Tone, Phonation, and the Phonology-Phonetics Interface in San Martín Peras Mixtec

Eischens, Benjamin

2022

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

TONE, PHONATION, AND THE PHONOLOGY-PHONETICS INTERFACE IN SAN MARTÍN PERAS MIXTEC

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

LINGUISTICS

by

Benjamin J. Eischens

September 2022

The Dissertation of Benjamin J. Eischens
is approved:

Professor Ryan Bennett, Chair

Professor Amanda Rysling

Professor Jaye Padgett

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Tone, phonation, and the phonology-phonetics interface in San Martín Peras Mixtec

by

Benjamin J. Eischens

The phonology-phonetics interface is broadly concerned with the question of how abstract knowledge of a language's sound system influences and is influenced by the physical processes of producing and perceiving speech. There are a multitude of models that aim to address this question, many of which make distinct and even contradictory claims. This dissertation uses evidence from tone and phonation in San Martín Peras Mixtec (SMPM; ISO: jmx) to narrow the hypothesis space of viable frameworks of the interface, arguing that two basic characteristics must hold of any successful approach. First, drawing from phonetic and phonological analysis of a highly-specific process of low tone spread in SMPM, I argue that phonological and phonetic units of representation are defined at distinct levels of granularity, and therefore that they constitute distinct levels of representation. By extension, if any framework is to successfully derive low tone spread, it must incorporate different levels of representation for phonology and phonetics. Second, I show that a rate-driven, phonetic process of laryngeal reduction in SMPM influences the application of a phonological alternation, which provides evidence that the language's phonetic system is able to non-trivially influence which phonological output makes it to the surface. I argue that, in order to account for laryngeal reduction in SMPM, any successful model of the interface must allow for phonetics to influence phonology, though this influence must be indirect. Finally, I outline a model of Phonetically-Informed Candidate Selections (PICS) that embodies these characteristics and compare it with other frameworks that implement the required characteristics somewhat differently, outlining the distinct typological predictions and empirical coverage of each approach. Ultimately, this dissertation aims to constrict the hypothesis space of potential approaches to the phonology-phonetics interface, with the ultimate goal of pushing forward the field's understanding of the nature of language users' multi-tiered knowledge about the sound systems of their languages.

To Emery Martha Eischens,

born the day I defended this dissertation.

Thanks for letting me become an uncle and a doctor on the same day!

Acknowledgments

This work has only one official author, which is a shame, because it does not come close to representing the immense amount of support and encouragement I had the privilege of receiving, both during and before my time in Santa Cruz. This support has come largely from three distinct groups: The Mixtec community in California and Oaxaca, the superb group of educators who have guided me to this point, and my family and friends.

First, I would like to express my heartfelt gratitude to the members of the community of San Martín Peras and Ahuejutla, both in situ and in diaspora, whose patience and enthusiasm in sharing their language with me has made this work possible. Principal among these collaborators is Natalia Gracida Cruz, whose tireless work in the Pajaro Valley School District has elevated the Mixtec community in Santa Cruz County, and whose dedication to helping her community shines through in her efforts to raise awareness of and appreciation for the various Mixtec languages spoken in and around the Monterey Bay Area. Thank you also to Roselia Durán Cruz for her continued support and participation in the various projects that make up this dissertation. Finally, thank you to Juan Gracida Ortiz, Margarita Cruz Salazar, Irma López Basurto, and Eraclio Gracida Cruz for their warm welcome and hospitality in Ahuejutla. Tasha'vindo!

The second group of people I would like to thank are the educators whose hard work and excellent pedagogy have been invaluable to me. Thank you to Ania Lubowicz for introducing me to linguistics, and to Claire Halpert for advising me through the very

beginnings of my journey as a fieldworker and on my way to graduate school. I am also grateful to Maziar Toosarvandani for his guidance in the initial years of my work on Mixtec, and as I transitioned from syntactic research to phonology. Thank you to Jaye Padgett for his support in my development as a teacher and a linguist, and to Amanda Rysling for the numerous hours of personalized instruction in statistics, experimental methods, and the connections between theoretical and experimental work. These experiences fundamentally altered my philosophical approach to linguistics. Finally, I am indebted to Ryan Bennett for the unbelievably attentive and endless support he has given me throughout my time as his advisee. He patiently explained to a very confused, first-year Ben what ‘feeding’ and ‘bleeding’ mean, and he has provided feedback, practical guidance, and encouragement at every step of the way since. Ryan, if I end up being half the advisor and linguist you are, I’ll consider it a success.

