In the current case of analyzing intensity, even with such method where a positive result has a better chance to appear, no positive results are returned by the Paired t-test. Therefore, it is

highly unlikely that the two derived Tone 3s differ in intensity. Although intensity, like f0, is measured as time-varying data, I did not employ Growth Curve Analysis (GCA) (Mirman, 2017; Mirman et al., 2008) to analyze intensity due to the lack of theoretical basis. Recall for f0, it is widely assumed in previous literature that the most complex f0 contour for Mandarin tones can only have one change of direction and produce U-shaped contours (Bao, 1990, 1992; Duanmu, 1994; inter alia). Based on this, I only considered up to second-order function to ensure that the final model is not more complex than a U-shape contour. In contrast, the study on the intensity of Mandarin tones is rare. It is not clear what the most complex intensity contour should look like. Therefore, I am not sure up to what level of higher-order function should be considered, and I chose a different method to do statistical analysis.

The difference in raw intensity is summarized in Table *17* (intensity at the post-lexical level - intensity at the lexical level). The largest difference at step 20 is only 0.38 dB. Again, it is hard to imagine that native speakers can make use of such small difference to distinguish derived Tone 3s at both levels.

Step	df	t	р	Step	df	t	р
0	4	-0.56	0.61	11	4	-0.52	0.63
1	4	-0.13	0.91	12	4	-0.62	0.57
2	4	0.30	0.78	13	4	-0.50	0.65
3	4	0.77	0.49	14	4	-0.36	0.74
4	4	0.74	0.50	15	4	-0.20	0.85

Table 16: Paired t-test results for all steps

(Intensity, baseline: derived Tone 3 at the post-lexical level)

Table 16: (cont'd)

5	4	0.25	0.82	16	4	-0.31	0.78
6	4	-0.14	0.90	17	4	-0.33	0.76
7	4	-0.31	0.77	18	4	-0.04	0.97
8	4	-0.39	0.72	19	4	0.11	0.92
9	4	-0.38	0.72	20	4	0.24	0.82
10	4	-0.32	0.77				

Step	intensity difference (dB)	Step	intensity difference (dB)			
0	-0.07	11	0.03			
1	-0.12	12	0.02			
2	-0.08	13	>-0.01			
3	-0.12	14	-0.08			
4	-0.13	15	-0.12			
5	-0.04	16	-0.08			
6	0.03	17	-0.10			
7	0.10	18	-0.24			
8	0.13	19	-0.29			
9	0.09	20	-0.38			
10	0.03					
Table 17: Intensity difference of each step in Experiment 3						

 Table 17: Intensity difference of each step in Experiment 3

Overall derived Tone 3s at both levels are indistinguishable also with regard to duration and intensity. Therefore I can be more sure that these two Tone 3s are indeed identical phonetically.

3.3 Interim discussion

With regard to using application rate of Tone 3 sandhi to infer phonological behavior, concerns may still be raised with regard to performance factors. Considering the number of performance factors that have been identified to affect speech production, the chance is high that there are a lot more unidentified performance factors. Therefore the possibility can never be ruled out that the differences in application rates may be caused by one of the performance factors. While such an argument being logical, it is certainly not precise or accurate enough. And the responsibility is on researchers to identify the exact performance source that causes the differences in application rates. Due to the limitation of the current dissertation, I cannot identify other reasonable performance factors for the Huai'an case. I believe it is at least reasonable at the current stage to see if an analysis of the differences in application rate of Tone 3 sandhi in terms of differences in phonological representation is possible. I will return to this issue in Chapter 4 under Section 4.2.

In the future, to further confirm that phonetically identical forms can have different phonological behaviors, more evidence from different languages are needed with different methods to establish phonological inequality.

CHAPTER 4 THE EXPLANATION FOR THE OBSERVED GAP BETWEEN PHONOLOGY AND PHONETICS³²

4.1 The explanation for incomplete neutralization

4.1.1 Desiderata for any explanation for incomplete neutralization

With the two clear cases of incomplete neutralization shown in Experiments 1 and 2, the next step is naturally the explanation for incomplete neutralization, i.e. why phonologically identical forms can have different phonetic distributions. First of all, I would like to lay out the desiderata stated in Du and Durvasula (accpeted) that any explanations of incomplete neutralization must achieve before discussing previous explanations and introduce the current explanation.

- (26) Desiderata for a theory of incomplete neutralization
 - a. The simplest explanation of why incomplete neutralization exists as a phenomenon.
 - b. An explanation for the actual distribution of effect sizes among different phonological processes.
 - c. An explanation of why 'over-neutralization' is never observed.
 - d. An explanation of how a feeding interaction is possible where the derived representation still incompletely neutralizes with the element that triggers the process.
 - e. Related to (d), an explanation of why incompletely neutralized segments can trigger the process, but other phonetically similar segments do not.

³² Part of this chapter comes from my collaborative work with Karthik Durvasula, see reference (Du & Durvasula, accepted).

First, to ensure the priority of a relatively simple theoretical model, explanations that can solve the problem while retaining a relatively simple phonological model should be considered first (Occam's razor/law of parsimony). Consequently, if independently needed performance mechanisms have the potential to account for the observation of incomplete phonetic neutralization, they should be prioritized. Consistent with this principle, in the current study, the difference in Tone 3 sandhi application rates is assigned to independently needed performance factors of phonological planning, and therefore there is no need to complicate our understanding of the relevant phonological (tonal) representations. For the explanation of incomplete neutralization, beyond previously identified factors such as orthography and task effects, performance factors in my opinion that need to be explored further include phonological planning (Kilbourn-Ceron & Goldrick, 2021; Tanner et al., 2017; Wagner, 2012), cascaded activation of morphemes during production (Goldrick & Blumstein, 2006) and variability of phonological process. I will show in this dissertation that variability can trigger incomplete neutralization with large effect size.

The second challenge facing theories of incomplete neutralization is the systematic disparity in effect sizes (26b). Any proposed theory should explain among the observed cases why effect sizes of incomplete neutralization are rather small in devoicing processes (as in German, Dutch, Russian...), but can be quite large as in Huai'an tone sandhis or Japanese vowel lengthening. Related to (26a), it is optimal to assign such disparity to independently needed performance mechanisms.

The third challenge is that the proposed explanation should not only predict cases of 'incomplete neutralization' where the derived category is phonetically close to an underlying category (and in fact, between the phonetic manifestation of two underlying categories – its own UR and the phonological representation it is putatively changing to), but also avoid predicting

cases of 'over-neutralization' where the degree of application is beyond the phonetic distribution of the underlying category it is neutralizing to (26c). Back to the case of German devoicing, under the scenario of 'incomplete neutralization', the phonetic cues of *derived* voiceless stops fall between underlying voiceless stops and underlying voiced stops. While under the scenario of 'overneutralization', the phonetic cues of *underlying* voiceless stops fall between *derived* voiceless stops and underlying voiced stops. However, only 'incomplete neutralization' has been observed in examined languages including Huai'an. This observation would be particularly problematic for purely exemplar representations (Brown & McNeill, 1966; Bybee, 1994; Goldinger, 1996, 1997; Port & Leary, 2005, Roettger et al., 2014; inter alia). Many previous theories account for the absence of 'over-neutralization' by proposing some mechanism where phonetically incomplete neutralization is simply intermediate between two representations as it results from a blend of all phonetic cues of two distinct representations (Anderson, 1975; Braver, 2019; Gafos & Benus, 2006; Nelson & Heinz, 2022; Smolensky et al., 2014; Van Oostendorp, 2008).³³ Such theories are either not specific enough, or other independently needed mechanisms need to incorporate to capture the systematic disparity in effect sizes illustrated in the previous challenge (26b).

The fourth challenge that any theory of incomplete neutralization faces is to explain how a feeding interaction is possible where the derived representation still incompletely neutralizes with the element that triggers the process (26d). In the case of Huai'an, the Tone 3 output of the high-register tone sandhi processes can feed the low-register Tone 3 sandhi process as in (10) despite incompletely neutralizing with underlying Tone 3 in the phonetics. Any categorical theory of phonological representations naturally accounts for this as is observed rule process/rule interactions. Of course, it is possible for a theory of gradient phonological representations to do so

³³ Note, typically, incomplete neutralization is argued to be a blend of the surface representation and the underlying representation, or the surface representation and a base representation, or two co-activated surface representations.

too; however, to assess the effectiveness of such a theory, one needs to grapple with the specifics of the representations and computations proposed. Back to the issue of Tone 3 sandhi application rate difference, if one were to propose that the differential application rates are a consequence of gradient phonological representations, where phonetic proximity triggers application of a process, then one has to address two things: first, why do we see the gradience in application rates with the derived category but not with the underlying category, though both vary in terms of phonetic manifestations?; second, we need to ensure that other phonetically similar sounds do not trigger the process too (26e). For example, in German, though both voiced obstruents and sonorants are phonetically voiced, only voiced obstruent devoice at the right edge of a prosodic word. One may grant that the distinction between obstruents and sonorants is a difference in phonological representations, however, by making use of such distinction, a view of category is implicitly implemented.

4.1.2 Previous explanations for incomplete neutralization

Recall in Section 1.2.1 that the definition of incomplete neutralization is two-fold and involves both the phonology and the phonetics. A neutralization process should be classified as incomplete neutralization only when it has been argued to be complete in the phonology but incomplete in the phonetics. ³⁴ Therefore, the explanation of incomplete neutralization can logically lie in the phonology or the phonetics or their interface. It turns out proposals have been made in all these three fields.

³⁴ What I mean here about phonological completeness is that the neutralization can be analyzed as complete under the traditional formal phonology where categorical phonological representation is assumed.

4.1.2.1 Explanations within phonology

The explanations inside phonology generally involve introducing gradience into the knowledge level (McCollum, 2019; Roettger et al., 2014; inter alia). Under such a framework, the assumption of categorical phonological representation in the Standard generative view of Phonology is dropped, and fine-grained gradient information is allowed inside phonology. A consensus has not been reached by previous studies on how to incorporate gradience inside formal phonology (Lionnet, 2017; Pierrehumbert et al., 2000; Silverman, 2006; Tucker & Warner, 2010), but McCollum (2019) argues that some form of continuously valued variables has to employed in order to do so. To apply this perspective to German final devoicing, phonology should not only direct a underlyingly voiced segment to devoice, but also state to what the degree the devoicing process should occur to distinguish the *derived* voiceless segment from its *underlying* counterpart.

Despite the fact that the observed effect of incomplete neutralization can get a straightforward explanation by incorporating gradience into phonology, by violating Occam's razor/law of parsimony (26a), the proposed new theory also becomes much weaker and predicts many more possible grammars. To appreciate this statement, under the Standard generative view of Phonology, only one grammar is possible for the final obstruent devoicing process like that in German, namely the [+voice] feature in the underlying representation should disappear completely in the surface. In contrast, under the proposed new theory of gradient phonology, an infinite number of grammars are possible, differentiating on the degree to which the devoicing is demanded to happen. The second issue with this framework is that it does not offer a satisfying explanation for the systematic disparity in effect sizes as stated in (26b). If an infinite number of grammars are available and are presumably equally possible, then it is highly unlikely that the actual distribution of grammars is like what has been discovered. The third issue is that this framework can potentially predict cases of 'over-neutralization' as stated in (26c) since phonology can directly demand to what the degree a phonological process apply in the phonetics. However, only 'incomplete neutralization' has been observed in previously examined languages, which is in contrast with this prediction. The last issue, as discussed under (26d), is that such framework cannot offer a satisfying explanation for how a feeding interaction is possible under the condition of incomplete phonetic neutralization.

4.1.2.2 Explanations inside phonology-phonetics interface

The explanations aiming to revise the mechanism of phonology-phonetics generally involves revising what phonetics can see inside phonology. As noticed by many previous researchers, the direction of incomplete neutralization is almost always towards the underlying representation before derivation (Gouskova & Hall, 2009; Van Oostendorp, 2008). Again to take the German devoicing case as an example, all examined phonetic cues of *derived* voiceless stops deviate from *underlying* voiceless stops and head towards *underlying* voiced stops. In the light of this, the proposal has been made that both underlying representation and surface representation should be available for performance. And an incomplete neutralized form is then generated by blending these two representations (Anderson, 1975; Goldrick & Blumstein, 2006; Nelson & Heinz, 2021; inter alia).³⁵ The first issue with this model is that it cannot predict when incomplete neutralization occurs and when it fails to occur. If both underlying representation and surface representation are always available to phonetics, incomplete neutralization should occur globally for any phonological processes. But in contrast to this prediction, there are reported cases of complete

³⁵ As pointed by Karthik Durvasula, under Oostendorp's view, the influence of underlying representation is directly on the surface representation, which is slightly different from Anderson and Nelson & Heinz's view that the influence of underlying representation is on the performance. These two views are highly similar in explaining the phenomenon of incomplete neutralization.

neutralization (e.g. Korean: Kim & Jongman, 1996). It is worth noting that later Lee (2016) extends the study of Kim and Jongman on manner neutralization in Korean with more speakers and measurements of more phonetic cues. Lee finds a weak effect of incomplete neutralization. Such difference in results may point to there being no complete neutralization in natural languages. However, the results difference in these two studies may also be due to language change considering there is a 20 years of time span between these two studies. The difference in results may also due to the influence of second language. Participants in Kim and Jongman's study are students of an US university, while no participants in Lee's study stayed in a English-speaking environment for more than a year.

The second issue is that this model also does not offers an satisfying explanation for the systematic disparity in effect sizes as stated in (26b) and discussed under (26c). With no constraints on the degree of influence from underlying representation, the produced sound can fall at any point on the spectrum between underlying and surface representations. This is again contrary to the observed facts. Lastly, this model is too vague to allow a understanding of the contours of derived tones in Huai'an tone sandhis. The results in Experiment 1 and 2 show that the pattern of incomplete neutralization cannot be simply seen as something intermediate between two representations. The derived Tone 3s in Experiments 1 and 2 are quite different, but characterizing their contours based on the intended surface Tone 3 and the underlying tone (Tone 1 or Tone 4 respectively) is non-trivial. The *derived* Tone 3 had an initial f0 value close to that of an *underlying* Tone 3, and the end point was similar to that of an underlying Tone 1. In contrast, in Experiment 2, the *derived* Tone 3 has an initial f0 value close to that of an *underlying* Tone 4, and the end point was intermediate between *underlying* Tone 4 and *underlying* Tone 3. These patterns cannot be outcomes of simply blending underlying representations and surface representations.

4.1.2.3 Explanations from phonetics

Finally, Braver (2019) proposed an explanation that falls in the realm of phonetics with the model of Weighted Phonetic Constraint (Flemming, 2001). Under such a constraint-based framework that is similar to Optimality Theory (Prince & Smolensky, 1993), the phonetic details are no longer just a consequence of Universal Phonetics (Chomsky & Halle, 1968). Therefore, phonetic values are not automatically decided and determined but can be different under the same phonetic context for the same piece of information transferred from phonology. And Flemming proposed that the actual phonetic value is computed by a compromise among a series of weighted constraints at the knowledge level. There are two main motivations for assuming there is a grammar at the knowledge level in the phonetics as well as in the phonology. First, by doing so, the issue of language-specific variation that is omitted by standard categorical phonological representation can be easily accounted for (Keating, 1985). Second, as pointed by Flemming (2001), potential parallels can be drawn between phonetics and phonology in many phenomena, and such parallels could be interpreted as suggesting that phonetics and phonology operates with similar mechanisms and may be treated in a unified framework. An example that is relevant to incomplete neutralization is assimilation and coarticulation. Both describe a situation where one segment neutralize towards a neighboring segment.³⁶

Utilizing this model, two constraints are proposed by Braver to solve the issue of incomplete neutralization. The first one is a Paradigm Uniformity constraint that requires the derived form to be similar to the morphologically related base (Benua, 1995; Burzio, 1994, 1998; Flemming, 1995; Kenstowicz, 1995; Kiparsky, 1978; Yu, 2007). And the second one is a markedness constraint that

³⁶ It is worth noting that by employing technical tools such as electromagnetic articulography (EMA) and eletropalatography (EPG), consistent gestural differences have been found between coarticulation and assimilation (Shaw et al., 2021 using EMA; Solé, 2002 using EPG), which means Flemming's claim is problematic. I put aside this flaw in the main text and continue the discussion to show other problems of the model by Flemming.

requires a derived form to be similar to a completely neutralized target. With the above constraints, phonetic incomplete neutralization is generated as a compromise between Paradigm Uniformity constraint and the markedness constraint. Also by dropping the principle of strict domination of constraints in classic Optimality Theory, different degrees of incomplete neutralization can be achieved by varying relative weight of constraints in the compromise. Therefore, similar to the previous two models, this model also does not provide a satisfying explanation for the systematic disparity in effect sizes as stated in (26b).

It is worth noting that the model of weighted phonetic constraint is similar to the gradience phonology model in the sense that both models attempt to obscure the line between phonology and phonetics. The model of weighted phonetic constraint introduces Optimality Theory style computation at the knowledge level into phonetics, while the gradient phonology model introduces gradience that is typical in phonetics into phonology. The consequences are also similar. Both types of models are much more complicated and weaker than traditional theories where categorical phonological representation and Universal Phonetics are assumed.

To summarize for Section 4.1.2, previous explanations inside phonology or phonetics generally involve complicating the overall knowledge system of phonology-phonetics greatly. Other explanations that do not involve changing the knowledge system lie inside phonology-phonetics interface mechanism, but they often cannot predict when incomplete neutralization occurs as well as the exact effect size (Anderson, 1975; Goldrick & Blumstein, 2006; Nelson & Heinz, 2021; inter alia). Therefore all previous explanations are not perfectly satisfying and the explanation for the phenomenon of incomplete neutralization remains an open question. I will then propose my explanation in Section 4.1.3 and present the empirical data that can support. I will also present a new experiment in Chapter 5 to provide evidence for my explanation.

4.1.3 The current explanation on incomplete neutralization

For this dissertation, I propose a different *performance* explanations for incomplete neutralization cases with small effect size and with large effect size.

It is worth noting that I claimed in Chapter 1 and Chapter 2 that incomplete neutralization cases with small effect size may not even be interpreted as incomplete phonological neutralization. Native speakers may not be able to distinguish phonological categories using such small phonetic difference and therefore are likely to analyze them to be in the same category in phonology anyway. However, also as stated in this Chapter 2, some previous studies on incomplete neutralization have shown that phonetic differences that are smaller than the relevant Just Noticeable Difference are still perceptually distinguishable (Port & O'Dell, 1985; Warner et al., 2004; inter alia). However, since stimuli in elicitation task always contain the minimal pairs in these studies, an unnatural speech may be brought out where the contrast between derived form and its underlying counterpart is exaggerated. And such unnatural speech may be the reason why native speakers can distinguish phonological categories in the following perceptual task. This means native speakers may not use natural speech where the phonetic contrast is rather small to distinguish phonological categories.

To be consistent with previous literature where incomplete neutralization cases with small effect size are typically still called incomplete neutralization, I continue to use the term 'incomplete neutralization with small effect size'. I employ Just Noticeable Difference as the cutting point between small effect size and large effect size. An incomplete neutralization with an effect size that is smaller than the corresponding Just Noticeable Difference will be called 'incomplete neutralization cases with small effect size'. In contrast, an incomplete neutralization with an effect size that is smaller than Just Noticeable Difference will be called 'incomplete neutralization with an effect size that is smaller than Just Noticeable Difference will be called 'incomplete neutralization with an effect size that is smaller than Just Noticeable Difference will be called 'incomplete neutralization with an effect size that is smaller than Just Noticeable Difference will be called 'incomplete neutralization with an effect size'.

For previously widely discovered cases of incomplete neutralization with small effect size (e.g. final devoicing in German, Dutch and Russian), phonological planning effect (Kilbourn-Ceron & Goldrick, 2021; Tanner et al., 2017; Wagner, 2012), as a performance factor, can offer a satisfying explanation without invoking any changes to the categorical phonological knowledge (Durvasula, 2021). The central claim of the explanation is that speakers incrementally plan out the phonological contents beyond the current morpheme or word. When this situation occurs and the phonological details of the next morpheme or word are not immediately available, the underlying representation of the current word will be planned as it is even though the described condition of a phonological process is met. As time transpires, when the phonological details of the next word become available, the phonological process applies and another surface representation emerges and can be planned. Therefore, speakers can have two antagonistic planned surface representations for the same underlying representation at the same time. And the output in production will be a blend of the two surface representations. Again to take German final devoicing as an example, the recently planned voiceless obstruent will blend with the previously surfaced voiced obstruent, causing the output to be more voiced than a *underlying* voiceless obstruent and resulting in incomplete neutralization. As Durvasula points out, the effect of more recently planned surface representation is likely stronger due to a recency effect, so the output in production is predicted to be closer to a typical surface representation that undergoes phonological processes, which results in a small effect size in incomplete neutralization. Since the mechanism I will propose for incomplete neutralization with a large effect size may also be understood as a phonological planning effect, I use to term 'Progressive Planning Effect' to refer to the mechanism introduced in previous literature (Kilbourn-Ceron & Goldrick, 2021; Tanner et al., 2017; Wagner, 2012). Again, such a mechanism states that speakers incrementally plan out the phonological contents beyond the current morpheme or word.

In contrast, the large effect size in incomplete neutralization is rooted in phonological processes that are inherently optional. In other words, optionality is the triggering factor of incomplete neutralization with a large effect size, and I propose that only inherently optional processes can have a large effect size in phonetic incomplete neutralization. It is worth clarifying that I am not claiming that effect size is correlated with application rate. I will talk more about this in Chapter 6 under Section 6.3. I believe other factors, especially phonological representations, may play an important role in the observed effect size of incomplete neutralization in cases of incomplete neutralization with a large effect size. In contrast with previously introduced 'Progressive Planning Effect', I will use to term 'Contemporary Planning Effect' to refer to planning effect caused by optionality.

Incomplete neutralization with a large effect size can be modelled under Heinz's (2020) theoretical framework where phonological processes can have multiple outputs simultaneously. When one of them is implemented by the phonetics, the other outputs still exert a substantial influence on the planning and the subsequent implementation. As a result, the implemented surface representation moves closer to other possible surface representations in production, which results in incomplete neutralization with a large effect size. In Huai'an Tone 1 sandhi, *underlying* Tone 1 can surface as it is or undergoes Tone 1 sandhi to become Tone 3. When Tone 3 is implemented by the phonology, the other possible surface representation (Tone 1) still plays an important role in production, causing *derived* Tone 3 to deviate from *underlying* Tone 3 and become similar to *underlying* Tone 1. The key difference between the current scenario and the previous scenario that causes a small effect size is that here the other possible surface representations are valid outputs
of the phonology. In Huai'an, a Tone 1 is also a valid surface representation even before Tone 1, while in German, a voiced obstruent is not a valid surface representation at the right edge of a prosodic word. Since Tone 1 and Tone 3 are both valid surface representations which are simultaneously generated by the phonology, Tone 1 can exert a strong influence on the planning and the implementation when Tone 3 is chosen to be implemented by the phonetics.

The current explanation satisfies all the desiderata listed in Section 4.1.1. Since only performance factors are employed to explain incomplete neutralization, the relatively simple phonological framework that assumes categorical representation can be kept (26a). Also, since I offer separate explanations for incomplete neutralization with large effect size and incomplete neutralization with small effect size, the distribution of effect sizes among discovered cases can be naturally explained (26b). With regard to why 'over-neutralization' is never observed in examined languages, I argue that incomplete neutralization is always caused by underlying category playing a role in speech production in the examined cases. For incomplete neutralization cases with small effect size, the underlying representation accidently surface due to phonological planning effect; While for Huai'an cases where the effect sizes are large, the underlying representations can surface as they are in optional phonological processes and can still exert influence on production even when derived forms are picked by the phonology. Therefore, the derived category is phonetically always close to an underlying category and 'over-neutralization' never happens in examined cases (26c). It is worth noting that under my proposed theoretical framework, 'over-neutralization' can still appear in certain cases. Imagine a language that has a high tone category, a middle tone category and a low tone category in the phonology, and there is a phonological process stating that the underlying high tone optionally undergo tone sandhi to become either a middle tone or a low tone in the surface. When the middle tone is picked up by the phonology, according to my theory,

both the high tone and the low tone can still exert influence on the phonetics. Therefore, if the influence of low tone on speech production is stronger, the derived middle may not fall between the underlying high tone and the underlying middle tone to become a normal incomplete neutralization, and 'over-neutralization' situation may occur when derived middle tone falls between and the underlying middle tone and the underlying low tone. Such situation is of course rare in natural languages, which explains why 'over-neutralization' is never observed so far. Future research is needed to verify the existence of 'over-neutralization' in natural languages. Moreover, since categorical phonological representation is still kept in my theory, as stated in Section 4.1.1, rule process/rule interactions can be naturally accounted for (26d & e).

Available empirical evidence does seem to support separate explanations for incomplete neutralization cases with small effect size and with large effect size. Incomplete neutralization with large effect size is only found in phonological processes that are inherently optional. Besides the two tone sandhi processes in Huai'an, another case is French schwa deletion (Fougeron & Steriade, 1997). An example is shown in (27). Here both [dəʁol] in (27a) and [dʁol] in (27b) can be the surface representations for the underlying /dəʁol/. Although (27b) and (27c) are claimed to be phonologically identical. The segment [d] in (27b) where the schwa is deleted is not phonetically identical to the its underlying counterpart in (27c). Moreover, the effect size is large as shown in Figure *17*. The crucial comparison is between the two central bars in each four-bar set. And large effect sizes are observed in all measured phonetic cues. It is worth pointing out that the values of all the measurements are not available in the original paper. Another obvious worry is that there were too few participants (only 2 speakers) in the experiment. Overall, more future research is needed on the case of French schwa deletion to confirm an incomplete neutralization with large effect size case.

(27) a. de rôle [dәкol] 'role'
b. d'rôle [dкol] 'role'
c. drôle [dкol] 'funny'



Figure 17: Comparison of phonetic cues among derived and underlying forms: (a) Amount of linguopalatal contact in [d]; (b) Duration of lingual occlusion gesture of [d]; (c) Frequency of lenition of [d]

I will present a new experiment in Chapter 5 to further support the large effect size in incomplete neutralization being caused by optionality.

It is worth noting that planning effect can serve as the cause for both optionality and incomplete neutralization with large effect size and explains the correlation between them. Under the planning framework, optionality indicates that the phonological contents beyond the current word are poorly planned, which means the phonological contents of the next word may be available at a very late stage or totally unavailable when the current word is planned out. If the phonological contents of the next word are unavailable, the underlying representation will surface as it is. And if the phonological contents of the next word are available at a very late stage, the recently planned surface representation can only correct the previously surfaced underlying representation in a limited fashion, causing a large effect size in incomplete neutralization. An obvious advantage of using planning effects to explain incomplete neutralization with large effect size is that a unified explanation can be provided for incomplete neutralization regardless of the effect size. However, the predictions of planning effect are not attested in incomplete neutralization cases with large effect size.

First, Progressive Planning Effect predicts that for a derived contour tone in Huai'an, as time transpires, incomplete neutralization effect size will become smaller.³⁷ The phonological contents of the next word becomes more available at later stage of the contour, so at later stage, the surface representation that undergoes tone sandhi has a better chance to be planned out and exerts a stronger influence on the contour. However, in the cases of Huai'an Tone 1 and Tone 4 sandhis, as shown in Figure *4* and Figure *7*, the effect sizes become even larger in the later stages of contours.

Second, Progressive Planning Effect predicts the correlation between application rate and effect size of incomplete neutralization at the individual speaker level. Speakers who have a relatively narrow planning window (indicated by a relatively low application rate) should have larger effect size in incomplete neutralization than speakers have relatively wide planning window (indicated by a relatively high application rate). However, there is no such correlation as shown in

³⁷ As a remainder, I use to term 'Progressive Planning Effect' to refer to the planning mechanism introduced in previous literature (Kilbourn-Ceron & Goldrick, 2021; Tanner et al., 2017; Wagner, 2012). Such a mechanism states that speakers incrementally plan out the phonological contents beyond the current morpheme or word.

Figure 18 and Figure 19. In both figures, every dot represents a speaker. X-axis indicates the application rate of tone sandhi (number of token with tone sandhi applied/total number of token) and y-axis indicates the effect size of incomplete neutralization, which is the f0 difference between *derived* Tone 3 and *underlying* Tone 3 on raw pitch (f0 of *derived* Tone 3 - f0 of *underlying* Tone 3). An average is taken across all steps to do this subtraction calculation. Non-parametric Spearman correlation analysis shows that there is no significant correlation for both Tone 1 sandhi ($\rho = 0.48$, p = 0.24) and Tone 4 sandhi ($\rho = 0.33$, p = 0.25). Same as the data analysis process in Experiment 1 and 2, only *derived* Tone 3s that actually trigger another tone sandhi process are considered as real *derived* Tone 3s and analyzed here.



Figure 18: Relationship between Tone 1 application rate and effect size of incomplete neutralization



Figure 19: Relationship between Tone 4 application rate and effect size of incomplete neutralization

Overall, there is no clear evidence that can support planning effect being the causing effect for both optionality and incomplete neutralization with large effect size. And there is even counterevidence for the such explanation in the tonal contours. Therefore, it is more reasonable to provide separate explanations for incomplete neutralization with small effect size and incomplete neutralization with large effect size.

The final point to make in this subsection is that my proposed explanations for incomplete neutralization with small effect size and incomplete neutralization with large effect size are compatible. For phonological processes that are inherently optional, both planning effect and optionality can exert influence in the direction of incomplete neutralization at the same time because they are independent resources. Since a planning effect can only cause a small effect size, the influence of planning effect is predicted to not be obvious when optionality effect exists. Therefore it is expected by my proposed explanations that the predictions of planning effect are not attested in Huai'an tone sandhi cases.

4.2 The explanation for phonetically identical form can have different phonological behaviors

As a reminder of the finding in Chapter 3, although there is no phonetic difference with regard to all important phonetic cues between derived Tone 3 from lexical Tone 4 sandhi and derived Tone 3 from post-lexical Tone 4 sandhi in Huai'an, they have arguably different phonological behaviors in the sense that post-lexically derived Tone 3 triggers Tone 3 sandhi process at a higher rate across the boundary between subject and predicate than lexically derived Tone 3.

For such cases where functionally phonetically identical form can have arguably different phonological behaviors, an obvious explanation can certainly be that the previous analysis of phonological representations is flawed. Although it is described in previous literature (Du & Lin, 2019; Y. Wang & Kang, 2012) that Tone 4 sandhi in both lexical and post-lexical levels will result in a Tone 3 category, at least one of two described surface Tone 3s may be another tonal surface representation in the phonology. These two surface tones are then brought together in production by non-linguistic performance factors in this special phonological context, i.e. before underlying Tone 4. Although this analysis offers a straightforward explanation to the finding in Experiment 3, it will also has some other undesired consequences. First, in the case of Huai'an, this analysis would mean that there can be two highly similar phonological processes in one language. Both phonological processes apply to the same underlying toneme (Tone 4), and result in very similar surface representations that are functionally phonetically identical and trigger the same tone sandhi (Tone 3 sandhi) process. Such cases are definitely rare in natural languages. The situation will be even worse considering Tone 1 sandhi process exists in Huai'an. Similar to Tone 4 sandhi, Tone 1 sandhi also applies at both lexical and post-lexical levels. If derived Tone 3s from Tone 1 sandhi at both levels are also functionally phonetically identical while having different arguably

phonological behaviors. Then there will be another pair of highly similar phonological processes in Huai'an. Second, since there are two derived surface tones from Tone 4 that can trigger Tone 3 sandhi process (at different rates) according to this analysis, at least one of them does not belong to any underlying toneme categories. This situation means that there will be a Tone 3 sandhi process in Huai'an that is triggered by a derived tone that does not belong to any underlying tonemes. This Tone 3 sandhi process will then be highly restricted, which makes the whole analysis suspicious. Lastly, it is certainly not clear how native speakers distinguish the two highly similar surface tones. The two surface tones are phonetically identical in all phonetic cues that has been shown to be used by native speakers to distinguish phonological tones (Howie, 1976; Tupper et al., 2020; for f0; Fu & Zeng, 2000; Whalen & Xu, 1992 for intensity; Blicher et al., 1990 for duration).

Exploring other ways to explain the finding in Experiment 3, another possibility is that the prosodic structure, not the tonal representation, can account for the observed difference in triggering rate of Tone 3 sandhi. In Chapter 3, I stated that since Tone 3 sandhi applies across the boundary between subject and predicate at both levels, prosodic structure may not have an influence on Tone 3 sandhi application rate. However, other factors from prosodic structure may still play a role. Under Duanmu's framework (2007), the difference in Tone 3 sandhi application rate in Huai'an can potentially be accounted for by the interaction between stress and tone sandhi domain. Duanmu proposes that tone sandhi domain may be built according to the position of stresses in Mandarin languages. Since Mandarin languages are trochaic, a disyllabic word may induce a stress on the first syllable in Huai'an. Therefore, for the trisyllabic stimuli used in Experiment 3, the second syllable may carry a stress when the last two syllables form a word (example: [u pa-tsæ] 'Mr. Wu forcibly occupy (something).' from (23a)). Then the tone sandhi

domain tends to group the stressed syllable and the following unstressed syllable together according to Duanmu, which may leave the first syllable unparsed. Under such a situation, the second syllable optionally trigger Tone 3 sandhi on the first syllable. Under Duanmu's theoretical framework, the rhythmic structures of the stimuli used in Experiment 3 are shown in (28). The metrical stress in Huai'an turns out to be realized abstractly and therefore have no phonetic consequences on the second syllable, which explains why there is no phonetic difference between the two described derived Tone 3s. In this case, the metrical stress is abstractly realized and has no influence in the phonetics (Duanmu, 2007). This means the metrical stress is only detectable from different phonological behaviors, namely the difference in application rates of Tone 3 sandhi on the previous syllables. Such cases where phonological properties are abstract realized are certainly rare in natural languages, which makes Duanmu's theoretical framework suspicious.