The third group that I would like to thank is my family. My parents have always put me and my siblings before themselves, working tirelessly so that we could succeed. Thank you to my dad, Don, who delivered mail everyday without complaint, throughout more than 25 sweltering summers and frozen winters, so that I and my siblings could live a comfortable life. Thank you to my mom, Dawn, who educated me and my siblings over a period of more than 15 years, so that we could become independent and critical thinkers. Thank you to my siblings Jess, Deanna, Andy, and Matt, for always supporting and loving me no matter our differences and no matter the distance.

Finally, I am grateful to those who have made Santa Cruz feel like home: The members of my small-but-mighty cohort, Richard Bibbs and Jérémie Beauchamp, who helped and encouraged me throughout my entire time here, and those who welcomed me into the department when I was but a wee lad—Jed Pizarro-Guevara, Deniz Rudin, Erik Zyman, Tom Roberts, Jake Vincent, Kelsey Sasaki, Kelsey Kraus, and Jason Ostrove—as well as various Slug n' Friends: Jack Duff and Stephanie Rich (who I do, in fact, think are cool), Max Kaplan (thanks for the many conversations about phonology and academia), Dan Brodkin (thanks for your enthusiasm and encouragement), Molly Meehan, Kara Wernick, Morwenna Hoeks, Anjelica Casey, Yaqing Cao, Vishal Arvindam, Lalitha Balachandran, Maya Wax-Cavallaro, Nick Van Handel, Netta Ben-Meir, Andrew Angeles, Myke Brinkerhoff, Taijing Xiao, Niko Webster, Sophia Stremel, Eli Sharf, and Elifnur Ulusoy. A special thank-you to Andrew Hedding for the innumerable cups and coffee and conversations about Mixtec—may there be many, many more! Thank you to Tommy Case and Ryan Baldwin for their jokes and continued friendship, even at such a distance. And finally, my sincere gratitude and love to my partner, Sienna Ballou, for her endless encouragement and support throughout the past four years. You saw I was a phonologist before I did, you've supported me even in my most insufferable moments, and you have kept me grounded even when feverishly writing for weeks on end. You truly are one of the smartest people I know! Thank you so much for these years, and for what's to come.

Chapter 1

Introduction

A perennial issue in all sub-fields of generative linguistics is the question of how abstract linguistic knowledge is related to the real-time processes of producing and understanding speech. Traditionally, linguists have made a distinction between these two levels of description and analysis, concurring with Chomsky and Halle (1968) that a language user's abstract knowledge of their language—for example, the units of representation and combinatorics of these units—is quite separate from their knowledge of how these units are produced and perceived in real time. This distinction, often referred to as the difference between 'competence' and 'performance,' has been the object of much debate. Though most researchers agree there should be *some* dividing line between the two types of linguistic knowledge, very few agree about the shape and location of that line. This dissertation is concerned with a microcosm of this larger issue, often referred to as the phonology-phonetics interface.

1.1 The issue

Within the realm of the study of the languages' sound patterns, the distinction between abstract linguistic knowledge and the use of that knowledge in real time is relativized to the nature of language users' knowledge about the sounds of their language. On the one hand, one can account for a great deal of a language's sound system by making use of a relatively small set of distinctions, codified, for example, in sets of phonemes or phonological features. On the other hand, each of these phonemes or features has a multitude of articulatory, acoustic, and perceptual correlates that vary contextually, even among members of the same class. Consider, for example, the distinction between voiced and voiceless obstruents in English. A single, binary phonological feature [+/-voice] is sufficient to account for the phonemic contrast between voiced and voiceless obstruents (e.g., 'bat' vs. 'pat'), as well as for phonological alternations like voicing assimilation (e.g., [k^{h}æt-s] vs. [dɑg-z]). However, to describe the physical realization of the distinction between [+/-voice] obstruents, a large number of physical measures are needed. Lisker (1986) lists burst intensity, voice-onset time (VOT), closure duration, length of the preceding vowel, F0, and F1 transition, among others, as perceptual cues to stop voicing in English, as schematized in Figure 1.1. What is more, stops with an identical specification for [+/-voice] may vary in their values for these measures. For instance, VOT varies across place of articulation for stops with the same voice feature (Lisker and Abramson, 1967).

These two levels of description and analysis—the level of the [+/-voice] fea-

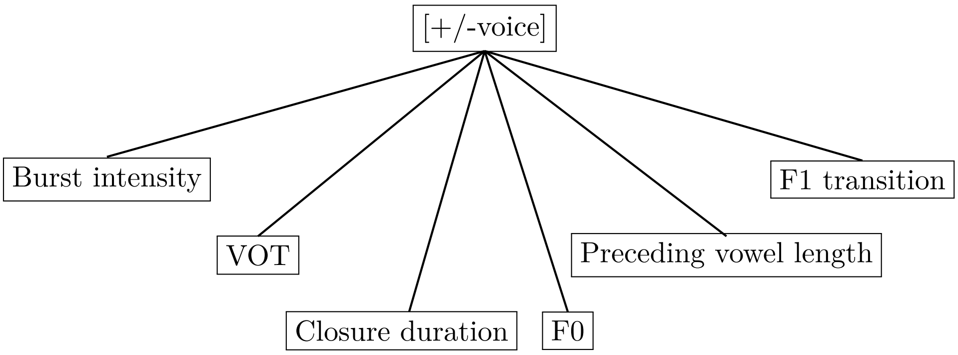


Figure 1.1: Perceptual cues to stop voicing in English

ture, and the level of the continuous, contextually-varying physical measures—are often referred to as phonology and phonetics. Phonology, generally speaking, concerns the abstract sound system of a language, described in terms of coarse-grained units like phonemes, segments, or features. Phonetics, on the other hand, is concerned with the physical realization of speech, and is usually described in terms of fine-grained, physical units. However, the existence of an inherent distinction in terms of units of representation between these two levels of representation has been questioned (Steriade, 2000; Flemming, 2001; Pierrehumbert, 2016, a.o.), and the nature of phonological and phonetic levels of representation is still up for debate. In fact, there exist a large number of proposed models of the relationship between phonology and phonetics, with most logically possible positions on the subject being occupied in some form or another. One useful way to organize these framework is along a continuum of the degree of overlap between phonology and phonetics, like that shown in Figure 1.2: On one end, there are frameworks like substance-free phonology (Reiss, 2017a), which claim that phonology and phonetics are completely distinct systems, with phonology being purely abstract and based on contrast, and the apparent interactions between phonology and phonetics being proposed to arise through phonetic pressures on diachronic sound change. On the other end of the continuum are frameworks like Flemming's (2001) unified model of the phonology and phonetics, which posits no separation at all between phonology and phonetics, and instead derives sound patterns through the interaction of constraints on contrast, articulatory effort, and perceptual distance. In a model like this, the physical measures associated with a sound are directly directly represented in the grammar. In between these two extremes, there are a large number of frameworks that make differing claims about the nature and degree of interaction between phonology and phonetics.[1]

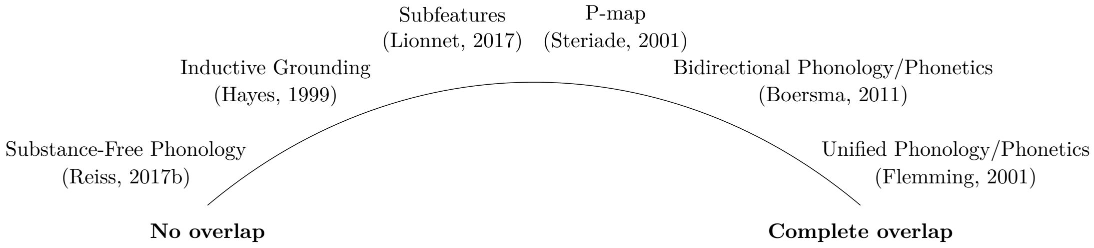


Figure 1.2: Continuum of frameworks of the phonology-phonetics interface

So, there exist myriad frameworks between two extremes, each with their own claims, predictions, strengths, and weaknesses. The goal of the present dissertation is to use evidence from tone and phonation in San Martín Peras Mixtec (SMPM, ISO: jmx) to argue for and discuss the implementation of some necessary characteristics of a successful model of the phonology-phonetics interface, with the aim of narrowing the

---

[1] This figure collapses a multi-dimensional space and, as a result, should not be taken as a literal claim about monotonic differences between the frameworks listed.

hypothesis space of potential frameworks in Figure 1.2. To that end, the dissertation strives to accomplish three major tasks, described in the following roadmap.