(28) The rhythmic structure of the stimuli used in Experiment 3

a. Tone 4 sandhi feeds Tone 3 sandhi at the lexical level

	Х
u	pa-tsæ
Mr. Wu	forcibly occupy

'Mr. Wu forcibly occupy (something).'

b. Tone 4 sandhi feeds Tone 3 sandhi at the post-lexical level

Х		Х
u	to	zəu
Mr. Wu	chop	meat

'Mr. Wu chops meat.'

For this subsection, two possible explanations are listed for the finding in Experiment 3. And both of them have concerning flaws. Future efforts are needed to verify or falsify these two explanations with more evidence. It is worth noting that if the second explanation employing prosodic structure turns out to be valid, then the statement that phonetically identical form can have different phonological behaviors cannot hold in Huai'an. In such a situation, I will leave it to the future research to verify my hypothesis that phonetically identical form can have different phonological behaviors in natural languages.

CHAPTER 5 EXPERIMENTAL EVIDENCE FOR SEPARATE EXPLANATIONS FOR INCOMPLETE NEUTRALIZATION WITH DIFFERENT EFFECT SIZES

5.1 Experiment 4

The new experiment serves to pinpoint optionality as the cause for incomplete neutralization with a large effect size. To reach this purpose, previously identified interacting factors, including speaker group variation, word frequency, prosodic structure (boundary strength) and speech rate, need to be controlled. The new experiment compares the effect sizes of incomplete neutralization in Tone 1, Tone 4 and Tone 3 sandhis processes. I examined these 3 phonological processes using exactly the same experimental paradigm on the same group of speakers in the same language (Huai'an). Therefore the influences from all previously identified interacting factors are expected to attenuate. The results do support that only optional phonological processes (Tone 1 and Tone 4 sandhis) have large effect sizes in incomplete neutralization.

5.2 Participants

I recruited 8 native speakers of Huai'an Mandarin also via personal relationships in Huai'an City. The age range was from 41 to 59 years old. Again, to minimize the influence of Standard Mandarin, I avoided younger speakers in this study. Among them, 4 self-identified as female, and 4 as male. All the participants were born and raised in Huai'an City. These speakers have not participated in any linguistic studies before or heard about the concept of incomplete neutralization.

5.3 Stimuli

The stimuli were organized in the same way as in Experiment 1. So three groups of stimuli are developed for Tone 1 sandhi, Tone 4 sandhi and Tone 3 sandhi separately. All the stimuli patterns are shown in (29), (30) and (31). The full stimulus list is summarized in APPENDIX D.

(29) First group of stimuli for Experiment 4

Four sets of stimuli for Tone 1 [the syllables crucial for the current comparison are underlined and boldface]

- a. underlying T3 following underlying T2: $/T2 T3 T1 / \rightarrow [T2 T3 T1]$
- b. underlying T3 following underlying T3: /T3 T3 T1/ \rightarrow [T2 T3 T1]
- c. derived T3 following underlying T2: $/T2 T1 T1/ \rightarrow [T2 T3 T1] \text{ or } [T2 T1 T1]$
- d. derived T3 following underlying T3: $/T3 T1 T1 / \rightarrow [T2 \underline{T3} T1]$ or [T3 T1 T1]

(30) Second group of stimuli for Experiment 4

Four sets of stimuli for Tone 4 [the syllables crucial for the current comparison are underlined and boldface]

- a. underlying T3 following underlying T2: /T2 T3 T4/ \rightarrow [T2 T3 T4]
- b. underlying T3 following underlying T3: /T3 T3 T4/ \rightarrow [T2 T3 T4]
- c. derived T3 following underlying T2: $/T2 T4 T4 / \rightarrow [T2 T3 T4]$ or [T2 T4 T4]
- d. derived T3 following underlying T3: $/T3 T4 T4 / \rightarrow [T2 T3 T4]$ or [T3 T4 T4]

(31) Third group of stimuli for Experiment 4

Four sets of stimuli for Tone 3 [the syllables crucial for the current comparison are underlined and boldface]

- a. underlying T2 following underlying T2: $/T2 T2 T3 / \rightarrow [T2 T2 T3]$
- b. underlying T2 following underlying T3: /T3 T2 T3/ \rightarrow [T3 T2 T3] or [T2 T2 T3]
- c. derived T2 following underlying T2: $/T2 T3 T3/ \rightarrow [T2 T2 T3]$
- d. derived T2 following underlying T3: $/T3 T3 T3/ \rightarrow [T3 T2 T3]$ or [T2 T2 T3]

To ensure the accuracy of the measurements of effect sizes of incomplete neutralization as much as possible, it is important to control the surface context as well as avoiding any potential annotation mistakes.

For the Tone 1 sandhi process, as with Experiment 1, the crucial comparison is between two tones in the second syllable. To be specific, the comparison is between the underlying Tone 3 in (29b) and the derived Tone 3 in (29d). This comparison allows me to perfectly control for the surface context, while also establishing that the two tones are indeed categorical Tone 3s since they trigger Tone 3 sandhi on the preceding tone. Again, the set of possibilities also allows me to look at an underlying Tone 1 in roughly the same surface context, as in the second possibility in (29c), for visual comparison.

Similar to the Tone 1 sandhi, for the Tone 4 sandhi process, the crucial comparison is between two tones in the second syllable. To be specific, the comparison is between the underlying Tone 3 in (30b) and the derived Tone 3 in (30d).

For Tone 3 sandhi process, since there are no tone sandhi processes in current Huai'an that can be triggered by Tone 2, it is impossible to establish derived Tone 2 as categorical Tone 2. However, I will show that derived Tone 2 is phonetically highly similar to underlying Tone 2, which can provide at least some support that the annotation is appropriate. Since there is variation between Tone 2 and Tone 3 on the first syllable, potential annotation mistake can occur in (31b) and (31d). To avoid this issue, the crucial comparison here is between the underlying Tone 2 in (31a) and the derived Tone 2 in (31c). This comparison allows me to perfectly control for the surface context. It is worth noting that in Experiment 4, Tone 3 sandhi was observed to apply in long distance, and underlying 'Tone 3 Tone 2 Tone 3' can surface as 'Tone 2 Tone 2 Tone 3'. Such observation is contrary to previous analysis of Tone 3 sandhi in Standard Mandarin where Tone 3 sandhi is only applicable for adjacent Tone 3 syllables (M. Y. Chen, 2000; Duanmu, 2007). Since the crucial comparison is in the second syllable and I controlled for the variation in the first syllable when measuring the effect size of incomplete neutralization, this new observation is tangential to the current study.

Each participant produced 4 repetitions of 72 test sentences at a natural speech rate, which means each participant read a total of 288 sentences. All stimuli were randomized for each participant.

5.4 Procedure

The procedure was identical to that of previous experiments.

5.5 Measurement

The recordings were also manually annotated by the author and with the same scheme as in Experiment 2. An example is shown in Figure 20.



Figure 20: Annotation scheme of Experiment 4 (multiple tone sandhi processes)

5.6 Results and statistical modelling

The number of tokens for each possible combination of Underlying Representation and Surface Representation is summarized in Table *18*, Table *19* and Table *20*. The data used to calculate application rates are marked with boldface. 36 tokens were not marked as 'good' and excluded for the Tone 1 process, which accounts for 4.7% of all test stimuli for Tone 1 sandhi. 25 tokens were not marked as 'good' and excluded for Tone 4 process, which accounts for 3.3% of all test stimuli for Tone 4 sandhi. 40 tokens were not marked as 'good' and excluded for Tone 3 sandhi.

The application rate of Tone 1 sandhi in the second syllable is 49.6%, the application rate of Tone 4 sandhi in the second syllable is 73.0%, and the application rate of Tone 3 sandhi in the second syllable is 96.5%.^{38 39} Therefore, it is safe to categorize Tone 1 and Tone 4 sandhi processes as optional and Tone 3 sandhi process as mandatory.

UR	SR	Number of tokens
T2T3T1	T2T3T1	188
T3T3T1	T3T3T1	7
T3T3T1	T2T3T1	180
T2T1T1	T2T1T1	111
T2T1T1	T2T3T1	66
T3T1T1	T3T1T1	69
T3T1T1	T3T3T1	20
T3T1T1	T2T3T1	91

Table 18: Number of Tokens for UR and SR combination in Experiment 4

(Tone 1 sandhi; Data for calculating application rates boldfaced)

³⁸ Since I cannot verify derived Tone 2 from Tone 3 on the position of second syllable by phonological behavior under any contexts, I chose to trust my annotation completely when calculating application rates. Therefore, in calculating application rates for Tone 1 and Tone 4 sandhi processes, I did not count derived Tone 3s that did not trigger Tone 3 sandhi process as Tone 1 or Tone 4. I counted derived Tone 3s that did not trigger Tone 3 sandhi process as Tone 3. I also did not exclude cases where the first syllables are underlying Tone 2s, i.e. a context where I cannot verify Tone 3 category on the second syllable by phonological behavior. By doing so, I am consistent in calculating application rates by applying the same standard for all phonological processes examined in Experiment 4. It is also worth pointing out that, since my annotation has been shown to be reliable in Experiment 2 by checking the first syllable, it is reasonable to keep trusting my annotation in Experiment 4. I will do the same check on the first syllable in Experiment 4, which I will show later in this subsection. It is also worth noting that, as shown in Figure 23, annotated derived Tone 2 from Tone 3 in the second syllable is highly similar with underlying Tone 2 in the same context with regard to f0 contour shape. This observation again provides evidence for my annotation being accurate.

³⁹ The application rate of 96.5% is viewed by me as effectively mandatory. The 3.5% of cases where Tone 3 sandhi fails to apply can certainly be due to speech error.

UR	SR	Number of tokens
T2T3T4	T2T3T4	185
T3T3T4	T3T3T4	14
T3T3T4	T2T3T4	166
T2T4T4	T2T4T4	78
T2T4T4	T2T3T4	111
T3T4T4	T3T4T4	24
T3T4T4	T3T3T4	113
T3T4T4	T2T3T4	52

Table 19: Number of Tokens for UR and SR combination in Experiment 4

UR	SR	Number of tokens
T2T2T3	T2T2T3	178
T3T2T3	T3T2T3	86
T3T2T3	T2T2T3	90
T2T3T3	T2T3T3	9
T2T3T3	T2T2T3	179
T3T3T3	T3T3T3	0
T3T3T3	T2T3T3	4
T3T3T3	T3T2T3	39
T3T3T3	T2T2T3	143

(Tone 4 sandhi; Data for calculating application rates boldfaced)

Table 20: Number of Tokens for UR and SR combination in Experiment 4

(Tone 3 sandhi; Data for calculating application rates boldfaced)

The z-score transformed f0 contours on the crucial second syllable are shown in Figure 21, Figure 22 and Figure 23. For the Tone 1 and Tone 4 sandhi processes, the crucial comparison is between derived Tone 3 and underlying Tone 3; while for the Tone 3 sandhi process, the crucial comparison is between derived Tone 2 and underlying Tone 2. For Tone 1 and Tone 4 sandhi processes, as with previous experiments, I also present the tone contour for an underlying Tone 1/Tone 4 in the same surface context for visual comparison.

Based on visual inspection of the data, the existence of incomplete neutralization is clear for Tone 1 and Tone 4 sandhi processes. In contrast, for Tone 3 sandhi process, derived Tone 2 and underlying Tone 2 are highly similar with regard to tonal contour, but there is a visible gap between them. I will show that incomplete neutralization exists for all three tone sandhi processes using statistical modelling.⁴⁰

It is also worth noting that the contour shape of the derived Tone 3 from Tone 1 in the current experiment is different from that in Experiment 1. As a reminder, in Experiment 1, the contour shape of derived Tone 3 from Tone 1 starts as an underlying Tone 3 and ends as an underlying Tone 1. However, in the current experiment, the starting point of derived Tone 3 from Tone 1 is between underlying Tone 3 and underlying Tone 1, the end point seems to be close to that of underlying Tone 1 but there is still clear gap between them. The contour shape of derived Tone 3 from Tone 3 from Tone 3 from Tone 4 in the current experiment remains consistent with that in Experiment 2.

⁴⁰ Again, to further address the concern that incomplete neutralization patterns identified in Experiment 4 may arise as a result of averaging the outcomes of an optional phonological process, the distributions of underlying Tone 1, derived Tone 3, and derived Tone 3 are shown for each time step in APPENDIX H, and the distributions of underlying Tone 4, derived Tone 3, and derived Tone 3 are shown for each time step in APPENDIX I. Again, crucially, the derived Tone 3 distribution is generally uni-modal, and distinct from the other two distributions, across the time-steps. Thus, there is no evidence of an averaging artifact over optional surface representations for the derived Tone 3 cases.



Figure 21: Contours comparison of the second syllable in Experiment 4 (Tone 1 sandhi)

(Error bars indicate standard error)



Figure 22: Contours comparison of the second syllable in Experiment 4 (Tone 4 sandhi)

(Error bars indicate standard error)



Figure 23: Contours comparison of the second syllable in Experiment 4 (Tone 3 sandhi) (Error bars indicate standard error)

The modelling method remains the same as in previous experiments. The observation of incomplete phonetic neutralization is supported by model comparisons for all tone sandhi processes (Tone 1/Tone 4/Tone 3 sandhis).

For Tone 1 sandhi process, the addition of a tone sandhi condition improves the model on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=455.19$, p<0.01), the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=6.77$, p<0.01) and the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)=7.51$, p<0.01). Figure 24 shows how the full model (Model 4) with the assumption of tone sandhi affecting every fixed effect fits the observed data. And the parameter estimates for full model are summarized in Table 21.



Figure 24: Observed data and Growth Curve Model fits for derived and underlying Tone 3 for

	Estimate	Std. Error	t	р
Intercept	0.10	0.02	4.78	<0.01
Linear	-17.60	1.52	-11.55	< 0.01
Quadratic	5.01	1.27	3.95	< 0.01
Tone Sandhi: Intercept	-0.67	0.03	-23.04	< 0.01
Tone Sandhi: Linear	0.47	1.18	0.40	0.6941
Tone Sandhi: Quadratic	4.45	1.18	3.78	< 0.01

Tone 1 sandhi process (Error bars indicate standard error)

Table 21: Parameter estimates of the full model (Model 4) for Tone 1 sandhi process with the assumption of tone sandhi affecting every fixed effect (baseline: derived Tone 3)

⁴¹ It is worth noting here that although tone sandhi condition improves the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=6.77$, p<0.01), such significance is missing in the full model (Model 4). I am not sure about the factor that causes the difference. However, the results are clear that there is incomplete neutralization in Tone 1 sandhi in the current experiment.

For Tone 4 sandhi process, the addition of a tone sandhi condition improves the model on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=929.04$, p<0.01), not on the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=0.31$, p=0.58) and on the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)=17.22$, p<0.01). Figure 25 shows how the full model (Model 4) with the assumption of tone sandhi affecting the intercept and the quadratic term fits the observed data. And the parameter estimates for full model are summarized in Table 22.



Figure 25: Observed data and Growth Curve Model fits for derived and underlying Tone 3 for Tone 4 sandhi process (Error bars indicate standard error)

	Estimate	Std. Error	t	р
Intercept	0.50	0.10	5.09	< 0.01
Linear	-16.98	1.39	-12.26	< 0.01
Quadratic	2.20	1.44	1.52	0.15
Tone Sandhi: Intercept	-1.18	0.03	-35.92	< 0.01
Tone Sandhi: Linear	0.71	1.23	0.56	0.58
Tone Sandhi: Quadratic	5.29	1.27	4.18	< 0.01

Table 22: Parameter estimates of the full model (Model 4) for Tone 4 sandhi process with the assumption of tone sandhi affecting every fixed effect (baseline: derived Tone 3)

For Tone 3 sandhi process, the addition of a tone sandhi condition improves the model on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=14.09$, p<0.01), but no on the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=1.53$, p=0.22) or on the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)=0.66$, p=0.42). Figure 26 shows how the full model (Model 4) with the assumption of tone sandhi affecting only the intercept fits the observed data. And the parameter estimates for full model are summarized in Table 23.



Figure 26: Observed data and Growth Curve Model fits for derived and underlying Tone 2 for

	Estimate	Std. Error	t	р
Intercept	< 0.01	0.03	8.26	0.98
Linear	19.92	2.79	-9.89	< 0.01
Quadratic	3.50	1.96	3.29	0.11
Tone Sandhi: Intercept	0.08	0.02	-43.58	< 0.01
Tone Sandhi: Linear	1.18	0.95	0.56	0.22
Tone Sandhi: Quadratic	0.78	0.95	4.18	0.42

Tone 3 sandhi process (Error bars indicate standard error)

Table 23: Parameter estimates of the full model (Model 4) for Tone 3 sandhi process with the assumption of tone sandhi affecting every fixed effect (baseline: derived Tone 2)

With regard to effect size, as predicted by my theory, the effect sizes of incomplete neutralization are large for optional phonological processes (Tone 1 and Tone 4 sandhis), while

the effect size of incomplete neutralization for mandatory phonological process (Tone 3) is very small.