1.2 Roadmap

First, after a discussion of necessary language background on SMPM in Chapter 2, Chapter 3 describes and analyzes a highly specific process of tone sandhi in the language. This process, termed low tone spread, only applies in a very restrictive environment and is amenable to an analysis as being triggered by cumulative, physical coarticulatory pressures. This means that it is a prime example for analysis in terms of a 'complete overlap' framework like Flemming's (2001) unified model. However, I argue that, because of its reliance on the physical structure of sounds, this type of framework is inadequate for the purposes of analyzing low tone spread in SMPM, with the reason being that it is unable to account for an opaque relationship between low tone spread and a separate tone sandhi process in the language. Specifically, two physically-identical but derivationally-distinct configurations are treated differently by the grammar. I conclude that, in order to satisfactorily model low tone spread in SMPM, any model of the phonology-phonetics interface must necessarily posit a distinction between two levels of representation, one of which defines its units of representation in a rather coarse-grained manner (e.g., segments, features, etc.) and one that defines its units of representation in a fine-grained manner (e.g., articulatory, acoustic, or perceptual measures).

Having established that the two levels of representation must be distinct, Chapter 4 considers a separate sound pattern in SMPM that suggests that the phonology and phonetics interact synchronically. Specifically, the chapter describes and analyzes a pattern of fast-speech reduction that applies to roots with laryngealized vowels. This pattern bears many of the characteristics of a non-phonological alternation, being gradient and driven by speech rate, a factor commonly considered to be non-phonological. Despite its phonetic nature, this alternation is correlated with a change in phonological representation, namely the deletion of a mora, as evidenced by an interaction between laryngeal reduction and a process of tone sandhi in the language. I argue that the phonological deletion of a mora is inextricably linked to laryngeal reduction and that, as a result, mora deletion is a phonological processes whose driving factors are non-phonological. Because of this interaction, I conclude that any framework of the interface that is to model laryngeal reduction and mora deletion in SMPM must necessarily allow some interplay between phonology and phonetics.

The combination of the conclusions from Chapters 3 and 4 results in a set of requirements that I argue must hold of any framework of the phonology-phonetics interface, if it is to adequately model both low tone spread and laryngeal reduction in SMPM. That is, a framework must (1) make a distinction between phonological and phonetic levels of representation, and (2) allow for some interaction between those two levels. While these requirements exclude the very ends of the continuum in Figure 1.2, they are rather broad and encompass a range of proposals that make differing, often contradictory claims about the nature of the relationship between phonology and phonetics. With the aim of elucidating the ways that the required characteristics might be implemented, Chapter 5 motivates and proposes a model of interaction between phonological phonetics, dubbed Phonetically-Informed Candidate Selection (PICS). The model posits that phonetics acts as a ‘filter’ of sorts over phonological outputs, choosing from a precompiled set of potential outputs the one that is most suitable to the circumstances at hand. Because it separates phonology and phonetics, it is able to derive low tone spread, and because it allows for an interaction of the two, it is able to derive the interaction between laryngeal reduction and mora deletion. With this model specified, it is then compared to a handful of other frameworks that broadly fulfill the characteristics that I argue are necessary for any model of the interface. Through theory comparison, I show that each of these models makes different predictions and has different ranges of empirical coverage. Ultimately, I conclude that something like the PICS model is empirically-motivated and has rather broad empirical coverage, but does not encompass the whole of phonology-phonetics interactions. The emergent picture is one in which phonology and phonetics are separate systems with multiple, constrained means of interaction.

Chapter 2

Background

2.1 Introduction

In order to motivate the theoretical claims made throughout this dissertation, it is necessary to contextualize the data and arguments within the larger picture of of SMPM's sound system. To that end, this section provides a brief overview of SMPM's phonological inventory. Given that the pieces of the sound system that are most important for this dissertation are tone and phonation, this section focuses mostly on these after a brief description of general research methods and the consonant and vowel phoneme inventory of the language.