The raw f0 difference (f0 of derived Tone 3 - f0 of underlying Tone 3 for Tone 1; f0 of derived Tone 3 - f0 of underlying Tone 3 for Tone 4; f0 of derived Tone 2 - f0 of underlying Tone 2 for Tone 3) of each step for Tone 1, Tone 4 and Tone 3 sandhis are summarized in Table 24, Table 25 and Table 26.

For Tone 1 sandhi, the mean difference in f0 between underlying Tone 3 and derived Tone 3 across all steps is 27 Hz, which is more than 4 times the Just Noticeable Difference of f0 value (7 Hz) for Mandarin speakers (Jongman et al., 2017). For Tone 4 sandhi, the mean difference in f0 between underlying Tone 3 and derived Tone 3 across all steps is 50 Hz, which is more than 7 times the Just Noticeable Difference of f0 value (7 Hz) for Mandarin speakers. Moreover, across the last 18 steps (step 3 to step 20) of Tone 1 sandhi, the f0 difference is over 22 Hz, which is more than 3 times the Just Noticeable Difference; and across all steps except the very first one (step 1 to step 20) of Tone 4 sandhi, the f0 difference is over 37 Hz, which is more than 5 times the Just Noticeable Difference, it is safe to define Tone 1 and Tone 4 sandhi processes as incomplete neutralization with large effect sizes.

In contrast, for Tone 3 sandhi, the mean difference in f0 between underlying Tone 2 and derived Tone 2 across all steps is only 1 Hz. Moreover, across all steps of Tone 3 sandhi, the f0 differences are much less than the Just Noticeable Difference. Therefore, it is safe to define Tone 3 sandhi process as incomplete neutralization with a small effect size.

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	14	11	35
1	14	12	33
2	18	13	31
3	22	14	30
4	25	15	29
5	26	16	28
6	28	17	26
7	31	18	24
8	32	19	23
9	33	20	23
10	35		
		1	

 Table 24: f0 Difference of each step in Experiment 4 (Tone 1)

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	18	11	56
1	38	12	56
2	46	13	56
3	46	14	55
4	48	15	53
5	52	16	51
6	54	17	52

 Table 25: f0 Difference of each step in Experiment 4 (Tone 4)

Table 25: (cont'd)

7	56	18	48
8	58	19	47
9	62	20	47
10	59		

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	3	11	2
1	2	12	1
2	1	13	2
3	1	14	2
4	1	15	1
5	1	16	0
6	1	17	0
7	2	18	-2
8	2	19	-2
9	2	20	-3
10	2		

 Table 26: f0 Difference of each step in Experiment 4 (Tone 3)

To summarize the results of Experiment 4, only optional phonological processes (Tone 1 and Tone 4 sandhis) have large effect sizes in incomplete neutralization. While for mandatory phonological process (Tone 3), the effect size is rather small. Again, by comparing Tone 1, Tone 4 and Tone 3 sandhis of Huai'an using exactly the same experimental paradigm on the same group of speakers, all previously identified interacting factors, including speaker group variation, word frequency, prosodic structure (boundary strength) and speech rate, are expected at least to attenuate.

Such results support my proposal that the large effect size in incomplete neutralization is rooted in phonological processes that are inherently optional. And different performance explanations are needed for incomplete neutralization cases with small effect size and incomplete neutralization cases with large effect size.

To address the concern of inaccurate annotation, as with Experiment 2, I will also show that derived Tone 2 triggered by derived Tone 3 from Tone 1 or Tone 4 is indeed phonetically highly similar with underlying Tone 2. Therefore, it is reasonable to analyze that derived Tone 2 triggered by derived Tone 3 is indeed phonologically identical to underlying Tone 2, which in turn proves that annotated derived Tone 3 from Tone 1 or Tone 4 in the second syllable is indeed phonologically identical to underlying Tone 3.

The tone contours of the z-score transformed f0 for the relevant first syllables are shown in Figure 27 and Figure 28. As with Experiment 2, I also present the tone contours for underlying Tone 3s in the first syllable that come from derived Tone 3s failing to trigger Tone 3 sandhi on the preceding syllables. By doing so, three-way visual comparisons are possible at the position of the first syllable under the same phonological environment, i.e. before derived Tone 3 (from either Tone 1 or Tone 4).



Figure 27: Contours comparison of the first syllable in Experiment 4 (Tone 1)



Figure 28: Contours comparison of the first syllable in Experiment 4 (Tone 4)

Based on the visual inspection of the data, the derived Tone 2s that undergo Tone 3 sandhi with reference to the following derived Tone 3s (from underlying Tone 1 and Tone 4) are phonetically highly similar to the corresponding underlying Tone 2s with regard to the f0 contour. The f0 contours of derived Tone 2 and underlying Tone 2 in both figures are phonetically very

different from those of corresponding underlying Tone 3s. Furthermore, as with the other tone sandhi processes discussed in this paper, there is incomplete phonetic neutralization in both cases of the derived Tone 2 and the underlying Tone 2 in the first syllable. Gaps between the derived Tone 2 and the underlying Tone 2 in both cases are obvious.

The modelling method remains the same for contour tones, and the results do support the observation of incomplete neutralization. For the case of the derived Tone 2 before the derived Tone 3 from Tone 1, the addition of a Tone Sandhi condition improves the model on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=66.41$, p=<0.01), but not on the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=3.20$, p=0.07) or the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)<0.01$, p=0.95).

For the case of the derived Tone 2 before the derived Tone 3 from Tone 4, the addition of a Tone Sandhi condition also only improves the model on the intercept as shown by comparing Model 1 and Model 2 ($x^2(1)=20.59$, p=<0.01), but not on the linear term as shown by comparing Model 2 and Model 3 ($x^2(1)=0.09$, p=0.76) or the quadratic term as shown by comparing Model 3 and Model 4 ($x^2(1)=0.10$, p=0.75).

Figure 29 and Figure 30 show how the full models (Model 4) with the assumption of tone sandhi affecting every fixed effect fit the observed data. And the parameter estimates for the full models are summarized in Table 27 and Table 28. The f0 difference (f0 of underlying Tone 2 - f0 of derived Tone 2) of each step is summarized in Table 29 and Table 30.



Figure 29: Observed data and Growth Curve Model fits for derived and underlying Tone 2 before derived Tone 3 from Tone 1 sandhi (Error bars indicate standard error)



Figure 30: Observed data and Growth Curve Model fits for derived and underlying Tone 2 before derived Tone 3 from Tone 4 sandhi (Error bars indicate standard error)

	Estimate	Std. Error	t	р
Intercept	-0.21	0.07	-3.02	0.02
Linear	15.40	3.34	4.60	< 0.01
Quadratic	2.09	1.66	1.25	0.24
Tone Sandhi: Intercept	0.24	0.03	8.26	0.13
Tone Sandhi: Linear	1.93	1.08	1.79	0.07
Tone Sandhi: Quadratic	0.06	1.07	0.06	0.95

Table 27: Parameter estimates of the full model (Model 4) with the assumption of tone sandhi affecting every fixed effect (data: Tone 1; baseline: derived Tone 2)

	Estimate	Std. Error	t	р
Intercept	-0.19	0.09	-2.13	0.06
Linear	15.69	1.93	8.11	< 0.01
Quadratic	2.25	1.23	1.82	0.09
Tone Sandhi: Intercept	0.13	0.03	4.56	< 0.01
Tone Sandhi: Linear	0.31	1.05	0.29	0.77
Tone Sandhi: Quadratic	0.33	1.04	0.32	0.75

Table 28: Parameter estimates of the full model (Model 4) with the assumption of tone sandhiaffecting every fixed effect (data: Tone 4; baseline: derived Tone 2)

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	19	11	11
1	12	12	11
2	6	13	11
3	6	14	12
4	7	15	12
5	7	16	12
6	7	17	12
7	8	18	12
8	9	19	12
9	10	20	13
10	11		

 Table 29: f0 Difference of each step for first syllable in Experiment 4 (Tone 4)

Step	f0 difference (Hz)	Step	f0 difference (Hz)
0	-1	11	-6
1	-7	12	-7
2	-10	13	-7
3	-11	14	-8
4	-11	15	-8
5	-10	16	-8
6	-8	17	-8

 Table 30: f0 Difference of each step for first syllable in Experiment 4 (Tone 4)

Table 30: (cont'd)

7	-7	18	-10
8	-6	19	-10
9	-5	20	-11
10	-5		

Despite the observed incomplete neutralization, the substantial phonetic difference between derived Tone 2 and underlying Tone 3 in both cases and the phonetic similarity between derived Tone 2 and underlying Tone 2 in both cases are difficult to account for by any mechanism known to me other than Tone 3 sandhi – it cannot simply be random variation or a co-articulatory change. Therefore, the impressionistic coding was in my opinion appropriate for Experiment 4.

5.7 Interim Discussion

Finally for this chapter, I would like to state explicitly about a potential confound in Experiment 4. As pointed out by Karthik Durvasula (personal communication), the large effect size cases are about derived Tone 3, and the small effect size case is always about derived Tone 2. Therefore, the tonal target can potentially predict a larger/smaller effect size of incomplete neutralization. Due to the lack of data, it is difficult to verify or falsify this hypothesis at the current stage, I will leave this to the future research.

CHAPTER 6 GENERAL DISCUSSION

6.1 Summary of the findings

This dissertation provides experimental evidence for the gap between phonology and phonetics from two aspects, namely phonologically identical surface forms can correspond with different phonetic distributions, and phonologically different surface forms can correspond with identical phonetic distribution. These findings obviously undermine the two statements in (1), which are also repeated here in (32). These two statements are developed from strict correspondence between phonological representations and phonetic patterns.

(32) The two statements about the strict correspondence between phonological representations and phonetic patterns

Statement 1: Phonologically identical surface forms necessarily correspond with identical phonetic distributions.

Statement 2: Phonologically different surface forms necessarily correspond with different phonetic distributions.

First, using the phenomenon of incomplete neutralization, I show that phonologically identical forms can correspond with different phonetic distributions. I provided two clear cases of incomplete neutralization based on data from Huai'an high-register tone sandhi processes. I observed robust phonetic differences (with large effect sizes) between a derived Tone 3 and an underlying Tone 3 in two independent experiments. This indicates that the observed effect is not likely to be a 'false positive' or functionally unimportant. Moreover, the cases of Huai'an

circumvent the potential interference of orthography by presenting stimuli in Chinese characters. Therefore, some previous criticisms related to experimental design and the interpretation of data do not apply to the current Huai'an evidence.

A crucial aspect of this part is that I first established that the relevant tone sandhi processes are in fact phonological processes. To establish this fact, I look at the phonological behavior of the derived tones, which to me is the best way of establishing phonological representations. More specifically, I looked at cases of tone sandhi that had feeding interactions, namely high-register tone sandhis including Tone 1 sandhi (Experiment 1) and Tone 4 sandhi (Experiment 2) feed Tone 3 sandhi in Huai'an Mandarin. This establishes the fact that the Tone 1 and Tone 4 sandhi processes are indeed cases of phonological neutralization. Despite this, I observed incomplete phonetic neutralization between underlying Tone 3 and derived Tone 3s stemming from the two tone sandhi processes. Consequently, my results establish the fact that phonologically identical forms can still be phonetically different.

Second, Statement 2 leads to a deduction that phonetic identical forms necessarily correspond with identical surface forms in phonology. Using also data from Huai'an, I show that phonetically identical form can have different phonological behaviors. Therefore the deduction of Statement 2 and Statement 2 itself cannot hold. I compared derived Tone 3 from lexical Tone 4 sandhi and derived Tone 3 from post-lexical Tone 4 sandhi, and find they are indistinguishable with regard to previously identified important phonetic cues, namely f0 contour, duration and intensity. To the best of the my knowledge, native speakers of Mandarin languages cannot solely use any other phonetic cues to distinguish phonemic tones.⁴² Therefore, I assume that phonetic identity between

⁴² It is worth noting that according to the observation of Duanmu (2007), low tone in Standard Mandarin is usually correlated with breathiness. Huai'an also has such correlation according to my observation. Therefore, in principle, native speakers of Huai'an should be able to distinguish Tone 3 with other tones by breathiness. Since f0 and breathiness can be reasonably analyzed as correlated in Huai'an, I do not list breathiness as a separate phonetic cue.
the two derived Tone 3s can be established. In this particular use, I use triggering rate of Tone 3 sandhi to indicate phonological behaviors of derived Tone 3s. First, Tone 3 sandhi applies across the boundary between subject and predicate before both lexically derived Tone 3 and post-lexically derived Tone 3. Second, I controlled the planning difficulty effect caused by long utterance. For Tone 3 sandhi to apply before both lexically derived Tone 3 and post-lexically derived Tone 3, a trisyllabic phonological phrase is needed. Overall, it is difficult to assign the difference in triggering rates to prosodic boundary between subject and predicate or utterance length, which are performance factors usually used to explain variation in phonological processes.⁴³ Therefore, I argue that phonological inequality between the two derived Tone 3s can be established between the two derived Tone 3s.

6.2 Variation in effect size

Although the crucial comparison in Experiment 4 is between incomplete neutralization sizes of Tone 1/Tone 4 sandhi and that of Tone 3 sandhi. Experiment 4 also functions to replicate Experiment 1 and 2 since almost identical experimental paradigm is employed to observe Tone 1 and Tone 4 sandhi processes. The effect sizes of incomplete neutralization among different experiments are summarized in Table *31*. It is obvious that the effect sizes in Experiment 4 are substantially larger than those in Experiment 1 and 2. Such difference may certainly be just a result of random variation. In another possibility, dialectal difference may also offer a reasonable explanation. The participants in Experiment 1 and 2 are mainly from Huaiyin District of Huai'an city while the participants in Experiment 4 are all from Qingjiangpu District of Huai'an City. There

⁴³ In Chapter 4, I talked about a possibility of explaining Huai'an situations using prosodic structures. Under Duanmu's framework, the phonological difference can be explained by the foot/stress domain. Such possibility is worth being further explored. An question needs to be answered is that why the metrical stress under Dunamu's framework can be abstractly realized and have no influence in the phonetics.

are no previous reports that languages in these two districts are different, but possibilities cannot be ruled out subtle differences exist below the conscious level of native speakers. Despite the difference in effect size, the effect sizes are always robust if Just Noticeable Difference is used as the reference. This again provides good evidence for the existence of incomplete neutralization as a phenomenon.

Experiment	Observed tone sandhi processes	Mean f0 difference (Hz)
1	Tone 1 sandhi	18
2	Tone 4 sandhi	17
4	Tone 1 sandhi	27
		- 0
4	Tone 4 sandhi	50

Table 31: The comparison of effect sizes of Tone 1/Tone 4 sandhi among different experiments

These experiments are worth being replicated again in the future with dialectal effect being controlled. By doing so, a more fruitful discussion can be expected on the issue of effect size in incomplete neutralization.

6.3 The relationship between effect size of incomplete neutralization and application rate

It is reasonable to hypothesize that the size of incomplete neutralization effect in tone sandhi is correlated with application rate of tone sandhi. Recall my proposal in the current dissertation is that incomplete neutralization with a large effect size can only appear in optional phonological processes but never in mandatory phonological process. It seems intuitive and natural to make a further claim that an optional phonological process with a lower application rate will have a larger effect size of incomplete neutralization. It may also be reasonable to claim, in the situation of feeding order, that a larger effect size of incomplete neutralization will trigger another phonological process at a lower rate. These two claims are summarized in (33). A larger effect size means the derived phonological elements are more far away from their underlying counterparts in speech production. And, under the gradient framework, such bigger difference in the phonetics may be interpreted as bigger difference in the phonology, which may be reflected in the ability of triggering another phonological process.

(33) Two further claims about the potential correlation between effect size of incomplete neutralization and application rate of tone sandhi
Claim 1: An optional phonological process with a lower application rate will have a larger effect size of incomplete neutralization.
Claim 2: A larger effect size of incomplete neutralization will trigger another phonological process at a lower rate.

Both claims are not fully supported by the data collected for this dissertation. The data are summarized in Table *32*, which include the incomplete neutralization size of Tone 1/Tone 4 sandhis, the application rates of Tone 1/Tone 4 sandhis on the second syllable and the applications of Tone 3 sandhi on the first syllable. The Tone 3 sandhi is triggered by derived Tone 3 from Tone 1/Tone 4 sandhis.

Claim 1 is buttressed by comparing Tone 1 sandhi process in different instantiations of the same comparison. Tone 1 sandhi in Experiment 4 has a lower application rate than that in Experiment 1. And as predicted by Claim 1, incomplete neutralization of Tone 1 sandhi in

Experiment 4 has a larger effect size than that in Experiment 1. However, Claim 1 is not supported by comparing Tone 4 sandhi process in different instantiations of the same comparison. Although Tone 4 sandhi has a lower application rate in Experiment 2 than that in Experiment 4, incomplete neutralization of Tone 4 sandhi in Experiment 2 has a smaller effect size than that in Experiment 4. Moreover, Claim 1 is not supported by comparing different phonological processes in the same experiment. In Experiment 4, although Tone 1 sandhi has a lower application rate than Tone 4 sandhi, the effect size of incomplete neutralization of Tone 1 sandhi is smaller than the effect size of incomplete neutralization of Tone 4 sandhi.

Claim 2 is supported by comparing different phonological processes in the same experiment. In Experiment 4, Tone 4 sandhi has a larger effect size than Tone 1 sandhi. And as predicted by Claim 2, derived Tone 3 from Tone 4 triggers Tone 3 sandhi at a lower rate than derived Tone 3 from Tone 1. However, Claim 2 is not supported by comparing the same phonological process in different installations of the same experiment. For both Tone 1 and Tone 4 sandhi processes, although the effect sizes of incomplete neutralization are larger in Experiment 4 than those in Experiment 1 and 2. Derived Tone 3 from Tone 1 or Tone 4 triggers Tone 3 sandhi at a higher rate in Experiment 4. Such results further undermine the basis of explaining incomplete neutralization under a gradient framework.

			Tone 1/Tone 4	Tone 3 sandhi
	Observed tone	Mean f0	sandhi rate on	rate triggered
Experiment		difference	the syllable of	by incomplete
	sandhi processes	(Hz) ⁴⁴	incomplete	neutralization
			neutralization	syllable
1	Tone 1 sandhi	18	72.1%	74.0%
2	Tone 4 sandhi	17	66.0%	24.2%
4	Tone 1 sandhi	27	49.6%	82.9%
4	Tone 4 sandhi	50	73.0%	31.5%

 Table 32: The comparison of effect size and application rate of Tone 1/Tone 4 sandhi among different experiments

Overall, both Claim 1 and Claim 2 are not supported by experimental data from Huai'an, however, I still cannot assert if the two claims are absolutely right or wrong at this stage. First, the amount of available data is too small. They are only from three experiments on a single language, namely Huai'an. Second, a lot of other factors may affect the data from Huai'an. These factors include variation among different experiments, variation among speaker groups, etc. To make more fruitful discussions of this topic, I urge more future research on more languages. Importantly,

⁴⁴ As a reminder, I only included derived Tone 3 tokens that actually trigger Tone 3 sandhi on the first syllable when measuring the effect size of incomplete neutralization. I did so to avoid potential issue of annotation mistake. Therefore I can make sure the measurement of effect size is as accurate as possible. In contrast, when calculating 'Tone 1/Tone 4 sandhi rate on the syllable of incomplete neutralization', I trusted my annotation completely and did not count derived Tone 3 tokens that do not trigger Tone 3 sandhi as Tone 1 or Tone 4. It is obvious from my experiments that derived Tone 3 from Tone 1 triggers Tone 3 sandhi at a substantially higher rate than derived Tone 3 from Tone 4. Therefore, if I count derived Tone 3 tokens that do not trigger Tone 3 sandhi as Tone 1 or Tone 4, the results will be heavily biased. When calculating 'Tone 3 sandhi rate triggered by incomplete neutralization syllable', I also trusted my annotation completely and did not count derived Tone 3 tokens that do not trigger Tone 3 sandhi as Tone 1 or Tone 4. The reason is obvious, if I did so, the results will always be 100%, which makes this column of data meaningless.

more detailed data need to be collected, which include both effect size of incomplete neutralization and application rate of phonological process.

6.4 Good-Enough Logic

As explicitly stated in Chapter 1, the main difficulty of using phonetic measurements to probe linguistic knowledge lies in that there are multiple performance factors that affect speech production. The number of such performance factors is definitely largely considering the expected large amount of unidentified factors that can be classified under categories like sociolinguistic interactions. Therefore, as Chomsky (1964) indicated, the effort to improve experimental techniques with the purpose of eliminating all interacting performance factors is almost meaningless. And of course there is no guarantee that phonetic measurements can necessarily inform of phonological representations.

However, in actual linguistic research, a good-enough logic is often implicitly employed. Under such a logic, the efforts of ruling out some or even all previously identified performance factors is viewed as good-enough, and the possibility of unidentified performance factors playing a role is completely ignored, until such factors are identified. As expected, such a logic often leads to a hasty conclusion about phonological representation. An obvious example from the current dissertation is incomplete neutralization. During the 40 years of research on this phenomenon, considerable efforts have been made to attenuate the potential impact from identified performance factors such as writing system as introduced in Chapter 2. As a result, many researchers consider the acoustic data from the improved experiments to be 'good enough' to directly inform of phonological knowledge. And incomplete neutralization has been interpreted to pose a challenge to the Standard generative view of Phonology where categorical phonological representation and a certain version of modular feed-forward model are assumed (Braver, 2019; Goldrick & Blumstein, 2006; McCollum, 2019; Port & Leary, 2005; Manaster Ramer, 1996; Roettger et al., 2014). However, as Huai'an cases clearly show in Experiment 1 and 2, even when all identified performance factors in previous literature are controlled, the gap between phonetic measurement and phonological knowledge is still great. And phonetic incomplete neutralization is compatible with phonological complete neutralization.

The good-enough logic is also used in other topics of linguistics where the results can be affected by an infinite number of factors. Another example that also comes from the current dissertation is the attempt to establish phonetic equality. As stated in Chapter 3, there are an infinite number of phonetic cues to be observed for any phonological elements, and phonetic identity requires identity in all phonetic cues.⁴⁵ In Huai'an, I have only showed derived Tone 3s at the lexical and the post-lexical levels are indistinguishable with regard to identified important phonetic cues including f0 contour, duration and intensity contour. An implicit logic is of course identity in these phonetic cues are 'good enough' to support the overall phonetic equivalence, which is potentially problematic. However, f0 contour is widely recognized as the main phonetic cue for lexical tone. Also, there is evidence that native speakers of Mandarin languages can only rely solely on f0 contour or duration or intensity to distinguish lexical tones. Therefore I made an argument in this dissertation that identity in f0 contour, duration and intensity may be viewed as being equivalent with general phonetic identity for Mandarin tones. I admit that there is no guarantee that my argument employing 'good enough' logic is necessarily valid. Here I am only proposing such hypothesis and leave it to the future research.

⁴⁵ This would mean that 'phonetic neutralization' or 'phonetic merger' is meaningless.

I state this underlying logic explicitly here and hope such an act will not be understood as just a criticism. The positive side of such logic should also be recognized. Although the available data may be affected by an infinite number of unidentified intervening factors, such a situation should not stop attempts to develop better theoretical framework that can account for the available data and predict other patterns. Theories based on imperfect data can still be inspiring and pave the way for better theoretical framework. And of course a theory that has prerequisites, i.e. only applicable under certain situations, is also valuable. Newton's laws of motion are based on restricted data due to the limited experimental techniques in the 17th century (Newton, 1687). Later Newton's laws are found to be largely wrong at the scale of atoms and subatomic particles. However, the large majority of scholars recognized the great value of Newton's theory, which sets up the basis for classical mechanics and inspires more advanced theories in multiple subfields of physics.

To better solve the issue of an infinite number of performance factors in linguistic research, it is worth thinking about methodology that can alleviate the influence from all performance factors across the board. A practice in Experiment 4 is to make a within-language comparison of different phonological processes. As recognized in Shaw et al. (2021), such a practice can attenuate language specific factors (including performance factors) across the board.

6.5 Not an accident

Again as commented by Chomsky (1964), the relation between phonetic measurements and phonological knowledge is rather remote, and it is hopeless to use just phonetic measurements to probe phonological knowledge even with developed data-processing techniques.

However, linking hypotheses between phonetics and phonology proposed in recent studies have been supported in multiple languages. And such success is unlikely to be just an accident, which means a direct link between phonetics and phonology may be possible under certain situations. To give an example, Shaw et al. (2021) studied a certain contrast between a single complex segment and a sequence of simplex phonological segments. The differentiation between these two phonological concepts is rather difficult on both sides of phonology and phonetics. They often involve identical units in phonology and identical gestures in speech production. Several examples in IPA symbols are shown in (34).

- (34) a. Segment sequences /pj/, /kw/, /kp/, /ps/
 - b. Complex segments $/p^{j}$, $/k^{w}$, $/\hat{kp}$, $/\hat{ps}$ /

Shaw et. al. (2021) proposed a linking hypothesis that the gestural coordinations in the phonetics are different for phonological single complex segment and phonological segment sequence. To be more specific, the gestures of a complex segment are coordinated with reference only to gesture onsets, while the gestures of segment sequence are coordinated with reference to the offset of the first gesture and the onset. This distinction is schematized in Figure *31*, in which (a) shows a complex segment timing, while (b) shows a segment sequence.



Figure 31: Gesture coordinations for single complex segment and segment sequence

The hypothesis has been shown to be valid in English and Russian. These two languages are randomly selected only because there is clear phonological evidence supporting the existence of complex segment involving palatalized consonants in Russian and the existence segment sequence case involves consonant–glide sequences in English. Such success is unlike to be just an accident although more evidence from other languages will make us more confident about this conclusion. Overall, the study of Shaw et. al. (2021) shows that it is possible in certain cases to use phonetic measurements to diagnose phonological representations.

From my point of view, although there are an infinite number of performance factors that may play a role in speech production, it is still logically possible that there are aspects of phonetic invariance in the data that are stable against performance factors. And in these aspects, phonetic measurements can reliably indicate phonological representations.

CHAPTER 7 CONCLUSION

7.1 Summary

The primary goal of this dissertation is provide experimental evidence for the gap between phonology and phonetics. I use Huai'an tone sandhi cases to undermine both statements developed from strict correspondence between phonological representations and phonetic patterns. The two statements are repeated here.

(35) Two statements about the strict correspondence between phonological representations and phonetic patterns

Statement 1: Phonologically identical forms necessarily correspond with identical phonetic distributions.

Statement 2: Phonologically different forms necessarily correspond with different phonetic distributions.

My results suggest that these seemingly obvious statements are in fact too strong. And the relationship between phonology and phonetics is rather remote. Furthermore, echoing the general advice in Roettger et al. (2014), I would like to encourage more work on the topic and on my particular claims, since the acceptance of any phenomenon should not be based on a single study or a single language, and only by accumulating converging evidence from different methodologies can we be more certain of it.

The observed phenomena in this dissertation also highlights a discrepancy between the Standard generative view of Phonology (Kenstowicz, 1994; Pierrehumbert, 2002), wherein the

output of phonological computation (the surface phonological representation) uniquely feeds into a phonetics module, and Classic generative view of Phonology, where phonology is seen as knowledge (Chomsky, 1965; Chomsky & Halle, 1965, 1968; inter alia). Note, both views represent feed-forward models, where phonological computation feeds into phonetic manifestations, but phonetic manifestations cannot feed into phonological computation. However, as per the latter view, linguistic performance is a multi-factorial problem, and linguistic knowledge (i.e., competence) is only one of the many factors involved (Chomsky, 1964, 1965; Schütze, 1996; Valian, 1982; Warner et al., 2004; inter alia).⁴⁶

My results from Huai'an are problematic for the Standard generative view of Phonology - if phonetic manifestations depend solely on the output of phonology and nothing else, then it is of course the case that such a view cannot account for cases where phonological neutralization can still result in distinctness in the phonetics or phonologically different forms can have identical phonetic distribution. However, my results are not in conflict with the Classic generative view of Phonology. Phonology, as per this latter view, is conceived of as grammatical knowledge that is used by a speaker to map a string of lexical items in a specific syntactic structure to articulations, and the use of this knowledge is affected by multiple other performance factors. Consequently, nonalignment between phonology and phonetics are predicted to widely exist.

⁴⁶ I am not aware of any explicit argumentation that has ever been put forward in support of the Standard generative view over the Classic generative view. So, I am at a loss as to precisely when and, more importantly, why this change in viewpoints occurred. Here, I simply note the discrepancy.

7.2 Lingering question

A lingering question about the Huai'an incomplete neutralization cases is why different phonetic cues neutralize in different fashions. The patterns found Experiment 1 and 2 are shown again here in Figure *32* and Figure *33*.



Figure 32: Contours comparison of the second syllable in Experiment 1 (Tone 1) (Error bars indicate standard error)



Figure 33: Contours comparison of the second syllable in Experiment 2 (Tone 4) (Error bars indicate standard error)

For Tone 1 sandhi, based on the visual inspection of the data, the derived Tone 3 seems to start as an underlying Tone 3 and ends as an underlying Tone 1. And the contour shape of the derived Tone 3 is close to that of an underlying Tone 3. While for Tone 4, the derived Tone 3 seems to start as an underlying Tone 4 and then gradually deviates from underlying Tone 4 through the whole contour.

Although the pattern of derived Tone 3 from Tone 1 is not fully replicated in Experiment 4. The almost perfect alignment found in onset positions of tones in Experiment 1 (Tone 1 sandhi) and Experiment 2 (Tone 4 sandhi) is unlikely to be just an accident. Due to the limitation of the current dissertation, an explanation cannot be provided for the tonal contour patterns and why onset and offset positions neutralize in different fashions. However, I would like to emphasize again that any proposed explanations should take into consideration of the desiderata for a theory of incomplete neutralization stated in Section 4.1.1. Most importantly, if these tonal contour patterns, it would be optimal to do so.

BIBLIOGRAPHY

- Anderson, Stephen. R. (1975). On the interaction of phonological rules of various types. *Journal* of *Linguistics*, 11(1), 39-62.
- Audacity Team. (2019). Audacity [Computer Program]. Version 2.3.2, retrieved October 3, 2019 from: https://www.audacityteam.org/.
- Badia Margarit, Antonio M. (1962). Gramática catalana, 1. Madrid: Editorial Gredos.
- Baldwin, Scott A. & Hoffmann, John P. (2002). The dynamics of self-esteem: A growth-curve analysis. *Journal of youth and adolescence*, *31*(2), 101-113.
- Bao, Zhiming. (1990). On the nature of tone [PhD Thesis]. Massachusetts Institute of Technology.
- Bao, Zhiming. (1992). Toward a typology of tone sandhi. *Annual Meeting of the Berkeley Linguistics Society*, 18(2), 1-12.
- Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steven. (2021). *Lme4: Linear Mixed-Effects Models Using 'Eigen' and S4*. https://CRAN.R-project.org/package=lme4.
- Benua, Laura. (1995). Identity Effects in Morphological truncation. In Beckman, Jill N., Dickey, Laura W. & Urbanczyk, Suzanne (eds.) *Papers in Optimality Theory*, 77-136 University of Massachusetts Occasional Papers, 18; GLSA, UMass Amherst.
- Blicher, Deborah L. & Diehl, Randy L. & Cohen, Leslie B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, *18*(1), 37-49.
- Boersma, Paul & Weenink, David. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.41, retrieved 25 March 2021 from http://www.praat.org/.
- Braver, Aaron. (2019). Modelling incomplete neutralisation with weighted phonetic constraints. *Phonology*, *36*(1), 1-36.
- Braver, Aaron & Kawahara, Shigeto. (2016). Incomplete neutralization in Japanese monomoraic lengthening. In *Proceedings of the Annual Meetings on Phonology* (Vol. 2).
- Brown, Roger & McNeill, David. (1966). The "tip of the tongue" phenomenon. *Journal of verbal learning and verbal behavior*, 5(4), 325-337.
- Burzio, Luigi. (1994). Principles of English Stress. Cambridge University Press.

Burzio, Luigi. (1998). Multiple Correspondence. Lingua, 104(1-2), 79-109.

- Bybee, Joan L. (1994). A view of phonology from a cognitive and functional perspective. *Cognitive Linguistics*, 5(4), 285-305.
- Chan, Marjorie K. (1991). Contour-tone spreading and tone sandhi in Danyang Chinese. *Phonology*, 237-259.
- Chao, Yuen-Ren. (1930). A System of Tone-Letters. Le Maître Phonétique, 30: 24-27.
- Charles-Luce, Jan & Dinnsen, Daniel A. (1987). A reanalysis of Catalan devoicing. *Journal of Phonetics*, 15(2), 187-190.
- Chen, Matthew Y. (1991). An overview of tone sandhi phenomena across Chinese dialects. *Journal of Chinese Linguistics Monograph Series*, *3*, 111-156.
- Chen, Matthew Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge University Press.
- Chen, Si & Zhang, Caicai & McCollum, Adam G. & Wayland, Ratree. (2017). Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal languages. *Speech Communication*, 88, 17-38.
- Chen, Si & Li, Bin. (2021). Statistical modeling of application completeness of two tone sandhi rules. *Journal of Chinese Linguistics*, 49(1), 106-141.
- Chinese Government. (1958). Hanyu pinyin fang'an [Scheme of the Chinese Phonetic Alphabet].
- Chomsky, Noam. (1964). The development of grammar in child language: Formal discussion. *Monographs of the Society for Research in Child Development*, 29, 35-39.
- Chomsky, Noam. (1965). Aspects of the theory of syntax. Cambridge, Massachusetts: MIT Press.
- Chomsky, Noam & Halle, Morris. (1965). Some controversial questions in phonological theory. *Journal of linguistics*, 1(2), 97-138.
- Chomsky, Noam & Halle, Morris. (1968). *The sound pattern of English*. Harper and Row, New York.
- Cohn, Abigail C. (1993). Nasalisation in English: Phonology or phonetics. *Phonology*, *10*(1), 43-81.
- Di Paolo, Marianna. (1988). Pronunciation and categorization in sound change. In Ferrara, Kathleen & Brown, Becky & Walters, Keith & Baugh, John. (eds.), Linguistic change and contact: *Proceedings of the 16th Annual Conference on New Ways of Analyzing Variation in Language*, 84-92. Austin: University of Texas.

- Dinnsen, Daniel A. & Charles-Luce, Jan. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics*, *12*(1), 49-60.
- Dmitrieva, Olga. (2005). Incomplete neutralization in Russian final devoicing: Acoustic evidence from native speakers and second language learners [PhD Thesis]. University of Kansas.
- Downer, Gordon. B. (1959). Derivation by tone-change in Classical Chinese. *Bulletin of the School of Oriental and African Studies*, 22, 35-78.
- Du, Naiyan & Durvasula, Karthik. (accepted). Phonetic incomplete neutralization can be phonologically complete: Evidence from Huai'an Mandarin. *Phonology*.
- Du, Naiyan & Lin, Yen-Hwei. (2019). Obligatory Contour Principle on tone register: A case study of Huai'an Mandarin [Conference Presentation]. The 27th Annual Conference of International Association of Chinese Linguistics, Kobe City University of Foreign Studies, Kobe, Japan.
- Du, Naiyan & Lin, Yen-Hwei. (2021). Post-Lexical tone 3 sandhi domain-building in Huai'an Mandarin: multiple domain types and free application. *University of Pennsylvania Working Papers in Linguistics*, 27(1), 6.
- Duanmu, San. (1994). Against contour tone units. *Linguistic Inquiry*, 25(4), 555-608.
- Duanmu, San. (2007). The phonology of standard Chinese. OUP Oxford.
- Durvasula, Karthik. (2021). Incomplete neutralisation and understanding the distinction between competence and performance [Class handout]. Michigan State University.
- Dunbar, Ewan. (2013). *Statistical Knowledge and Learning in Phonology* [PhD Thesis]. University of Maryland, College Park.
- Ernestus, Mirjam & Baayen, R. Harald. (2006). The functionality of incomplete neutralization in Dutch: The case of past tense formation. In Goldstein, Louis M., Whalen, Douglas H. & Best, Catherine T. (eds.), *Papers in Laboratory Phonology VIII*. Mouton de Gruyter, 27-49.
- Ernestus, Mirjam & Lahey, Mybeth & Verhees, Femke & Baayen, R. Harald. (2006). Lexical frequency and voice assimilation. *Journal of the Acoustical Society of America*, *120*, 1040-1051.
- Ferreira, Fernanda & Swets, Benjamin. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46(1), 57-84.

Flemming, Edward. (1995). Auditory features in Phonology [PhD Thesis]. UCLA.

- Flemming, Edward. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, *18*(1), 7-44.
- Fougeron, Cécile & Steriade, Donca. (1997). Does Deletion of French Schwa Lead to Neutralization of Lexical Distinctions? In *Proceedings of Eurospeech*, 97(2) 943–46.
- Fourakis, Marios & Iverson, Gregory K. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica*, 41(3), 140-149.
- Fu, Qian-Jie & Zeng, Fan-Gang. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing*, 5(1), 45-57.
- Gafos, Adamantios I. & Benus, Stefan. (2006). Dynamics of phonological cognition. *Cognitive Science*, *30*(5), 905-943.
- Goldinger, Stephen D. (1996). Words and voices: Episodic traces in spoken word identification and recognition in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166-1183.
- Goldinger, Stephen D. (1997). Words and voices: Perception and production in an episodic lexicon. In Johnson, Keith & Mullennix, John W. (eds.), *Talker Variability in Speech Processing*. San Diego: Academic Press, 33-65.
- Goldrick, Matthew & Blumstein, Sheila E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*(6), 649-683.
- Gouskova, Maria & Hall, Nancy. (2009). Acoustics of epenthetic vowels in Lebanese Arabic. In Steve Parker (ed.) *Phonological argumentation: essays on evidence and motivation*. 203-225. London: Equinox.
- Hale, Mark & Kissock, Madelyn & Reiss, Charles. (2007). Microvariation, variation, and the features of universal grammar. *Lingua*, *117*(4), 645-665.
- Harris, John. (1985). *Phonological variation and change: studies in Hiberno-English*. Cambridge: Cambridge University Press.
- Hastie, Trevor & Tibshirani, Robert. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Heinz, Jeffrey. (2020). Deterministic Analysis of Optional Processes [Colloquium Talk]. University of California, Irvine.
- Hou, Jingyi. (1983). Changzhi fangyan jilüe [Notes on the Changzhi dialect]. *Fangyan [Dialect]*, *1983*(4), 260-274.

- Howie, John. M. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.
- Huang, Borong & Liao, Xudong. (2017). *Xiandai hanyu [Contemporary Chinese]*. Higher Education Press.
- Huang, C.-T. James. (1984). Phrase structure, lexical integrity, and Chinese compounds. *Journal* of the Chinese Language Teachers Association, 19(2), 53-78.
- Hyman, Larry M. (1975). Nasal States and Nasal Processes. In Ferguson Charles A. & Hyman, Larry M. & Ohala, John J. (eds.), *Nasalfest: Papers from a Symposium on Nasals and Nasalization*, 249-264. Special Publication, Stanford University Universals Project.
- Itô, Junko. (1990). Prosodic minimality in Japanese. CLS, 26(2), 213-239.
- Itô, Junko & Armin Mester. (1999). The phonological lexicon. In Tsujimura, Natsuko. (eds.), *The handbook of Japanese linguistics*. Maiden, Mass. & Oxford: Blackwell, 62-100.
- Itô, Junko & Mester, Armin. (2003). Weak layering and word binarity. In Honma, Takeru & Okazaki, Masao & Tabata, Toshiyuki & Tanaka, Shin'ichi (eds.), *A new century of phonology and phonological theory: a Festschrift for Professor Shosuke Haraguchi on the occasion of his sixtieth birthday*. Tokyo: Kaitakusha, 26-65.
- Itô, Junko & Mester, Armin. (2013). Prosodic subcategories in Japanese. Lingua, 124, 20-40.
- Japan Broadcasting Corporation. (1998). *The Japanese Language Pronunciation and Accent Dictionary*. Tokyo: NHK.
- Jassem, Wiktor & Richter, Lutosława. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, 17(4), 317-325.
- Jiao, Lidong. (2004). *Huai'an fangyan de shengdiao fenxi [An analysis of tones in Huai'an dialect]* [Master Thesis]. Tianjin Normal University.
- Jongman, Allard & Qin, Zhen & Zhang, Jie & Sereno, Joan A. (2017). Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America*, *142*(2), EL163-EL169.
- Jongman, Allard & Wang, Yue & Moore, Corinne B. & Sereno, Joan A. (2006). Perception and production of Mandarin Chinese tones. In Li, Ping & Tan, Lihai & Bates, Elizabeth & Tzeng, Ovid J. L. (eds.), *Handbook of East Asian psycholinguistics*. Cambridge, UK: Cambridge University Press, 209-217.
- Kam, Tak Him. (1977). Derivation by tone change in Cantonese: a preliminary survey. *Journal* of *Chinese Linguistics*, *5*, 186-210.

Kenstowicz, Michael J. (1994). Phonology in generative grammar. Cambridge, MA: Blackwell.

- Kenstowicz, Michael J. (1995). Morpheme Invariance and Uniform Exponence. *Ms. MIT. and Rutgers Optimality Archive.*
- Kharlamov, Viktor. (2012). Incomplete neutralization and task effects in experimentally-elicited speech: Evidence from the production and perception of word-final devoicing in Russian [PhD Thesis]. Université d'Ottawa/University of Ottawa.
- Keating, Patricia A. (1985). Universal phonetics and the organization of grammars. In Fromkin, Victoria A. (eds.), *Phonetic linguistics: essays in honor of Peter Ladefoged*, 115-132. New York: Academic Press.
- Kilbourn-Ceron, Oriana & Goldrick, Matthew. (2021). Variable pronunciations reveal dynamic intra-speaker variation in speech planning. *Psychonomic Bulletin & Review*, 28, 1365-1380.
- Kim, Hyunsoon & Jongman, Allard. (1996). Acoustic and Perceptual Evidence for Complete Neutralization of Manner of Articulation in Korean. *Journal of Phonetics*, 24(3) 295-312.
- Kiparsky, Paul. (1978). Analogical Change as a Problem for Linguistic Theory. *Studies in the Linguistic Sciences Urbana*, *III*, 8(2) 77-96.
- Labov, William. (1963). The social motivation of a sound change. Word, 19(3), 273-309.
- Labov, William. (1971). Methodology. In Dingwall, William O. (eds.), *A survey of linguistic science*, 412-497. College Park: University of Maryland Linguistics Program.
- Labov, William & Yaeger, Malcah & Steiner, Richard (1972). A quantitative study of sound change in progress. Philadelphia: U.S. Regional Survey.
- Labov, William. (1975). On the use of the present to explain the past. In Heilmann, Luigi (eds.), *Proceedings of the 11th International Congress of Linguists*. Vol. 2. Bologna: Mulino. 825-851.
- Laplace, Pierre S. (1820). *Théorie analytique des probabilités*. Courcier.
- Leben, William R. (1973). *Suprasegmental phonology* [PhD Thesis]. Massachusetts Institute of Technology.
- Lee, Kyounghee. (2016). *Neutralization of Coda Obstruents in Korean: Evidence in Production and Perception*. [PhD Thesis] Northwestern University.
- Lehiste, Ilse. (1970). Suprasegmentals. MIT Press.
- Li, Rong. (1989). Hanyu fangyan de fenqu [The geographic division of Chinese dialects]. *Fangyan [Dialect]*, *4*, 19.

- Lionnet, Florian. (2017). A theory of subfeatural representations: The case of rounding harmony in Laal. *Phonology*, *34*(3), 523-564.
- Lobanov, Boris M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, 49(2B), 606-608.
- Lü, Shuxiang. (1980). Danyang fangyan de shengdiao xitong [The tonal system of the Danyang dialect]. *Fangyan* [Dialect], 1980(2), 85-122.
- McCarthy, John J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, *17*(2), 207-263.
- Manaster Ramer, Alexis. (1996). A letter from an incompletely neutral phonologist. *Journal of Phonetics*, 24(4), 477-489.
- Mascaró, Joan. (1987). Underlying voicing recoverability of finally devoiced obstruents in Catalan. *Journal of Phonetics*, 15(2), 183-186.
- Matsui, Mayuki. (2015). Roshiago ni okeru yuuseisei no tairitsu to tairitsu no jakka: Onkyo to chikaku [Voicing contrast and contrast reduction in Russian: Acoustics and perception] [PhD Thesis]. Hiroshima University.
- McArdle, John J. & Nesselroade, John R. (2003). Growth curve analysis in contemporary psychological research. In Velicer, Wayne F. & Schinka, John A. (eds.), *Handbook of psychology: Research methods in psychology*, 2. New York: Wiley, 447-480.
- McCollum, Adam. (2019). Gradient morphophonology: Evidence from Uyghur vowel harmony. In *Proceedings of the Annual Meetings on Phonology* (Vol. 7).
- Mester, Armin. (1990). Patterns of truncation. *Linguistic Inquiry*, 21(3), 478-485.
- Milroy, James. (2001). Language ideologies and the consequences of standardization. *Journal of sociolinguistics*, *5*(4), 530-555.
- Milroy, James & John Harris. (1980). When is a merger not a merger? The MEAT/MATE problem in a present-day English vernacular. *English World-Wide*, *1*, 199-210.
- Mirman, Daniel. (2017). *Growth curve analysis and visualization using R*. Chapman and Hall/CRC.
- Mirman, Daniel & Dixon, James A. & Magnuson, James S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of memory and language*, 59(4), 475-494.

- Mori, Yoko. (2002). Lengthening of Japanese monomoraic nouns. *Journal of Phonetics*, 30(4), 689-708.
- Nelson, Scott & Heinz, Jeffrey. (2022). Incomplete neutralization and blueprint model of production. *Proceedings of the Annual Meetings on Phonology (Vol. 9)*.

Newton, Isaac. (1687). Philosophiae naturalis principia mathematica. G. Brookman.

- Nunberg, Geoffrey (1980). A falsely reported merger in eighteenth-century English: a study in diachronic variation. In Labov, William. (eds.), *Locating language in time and space*, 221-250. New York: Academic Press.
- Pierrehumbert, Janet B. (2002). Word-specific phonetics. In Gussenhoven, Carlos & Warner, Natasha. (eds.), *Papers in Laboratory Phonology VII*. Mouton de Gruyter, 101-139.
- Pierrehumbert, Janet B., Beckman, Mary E. & Ladd, D. Robert. (2000). Conceptual foundations of phonology as a laboratory science. *Phonological Knowledge: Conceptual and Empirical Issues*, 273-304.
- Piroth, Hans G. & Janker, Peter M. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics*, 32(1), 81-109.
- Port, Robert F. & Leary, Adam P. (2005). Against formal phonology. Language, 81(4), 927-964.
- Port, Robert F. & O'Dell, Michael L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13(4), 455-471.
- Poser, William J. (1990). Evidence for foot structure in Japanese. Language, 66(1), 78-105.
- Prince, Alan & Paul Smolensky. (1993). *Optimality Theory*, Rutgers University Center for Cognitive Science Technical Report 2, Rutgers University, New Brunswick, New Jersey.
- Pulleyblank, Douglas. (1986). Underspecification and low vowel harmony in Okpe. *Studies in African linguistics*, *17*(2), 119-154.
- R Core Team. (2021). R: A language and environment for statistical computing. [Computer program] Version 1.4.1106, retrieved 25 March, 2021 from https://www.rstudio.com.
- Ramsey, S. Robert. (1989). The languages of China. Princeton University Press.
- Roettger, Timo B. & Winter, Bodo & Grawunder, Sven & Kirby, James & Grice, Martine. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics*, 43, 11-25.
- Schütze, Carson T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

- Scott, Norman C. (1957). Notes on the pronunciation of Sea Dayak. Bulletin of the School of Oriental and African Studies, 20(1), 509-512.
- Scott, Norman C. (1964). Nasal consonants in Land Dayak (Bukar-Sadong). In D. Abercrombie. (eds.), *In Honour of Daniel Jones*, Longmans.
- Shaw, Jason A. & Oh, Sejin & Durvasula, Karthik & Kochetov, Alexei. (2021). Articulatory coordination distinguishes complex segments from segment sequences. *Phonology*, *38*(3), 437-477.
- Shen, Xiaonan S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics*, 18(2), 281-295.
- Silverman, Daniel. (2006). A critical introduction to phonology: Of sound, mind, and body. A&C Black.
- Slowiaczek, Louisa M. & Dinnsen, Daniel A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, *13*(3), 325-341.
- Slowiaczek, Louisa. M., & Szymanska, Helena J. (1989). Perception of word-final devoicing in Polish. *Journal of Phonetics*, 17(3), 205-212.
- Smolensky, Paul & Goldrick, Matthew & Mathis, Donald. (2014). Optimization and quantization in gradient symbol systems: A framework for integrating the continuous and the discrete in cognition. *Cognitive Science*, 38(6), 1102-1138.
- Solé, Maria-Josep. (2002). Assimilatory processes and aerodynamic factors. *Laboratory phonology* 7, 351-386.
- Tanner, James & Sonderegger, Morgan & Wagner, Michael. (2017). Production planning and coronal stop deletion in spontaneous speech. Laboratory Phonology: *Journal of the Association for Laboratory Phonology*, 8(1), p.15.
- Trudgill, Peter (1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Tucker, Benjamin V. & Warner, Natasha. (2010). What it means to be phonetic or phonological: The case of Romanian devoiced nasals. *Phonology*, 27(2), 289-324.
- Tupper, Paul & Leung, Keith & Wang, Yue & Jongman, Allard & Sereno, Joan A. (2020). Characterizing the distinctive acoustic cues of Mandarin tones. *The Journal of the Acoustical Society of America*, 147(4), 2570-2580.
- Valian, Virginia. (1982). Psycholinguistic experiment and linguistic intuition. In Simon, Thomas W. & Scholes, Robert J. (eds.), *Language, Mind, and Brain, Hillsdale*. NJ: Lawrence Erlbaum, 179-188.

- Van Oostendorp, Marc. (2008). Incomplete devoicing in formal phonology. *Lingua*, 118(9), 1362-1374.
- Wagner, Michael. (2002). The role of prosody in laryngeal neutralization. *MIT Working Papers in Linguistics*, 42, 373-392.
- Wagner, Michael. (2012). Locality in phonology and production planning. *McGill working* papers in linguistics, 22(1), 1-18.
- Wagner, Valentin & Jescheniak, Jörg D. & Schriefers, Herbert. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 423.
- Wang, Chiung-Yao & Lin, Yen-Hwei. (2011). Variation in Tone 3 Sandhi: The case of prepositions and pronouns. In *Proceedings of the 23rd North American Conference on Chinese Linguistics*, 138-155.
- Wang, Yuedong. (1998). Mixed-Effects Smoothing Spline ANOVA. *Journal of the Royal Statistical Society, Series. B*, 60(1), 159-174.
- Wang, Yifeng & Kang, Jian. (2012). Huai'an nanpian fangyan liangzizu lianxu biaodiao fenxi [An analysis of disyllabic tone sandhi in southern Huai'an dialect]. *Journal of Shenyang Institute of Engineering (Social Sciences)*, 8(3), 357-359.
- Warner, Natasha & Jongman, Allard & Sereno, Joan & Kemps, Rachèl. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, *32*(2), 251-276.
- Whalen, Douglas H. (1991). Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America*, 90(4), 2311.
- Whalen, Douglas H. (1992). Further results on the duration of infrequent and frequent words. *Journal of the Acoustical Society of America*, 91(4), 2339-2340.
- Whalen, Douglas H. & Xu, Yi. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica*, 49(1), 25-47.
- Wickham, Hadley & Averick, Mara & Bryan, Jennifer & Chang, Winston & McGowan, Lucy D. A. & François, Romain ... & Yutani, Hiroaki. (2019). Welcome to the Tidyverse. *Journal of open source software*, 4(43), 1686.
- Wright, Richard. (2004). Factors of lexical competition in vowel articulation. In Local, John & Ogden, Richard & Temple, Rosalind. (eds.), *Papers in laboratory phonology VI*, 75-87. Cambridge: Cambridge University Press.

- Xu, Yi. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95(4), 2240-2253.
- Xu, Yi. (1997). Contextual tonal variations in Mandarin. Journal of phonetics, 25(1), 61-83.
- Yang, Jilin. (1995). Zhongguo zhongxiaoxue baikequanshu [Encyclopedia for elementary school and middle school students]. Harbin Publishing House.
- Yip, Moira J. (1980). *The tonal phonology of Chinese* [PhD Thesis]. Massachusetts Institute of Technology.
- Yip, Moira J. (1989). Contour tones. *Phonology*, 6(1), 149-174.
- Yip, Moira J. (2002). Tone. Cambridge University Press.
- Yu, Alan C. L. (2007). Understanding near Mergers: The Case of Morphological Tone in Cantonese. *Phonology*, 24(1): 187-214.
- Yu, Alan C. L. (2011). Contrast reduction. In Goldsmith, John & Riggle, Jason & Yu, Alan C. L. (eds.), *The Handbook of Phonological Theory*, 291-318.
- Zhang, Ning. (1997). The avoidance of the third tone sandhi in Mandarin Chinese. *Journal of East Asian Linguistics*, 6(4), 293-338.

Sentence	IPA	Pinyin	Word-by- word gloss	Translation of the whole sentence	UR	SR
吴把车	u pa tci	wu ba che	'Mr. Wu' 'play' 'car'	'Mr. Wu plays with cars.'	T2T3T1	T2T3T1
吴鼓分	u ku fən	wu gu fen	'Mr. Wu' 'encourage' 'points'	'Mr. Wu tries to increase points.'	T2T3T1	T2T3T1
吴打车	u ta tei	wu da che	'Mr. Wu' 'call' 'car'	'Mr. Wu calls for a taxi.'	T2T3T1	T2T3T1
吴把虾	u pa xa	wu ba xia	'Mr. Wu' 'play' 'shrimp'	'Mr. Wu plays with shrimp.'	T2T3T1	T2T3T1
吴摆虾	u pe xa	wu bai xia	'Mr. Wu' 'place' 'shrimp'	'Mr. Wu places shrimp (in a plate).'	T2T3T1	T2T3T1
吴保车	u pə tei	wu bao che	'Mr. Wu' 'protect' 'car'	'Mr. Wu protects cars.'	T2T3T1	T2T3T1
吴扒车	u pa tei	wu ba che	'Mr. Wu' 'grasp' 'car'	'Mr. Wu catches cars.'	T2T1T1	T2T1T1/ T2T3T1
吴估分	u ku fən	wu gu fen	'Mr. Wu' 'estimate' 'scores'	'Mr. Wu estimates scores.'	T2T1T1	T2T1T1/ T2T3T1
吴搭车	u ta tei	wu da che	'Mr. Wu' 'take' 'cars'	'Mr. Wu gets a ride.'	T2T1T1	T2T1T1/ T2T3T1
吴扒虾	u pa xa	wu ba xia	'Mr. Wu' 'smash' 'shrimp'	'Mr. Wu smashes shrimp (to eat).'	T2T1T1	T2T1T1/ T2T3T1
吴掰虾	u pe xa	wu bai xia	'Mr. Wu' 'break off' 'shrimp'	'Mr. Wu breaks off shrimp (to eat).'	T2T1T1	T2T1T1/ T2T3T1
吴包车	u po tei	wu bao che	'Mr. Wu' 'rent' 'car'	'Mr. Wu rents cars.'	T2T1T1	T2T1T1/ T2T3T1
武把车	u pa tei	wu ba che	'Mr. Wu' 'play' 'car'	'Mr. Wu plays with cars.'	T3T3T1	T2T3T1
武鼓分	u ku fən	wu gu fen	'Mr. Wu' 'encourage' 'points'	'Mr. Wu tries to increase points.'	T3T3T1	T2T3T1
武打车	u ta tei	wu da che	'Mr. Wu' 'call' 'car'	'Mr. Wu calls for a taxi.'	T3T3T1	T2T3T1

APPENDIX A: STIMULI FOR EXPERIMENT 1 ON TONE 1 SANDHI

Table 33: Stimuli for Experiment 1 on Tone 1 sandhi

武把虾	u pa xa	wu ba xia	'Mr. Wu' 'play' 'shrimp'	'Mr. Wu plays with shrimp.'	T3T3T1	T2T3T1
武摆虾	u pe xa	wu bai xia	'Mr. Wu' 'place' 'shrimp'	'Mr. Wu places shrimp (in a plate).'	T3T3T1	T2T3T1
武保车	u po tei	wu bao che	'Mr. Wu' 'protect' 'car'	'Mr. Wu protects cars.'	T3T3T1	T2T3T1
		1	1		T	
武扒车	u pa tei	wu ba che	'Mr. Wu' 'grasp' 'car'	'Mr. Wu catches cars.'	T3T1T1	T3T1T1/ T2T3T1
武估分	u ku fən	wu gu fen	'Mr. Wu' 'estimate' 'scores'	'Mr. Wu estimates scores.'	T3T1T1	T3T1T1/ T2T3T1
武搭车	u ta tei	wu da che	'Mr. Wu' 'take' 'cars'	'Mr. Wu gets a ride.'	T3T1T1	T3T1T1/ T2T3T1
武扒虾	u pa xa	wu ba xia	'Mr. Wu' 'smash' 'shrimp'	'Mr. Wu smashes shrimp (to eat).'	T3T1T1	T3T1T1/ T2T3T1
武掰虾	u pe xa	wu bai xia	'Mr. Wu' 'break off' 'shrimp'	'Mr. Wu breaks off shrimp (to eat).'	T3T1T1	T3T1T1/ T2T3T1
武包车	u po tei	wu bao che	'Mr. Wu' 'rent' 'car'	'Mr. Wu rents cars.'	T3T1T1	T3T1T1/ T2T3T1

Table 33: (cont'd)

Sentence	IPA	Pinyin	Word-by- word gloss	Translation of the whole sentence	UR	SR
吴保税	u po suei	wu bao shui	'Mr. Wu' 'protect' 'tax'	'Mr. Wu is under bond.'	T2T3T4	T2T3T4
吴躲肉	u to zəu	wu duo rou	'Mr. Wu' 'avoid' 'meat'	'Mr. Wu avoids eating meat.'	T2T3T4	T2T3T4
吴把脉	u pa me	wu ba mai	'Mr. Wu' 'touch' 'blood vessel'	'Mr. Wu diagnoses by touching blood vessels.'	T2T3T4	T2T3T4
吴逮象	u te ciæ	wu dai xiang	'Mr. Wu' 'catch' 'elephant'	'Mr. Wu catches elephants.'	T2T3T4	T2T3T4
吴补炮	u pu pʰɔ	wu bu pao	'Mr. Wu' 'replenish' 'cannons'	'Mr. Wu replenishes the stock of cannons.'	T2T3T4	T2T3T4
吴报税	u po suei	wu bao shui	'Mr. Wu' 'declare' 'tax'	'Mr. Wu does taxes'	T2T4T4	T2T4T4/ T2T3T4
吴剁肉	u to zəu	wu duo rou	'Mr. Wu' 'chop' 'meat'	'Mr. Wu chops meat.'	T2T4T4	T2T4T4/ T2T3T4
吴罢卖	u pa me	wu ba mai	'Mr. Wu' 'stops' 'sell'	'Mr. Wu stops selling (to protest).'	T2T4T4	T2T4T4/ T2T3T4
吴带象	u te ciæ	wu dai xiang	'Mr. Wu' 'take along' 'elephant'	'Mr. Wu takes along elephants.'	T2T4T4	T2T4T4/ T2T3T4
吴布炮	u pu p ^h o	wu bu pao	'Mr. Wu' 'deploy' 'cannons'	'Mr. Wu deploys cannons.'	T2T4T4	T2T4T4/ T2T3T4
武保税	u po suei	wu bao shui	'Mr. Wu' 'protect' 'tax'	'Mr. Wu is under bond.'	T3T3T4	T2T3T4
武躲肉	u to zəu	wu duo rou	'Mr. Wu' 'avoid' 'meat'	'Mr. Wu avoids eating meat.'	T3T3T4	T2T3T4
武把脉	u pa me	wu ba mai	'Mr. Wu' 'touch' 'blood vessel'	'Mr. Wu diagnoses by touching blood vessels.'	T3T3T4	T2T3T4
武逮象	u te ciæ	wu dai xiang	'Mr. Wu' 'catch' 'elephant'	'Mr. Wu catches elephants.'	T3T3T4	T2T3T4

APPENDIX B: STIMULI FOR EXPERIMENT 2 ON TONE 4 SANDHI

Table 34: Stimuli for Experiment 2 on Tone 4 sandhi

武补炮	u pu p ^h ə	wu bu pao	'Mr. Wu' 'replenish' 'cannons'	'Mr. Wu replenishes the stock of cannons.'	T3T3T4	T2T3T4
		1				
武招税	u na sugi	wu bao shui	'Mr. Wu'	'Mr. Wu does	Т3Т4Т4	T3T4T4/
	u po suci	wa bao shar	'declare' 'tax'	taxes'	131414	T2T3T4
武则内	u to zou	wu duo rou	'Mr. Wu'	'Mr. Wu chops	T2T4T4	T3T4T4/
此利内	u to zəm	wu uuo iou	'chop' 'meat'	meat.'	131414	T2T3T4
武罢卖	u pa mɛ	wu ba mai	'Mr. Wu' 'stops' 'sell'	'Mr. Wu stops selling (to protest).'	T3T4T4	T3T4T4/ T2T3T4
武带象	u te ciæ	wu dai xiang	'Mr. Wu' 'take along' 'elephant'	'Mr. Wu takes along elephants.'	T3T4T4	T3T4T4/ T2T3T4
武布炮	u pu p ^h ə	wu bu pao	'Mr. Wu' 'deploy' 'cannons'	'Mr. Wu deploys cannons.'	T3T4T4	T3T4T4/ T2T3T4

Table 34: (cont'd)

APPENDIX C: STIMULI FOR EXPERIMENT 3 ON TONE 4 SANDHI AT THE LEXICAL AND POST-LEXICAL LEVELS

Sentence	IPA	Pinyin	Word-by- word gloss	Translation of the whole sentence	UR	SR
吴改宿	u ke su	wu gai su	'Mr. Wu' 'renovate' 'dorm'	'Mr. Wu is under bond.'	T2T3T4	T2T3T4
吴保宋	u pə suŋ	wu bao song	'Mr. Wu' 'protect' 'Mr. Song'	'Mr. Wu avoids eating meat.'	T2T3T4	T2T3T4
吴改剑	u te teĩ	wu gai jian	'Mr. Wu' 'refurbish' 'sword'	'Mr. Wu refurbishes swords.'	T2T3T4	T2T3T4
吴搞付	u kə fu	wu gao fu	'Mr. Wu' 'find fault' 'Mr. Fu'	'Mr. Wu finds fault with Mr. Fu.'	T2T3T4	T2T3T4
吴打饭	u ta fæ	wu da fan	'Mr. Wu' 'buy' 'meal'	'Mr. Wu bus meals.'	T2T3T4	T2T3T4
			'Mr Wu'			
吴盖宿	u te su	wu gai su	'build' 'dorm'	'Mr. Wu builds dorm.'	T2T4T4	T2T4T4/ T2T3T4
吴抱宋	u pə suŋ	wu bao song	'Mr. Wu' 'hug' 'Mr. Song'	'Mr. Wu hugs Mr. Song.'	T2T4T4	T2T4T4/ T2T3T4
吴带剑	u te teĩ	wu dai jian	'Mr. Wu' 'bring' 'sword'	'Mr. Wu brings sword.'	T2T4T4	T2T4T4/ T2T3T4
吴告付	u kə fu	wu gao fu	'Mr. Wu' 'sue' 'Mr. Fu'	'Mr. Wu sues Mr. Fu.'	T2T4T4	T2T4T4/ T2T3T4
吴罢战	u pa tsæ	wu ba zhan	'Mr. Wu' 'stop' 'war'	'Mr. Wu stops wars.'	T2T4T4	T2T4T4/ T2T3T4
			'Mr. Wu'	Mr. Wu is under		
武改宿	u ke su	wu gai su	'renovate' 'dorm'	bond.'	T3T3T4	T2T3T4
武保宋	u po suŋ	wu bao song	'Mr. Wu' 'protect' 'Mr. Song'	'Mr. Wu avoids eating meat.'	T3T3T4	T2T3T4
武改剑	u te teĩ	wu gai jian	'Mr. Wu' 'refurbish' 'sword'	'Mr. Wu refurbishes swords.'	T3T3T4	T2T3T4

Table 35: Tone 4 sandhi at the post-lexical level (Experiment 3)

武搞付	u ko fu	wu gao fu	'Mr. Wu' 'find fault' 'Mr. Fu'	'Mr. Wu finds fault with Mr. Fu.'	T3T3T4	T2T3T4
武打饭	u ta fæ	wu da fan	'Mr. Wu' 'buy' 'meal'	'Mr. Wu bus meals.'	T3T3T4	T2T3T4
	I	I			I	I
武盖宿	u te su	wu gai su	'Mr. Wu' 'build' 'dorm'	'Mr. Wu builds dorm.'	T3T4T4	T3T4T4/ T2T3T4
武抱宋	u pə suŋ	wu bao song	'Mr. Wu' 'hug' 'Mr. Song'	'Mr. Wu hugs Mr. Song.'	T3T4T4	T3T4T4/ T2T3T4
武带剑	u te teĩ	wu dai jian	'Mr. Wu' 'bring' 'sword'	'Mr. Wu brings sword.'	T3T4T4	T3T4T4/ T2T3T4
武告付	u kə fu	wu gao fu	'Mr. Wu' 'sue' 'Mr. Fu'	'Mr. Wu sues Mr. Fu.'	T3T4T4	T3T4T4/ T2T3T4
武罢战	u pa tsæ	wu ba zhan	'Mr. Wu' 'stop' 'war'	'Mr. Wu stops wars.'	T3T4T4	T3T4T4/ T2T3T4

Table 35: (cont'd)

Sentence	IPA	Pinyin	Word-by- word gloss	Translation of the whole sentence	UR	SR
吴改述	u ke su	wu gai shu	'Mr. Wu' 'paraphrase'	'Mr. Wu paraphrases.'	T2T3T4	T2T3T4
武保送	u po suŋ	wu bao song	'Mr. Wu' 'be admitted test-free'	'Mr. Wu is admitted (by college/university /etc.) test-free.'	T2T3T4	T2T3T4
吴改建	u ke teî	wu gai jian	'Mr. Wu' 'rebuild'	'Mr. Wu rebuilds (some buildings).'	T2T3T4	T2T3T4
吴保护	u po xu	wu bao hu	'Mr. Wu' 'protect'	'Mr. Wu protects.'	T2T3T4	T2T3T4
吴打仗	u ta tsæ	wu da zhang	'Mr. Wu' 'fight a battle'	'Mr. Wu fights a battle.'	T2T3T4	T2T3T4
吴概述	u pe su	wu gai shu	'Mr. Wu' 'summarize'	'Mr. Wu summarizes.'	T2T4T4	T2T4T4/ T2T3T4
吴报送	u po suŋ	wu bao song	'Mr. Wu' 'submit'	'Mr. Wu submits (something).'	T2T4T4	T2T4T4/ T2T3T4
吴待见	u te teĩ	wu dai jian	'Mr. Wu' 'likes'	'Mr. Wu likes (someone).'	T2T4T4	T2T4T4/ T2T3T4

Table 36: Tone 4 sandhi at the lexical level (Experiment 3)

吴告负	u ko fu	wu gao fu	'Mr. Wu' 'catch' 'elephant'	'Mr. Wu catches elephants.'	T2T4T4	T2T4T4/ T2T3T4
吴霸占	u pa tsæ	wu ba zhan	'Mr. Wu' 'lose'	'Mr. Wu lost.'	T2T4T4	T2T4T4/ T2T3T4
武改述	u ke su	wu gai shu	'Mr. Wu' 'paraphrase'	'Mr. Wu paraphrases.'	T3T3T4	T2T3T4
武保送	u po suŋ	wu bao song	'Mr. Wu' 'be admitted test-free'	'Mr. Wu is admitted (by college/university /etc.) test-free.'	T3T3T4	T2T3T4
武改建	u ke teĩ	wu gai jian	'Mr. Wu' 'rebuild'	'Mr. Wu rebuilds (some buildings).'	T3T3T4	T2T3T4
武保护	u po xu	wu bao hu	'Mr. Wu' 'protect'	'Mr. Wu protects.'	T3T3T4	T2T3T4
武打仗	u ta tsæ	wu da zhang	'Mr. Wu' 'fight a battle'	'Mr. Wu fights a battle.'	T3T3T4	T2T3T4
	1	1		1		1
武概述	u ke su	wu bai shu	'Mr. Wu' 'summarize'	'Mr. Wu summarizes.'	T3T4T4	T3T4T4/ T2T3T4
武报送	u pə suŋ	wu bao song	'Mr. Wu' 'submit'	'Mr. Wu submits (something).'	T3T4T4	T3T4T4/ T2T3T4
武待见	u te teĩ	wu dai jian	'Mr. Wu' 'likes'	'Mr. Wu likes (someone).'	T3T4T4	T3T4T4/ T2T3T4
武告负	u ko fu	wu gao fu	'Mr. Wu' 'catch' 'elephant'	'Mr. Wu catches elephants.'	T3T4T4	T3T4T4/ T2T3T4
武霸占	u pa tsæ	wu ba zhan	'Mr. Wu' 'lose'	'Mr. Wu lost.'	T3T4T4	T3T4T4/ T2T3T4

Table 36: (cont'd)

APPENDIX D: STIMULI FOR EXPERIMENT 4 ON TONE 1/TONE 4/TONE 3 SANDHIS

Sentence	IPA	Pinyin	Word-by- word gloss	Translation of the whole sentence	UR	SR
吴把车	u pa tei	wu ba che	'Mr. Wu' 'play' 'car'	'Mr. Wu plays with cars.'	T2T3T1	T2T3T1
吴鼓分	u ku fən	wu gu fen	'Mr. Wu' 'encourage' 'points'	'Mr. Wu tries to increase points.'	T2T3T1	T2T3T1
吴把虾	u pa xa	wu ba xia	'Mr. Wu' 'play' 'shrimp'	'Mr. Wu plays with shrimp.'	T2T3T1	T2T3T1
吴摆虾	u pe xa	wu bai xia	'Mr. Wu' 'place' 'shrimp'	'Mr. Wu places shrimp (in a plate).'	T2T3T1	T2T3T1
吴保车	u po tei	wu bao che	'Mr. Wu' 'protect' 'car'	'Mr. Wu protects cars.'	T2T3T1	T2T3T1
吴保书	u po su	wu bao shu	'Mr. Wu' 'protect' 'book'	'Mr. Wu protects books.'	T2T3T1	T2T3T1
吴扒车	u pa tei	wu ba che	'Mr. Wu' 'grasp' 'car'	'Mr. Wu catches cars.'	T2T1T1	T2T1T1/ T2T3T1
吴估分	u ku fən	wu gu fen	'Mr. Wu' 'estimate' 'scores'	'Mr. Wu estimates scores.'	T2T1T1	T2T1T1/ T2T3T1
吴扒虾	u pa xa	wu ba xia	'Mr. Wu' 'smash' 'shrimp'	'Mr. Wu smashes shrimp (to eat).'	T2T1T1	T2T1T1/ T2T3T1
吴掰虾	и ре ха	wu bai xia	'Mr. Wu' 'break off' 'shrimp'	'Mr. Wu breaks off shrimp (to eat).'	T2T1T1	T2T1T1/ T2T3T1
吴包车	u po tei	wu bao che	'Mr. Wu' 'rent' 'car'	'Mr. Wu rents cars.'	T2T1T1	T2T1T1/ T2T3T1
吴包书	u po su	wu bao shu	'Mr. Wu' 'cover' 'book'	'Mr. Wu covers books (with book cover).'	T2T1T1	T2T1T1/ T2T3T1
武把车	u pa tci	wu ba che	'Mr. Wu'	'Mr. Wu plays with	T3T3T1	T2T3T1
武鼓分	u ku fən	wu gu fen	'Mr. Wu' 'encourage' 'points'	'Mr. Wu tries to increase points.'	T3T3T1	T2T3T1

Table 37: Tone 1 sandhi (Experiment 4)

│ 武把虾	u pa xa	wu ba xia	'Mr. Wu' 'play' 'shrimp'	'Mr. Wu plays with shrimp.'	T3T3T1	T2T3T1
武摆虾	u pe xa	wu bai xia	'Mr. Wu' 'place' 'shrimp'	'Mr. Wu places shrimp (in a plate).'	T3T3T1	T2T3T1
武保车	u po tei	wu bao che	'Mr. Wu' 'protect' 'car'	'Mr. Wu protects cars.'	T3T3T1	T2T3T1
武保书	u po su	wu bao shu	'Mr. Wu' 'protect' 'book'	'Mr. Wu protects books.'	T3T3T1	T2T3T1
						1
武扒车	u pa tei	wu ba che	'Mr. Wu' 'grasp' 'car'	'Mr. Wu catches cars.'	T3T1T1	T3T1T1/ T2T3T1
武估分	u ku fən	wu gu fen	'Mr. Wu' 'estimate' 'scores'	'Mr. Wu estimates scores.'	T3T1T1	T3T1T1/ T2T3T1
武扒虾	u pa xa	wu ba xia	'Mr. Wu' 'smash' 'shrimp'	'Mr. Wu smashes shrimp (to eat).'	T3T1T1	T3T1T1/ T2T3T1
武掰虾	u pe xa	wu bai xia	'Mr. Wu' 'break off' 'shrimp'	'Mr. Wu breaks off shrimp (to eat).'	T3T1T1	T3T1T1/ T2T3T1
武包车	u po tei	wu bao che	'Mr. Wu' 'rent' 'car'	'Mr. Wu rents cars.'	T3T1T1	T3T1T1/ T2T3T1
武包书	u po su	wu bao shu	'Mr. Wu' 'cover' 'book'	'Mr. Wu covers books (with book cover).'	T3T1T1	T3T1T1/ T2T3T1

Table 37: (cont'd)

Sentence	IPA	Pinyin	Word-by- word gloss	Translation of the whole sentence	UR	SR
吴保税	u po suei	wu bao shui	'Mr. Wu' 'protect' 'tax'	'Mr. Wu is under bond.'	T2T3T4	T2T3T4
吴躲肉	u to zəu	wu duo rou	'Mr. Wu' 'avoid' 'meat'	'Mr. Wu avoids eating meat.'	T2T3T4	T2T3T4
吴把脉	u pa mɛ	wu ba mai	'Mr. Wu' 'touch' 'blood vessel'	'Mr. Wu diagnoses by touching blood vessels.'	T2T3T4	T2T3T4

 Table 38: Tone 4 sandhi (Experiment 4)

Table 38: (cont'd)

吴逮象	u te ciæ	wu dai xiang	'Mr. Wu' 'catch' 'elephant'	'Mr. Wu catches elephants.'	T2T3T4	T2T3T4
吴补炮	u pu p ^h ə	wu bu pao	'Mr. Wu' 'replenish' 'cannons'	'Mr. Wu replenishes the stock of cannons.'	T2T3T4	T2T3T4
吴举肉	u tcu ຊອເມ	wu ju rou	'Mr. Wu' 'hold' 'meat'	'Mr. Wu holds meat.'	T2T3T4	T2T3T4
吴报税	u po suεi	wu bao shui	'Mr. Wu' 'declare' 'tax'	'Mr. Wu does taxes.'	T2T4T4	T2T4T4/ T2T3T4
吴剁肉	u to zəu	wu duo rou	'Mr. Wu' 'chop' 'meat'	'Mr. Wu chops meat.'	T2T4T4	T2T4T4/ T2T3T4
吴罢卖	u pa me	wu ba mai	'Mr. Wu' 'stops' 'sell'	'Mr. Wu stops selling (to protest).'	T2T4T4	T2T4T4/ T2T3T4
吴带象	u te ciæ	wu dai xiang	'Mr. Wu' 'take along' 'elephant'	'Mr. Wu takes along elephants.'	T2T4T4	T2T4T4/ T2T3T4
吴布炮	u pu pʰɔ	wu bu pao	'Mr. Wu' 'deploy' 'cannons'	'Mr. Wu deploys cannons.'	T2T4T4	T2T4T4/ T2T3T4
吴拒肉	u teu ຊອເມ	wu ju rou	'Mr. Wu' 'refuse' 'meat'	'Mr. Wu refuses to eat meat'	T2T4T4	T2T4T4/ T2T3T4
武保税	u po suei	wu bao shui	'Mr. Wu' 'protect' 'tax'	'Mr. Wu is under bond.'	T3T3T4	T2T3T4
武躲肉	u to zəu	wu duo rou	'Mr. Wu' 'avoid' 'meat'	'Mr. Wu avoids eating meat.'	T3T3T4	T2T3T4
武把脉	u pa me	wu ba mai	'Mr. Wu' 'touch' 'blood vessel'	'Mr. Wu diagnoses by touching blood vessels.'	T3T3T4	T2T3T4
武逮象	u te ciæ	wu dai xiang	'Mr. Wu' 'catch' 'elephant'	'Mr. Wu catches elephants.'	T3T3T4	T2T3T4
武补炮	u pu p ^h ə	wu bu pao	'Mr. Wu' 'replenish' 'cannons'	'Mr. Wu replenishes the stock of cannons.'	T3T3T4	T2T3T4

Table 38: (c	ont'd)						
武举肉	u teu zətu	wu ju rou	'Mr. Wu' 'hold' 'meat'	'Mr. Wu holds meat.'	T3T3T4	T2T3T4	
武报税	u po suei	wu bao shui	'Mr. Wu' 'declare' 'tax'	'Mr. Wu does taxes.'	T3T4T4	T3T4T4/ T2T3T4	
武剁肉	u to zəu	wu duo rou	'Mr. Wu' 'chop' 'meat'	'Mr. Wu chops meat.'	T3T4T4	T3T4T4/ T2T3T4	
武罢卖	u pa me	wu ba mai	'Mr. Wu' 'stops' 'sell'	'Mr. Wu stops selling (to protest).'	T3T4T4	T3T4T4/ T2T3T4	
武带象	u te ciæ	wu dai xiang	'Mr. Wu' 'take along' 'elephant'	'Mr. Wu takes along elephants.'	T3T4T4	T3T4T4/ T2T3T4	
武布炮	u pu p ^h ə	wu bu pao	'Mr. Wu' 'deploy' 'cannons'	'Mr. Wu deploys cannons.'	T3T4T4	T3T4T4/ T2T3T4	
武拒肉	u teu zəti	wu ju rou	'Mr. Wu' 'refuse' 'meat'	'Mr. Wu refuses to eat meat.'	T3T4T4	T3T4T4/ T2T3T4	

Word-by-Translation of the Sentence IPA Pinyin UR SR word gloss whole sentence 'Mr. Wu' 'Mr. Wu captures T2T2T3 吴俘沈 u fu sən wu fu shen 'capture' T2T2T3 Mr. Shen.' 'Mr. Shen' 'Mr. Wu' 'Mr. Wu brings 吴携果 u ci ko wu xie guo 'bring' T2T2T3 T2T2T3 candy.' 'candy' 'Mr. Wu pastes 'Mr. Wu' u xu 吴糊口 wu hu kou 'paste' mouths (meaning: T2T2T3 T2T2T3 kʰəɯ to feed someone).' 'mouth' 'Mr. Wu' 'Mr. Wu moves 吴移沈 u i sən wu yi shen 'move' T2T2T3 T2T2T3 Mr. Shen.' 'Mr. Shen' 'Mr. Wu' 'Mr. Wu helps 吴扶许 T2T2T3 u fu cy wu fu xu 'help' T2T2T3 Mr. Xu.' 'Mr. Xu' 'Mr. Wu' 'Mr. Wu buries 吴埋果 u me ko wu mai guo T2T2T3 **T2T2T3** 'bury' 'fruits' fruits.'

Table 39: Tone 3 sandhi (Experiment 4)
Table 39: (cont'd)

吴辅沈	u fu sən	wu fu shen	'Mr. Wu' 'assist' 'Mr. Shen'	'Mr. Wu assists Mr. Shen.'	T2T3T3	T2T2T3
吴洗果	u ci ko	wu xi guo	'Mr. Wu' 'wash' 'fruits'	'Mr. Wu washes fruits.'	T2T3T3	T2T2T3
吴唬狗	u xu kətti	wuhugou	'Mr. Wu' 'frighten' 'dog'	'Mr. Wu frightens dogs'	T2T3T3	T2T2T3
吴倚沈	u i sən	wuyishen	'Mr. Wu' 'lean on' 'Mr. Shen'	'Mr. Wu leans on Mr. Shen.'	T2T3T3	T2T2T3
吴腐许	u fu cy	wufuxu	'Mr. Wu' 'spoil' 'Mr. Xu'	'Mr. Wu spoils Mr. Xu.'	T2T3T3	T2T2T3
吴买果	u mɛ ko	wumaiguo	'Mr. Wu' 'buy' 'fruits'	'Mr. Wu buys fruits.'	T2T3T3	T2T2T3
			'Mr Wu'			
武俘沈	u fu sən	wu fu shen	'capture' 'Mr. Shen'	'Mr. Wu captures Mr. Shen.'	T3T2T3	T3T2T3/ T2T2T3
武携果	u ci ko	wu xie guo	'Mr. Wu' 'bring' 'candy'	'Mr. Wu brings candy.'	T3T2T3	T3T2T3/ T2T2T3
武糊口	u xu k ^h ətt	wu hu kou	'Mr. Wu' 'paste' 'mouth'	'Mr. Wu pastes mouths (meaning: to feed someone).'	T3T2T3	T3T2T3/ T2T2T3
武移沈	u i sən	wu yi shen	'Mr. Wu' 'move' 'Mr. Shen'	'Mr. Wu moves Mr. Shen.'	T3T2T3	T3T2T3/ T2T2T3
武扶许	u fu cy	wu fu xu	'Mr. Wu' 'help' 'Mr. Xu'	'Mr. Wu helps Mr. Xu.'	T3T2T3	T3T2T3/ T2T2T3
武埋果	u me ko	wu mai guo	'Mr. Wu' 'bury' 'fruits'	'Mr. Wu buries fruits.'	T3T2T3	T3T2T3/ T2T2T3
			'Mr Wu'			
武辅沈	u fu sən	wu fu shen	'assist' 'Mr. Shen'	'Mr. Wu assists Mr. Shen.'	ТЗТЗТЗ	T3T2T3/ T2T2T3
武洗果	u ci ko	wu xi guo	'Mr. Wu' 'wash' 'fruits'	'Mr. Wu washes fruits.'	T3T3T3	T3T2T3/ T2T2T3
武唬狗	u xu kəti	wu hu gou	'Mr. Wu' 'frighten' 'dog'	'Mr. Wu frightens dogs'	T3T3T3	T3T2T3/ T2T2T3

Tuble 59. (cont d)										
武倚沈	u i sən	wu yi shen	'Mr. Wu' 'lean on' 'Mr. Shen'	'Mr. Wu leans on Mr. Shen.'	T3T3T3	T3T2T3/ T2T2T3				
武腐许	u fu cy	wu fu xu	'Mr. Wu' 'spoil' 'Mr. Xu'	'Mr. Wu spoils Mr. Xu.'	T3T3T3	T3T2T3/ T2T2T3				
武买果	u me ko	wu mai guo	'Mr. Wu' 'buy' 'fruits'	'Mr. Wu buys fruits.'	T3T3T3	T3T2T3/ T2T2T3				

Table 39: (cont'd)





Figure 34: Distribution of underlying Tone 1, derived Tone 3 and underlying Tone 3 in each step in Experiment 1







Figure 35: Distribution of underlying Tone 4, derived Tone 3 and underlying Tone 3 in each step in Experiment 2





APPENDIX G: COMPARISON OF RAW DURATION OF THE TWO TONE 3S IN EXPERIMENT 3 BY SPEAKER

Figure 36: Comparison of raw duration of the two Tone 3s in Experiment 3 by speaker (Note: the numbers on the top indicate speaker)





Figure 37: Distribution of underlying Tone 1, derived Tone 3 and underlying Tone 3 in each step in Experiment 4





APPENDIX I: DISTRIBUTION OF UNDERLYING TONE 4, DERIVED TONE 3 AND UNDERLYING TONE 3 IN EACH STEP IN EXPERIMENT 4

Figure 38: Distribution of underlying Tone 4, derived Tone 3 and underlying Tone 3 in each step in Experiment 4