2.2 Language family and communities

SMPM is an Otomanguean language spoken by about 10,000 people in and around the municipality of San Martín Peras in western Oaxaca, Mexico (Instituto

Nacional de Estadística y Geografía, 2010). Additionally, there are an estimated 350,000 indigenous Oaxacans living in California (Rabadán and Salgado, 2018), many of whom speak one of the multitude of indigenous Oaxacan languages. Speakers of the San Martín Peras variety of Mixtec are concentrated principally in the towns of Salinas, Oxnard, and Santa María (Mendoza, 2020), as well as Watsonville. The language is part of the Southern Baja group identified by Josserand (1983). It has default VSO word order, though arguments regularly front to a pre-verbal position through various processes related to information structure (Ostrove, 2018; Mendoza, 2020), as is the case in other Mixtec languages (i.e., León Vázquez, 2017 on Yucuquimi de Ocampo Mixtec; Macaulay, 1996 for Chalcatongo Mixtec).

2.3 Research Methods

Unless otherwise marked, all SMPM data in this dissertation come from my fieldwork with SMPM language consultants, either in Watsonville, California or in Ahuejutla, Mexico. Ahuejutla is a town of about 1,000-1,500 within the municipality of San Martín Peras. Most data points come from two consultants, both of whom currently live in Watsonville, California and use SMPM on a daily basis. Consultant 1 grew up in Ahuejutla, and Consultant 2 grew up in the town of San Martín Peras, which is the main town in the municipality. The contact language for elicitation sessions with both consultants was Spanish.

Methods for general data collection included translations of single words or full phrases from Spanish to SMPM and from SMPM to Spanish, eliciting well-formedness judgments for target sentences in an appropriate discourse context and asking for repetitions, and an informal forced-choice task in which a consultant was presented with two grammatical sentences uttered by the linguist and asked which sentence sounded most natural. Methods for determining the phonemic category of tones in relevant words included eliciting target words in tone frame sentences (Pike, 1948; pp. 50-52), viewing pitch tracks of target words in frame sentences in Praat (Boersma and Weenink, 2020) and aggregate pitch plots in R (R Core Team, 2013), and using tone sandhi processes that are sensitive to certain tonal specifications. Task-specific methods are described at their relevant points throughout the dissertation. Audio was recorded on zencastr.com (48 KHz, 16-bit) using a Cooler Master MH630 headset microphone.

It is worth noting that the number of language consultants whose productions and judgments are discussed in this work is lower than what is found in most phonetic studies. The principle reason for this is that the vast majority of this dissertation research was carried out online due to restrictions on in-person meetings during the COVID-19 pandemic. Additionally, travel restrictions prevented the collection of data in Oaxaca, where it would have been feasible to work with a larger number of language consultants. In order to compensate for the low number of consultants, a large number of observations were gathered across multiple tasks whenever possible. For example, the speech rate analysis later in this chapter analyzes 453 productions from Consultant 1, the analysis of the amplitude contours of laryngelized roots in Chapter 3 is conducted over 498 productions from Consultant 1, and analysis of the gradience of laryngeal reduction in Chapter 3 examines 145 productions from Consultant 1 and 164 from Consultant 2. There are many other portions of the dissertation that rely on phonetic analysis of multiple examples, and in all such cases where aggregated data is used for analysis, the number of observations is explicitly marked. In this way, the small number of consultants for this dissertation is made up for by the collection and analysis of a high number of observations per consultant across a number of distinct tasks over a long time period.

2.4 Phoneme inventory

SMPM's consonant inventory, which has been adapted from Peters (2018), is given in Table 2.1. There is not a simple voiced/voiceless distinction; instead, plain voiceless stops and affricates contrast with pre-nasalized stops and affricates (c.f., Iverson and Salmons, 1996). In the consonant table, pre-nasalized stops and affricates are listed to the right of non-pre-nasalized stops and affricates with the same place and manner of articulation.[1] Post-palatalization and post-labialization are contrastive on stops, but not on affricates–though both [ts] and [tsʲ] are used in SMPM, they are allophones of the same phoneme, with [ts] occuring before [i] and [tsʲ] occurring before all other vowels (Stremel, 2022). Post-palatalized and post-labialized stops are contrastive, though, and as a result separate columns are given for post-palatalized and post-labialized stops in Table 2.1. The prenasalized affricates are represented with a

---

[1] There is no pre-nasalized [k]. The only case of this I have encountered is *ĩ*ⁿkà, which is likely a fossilized combination of the numeral *ĩĩ* ('one') and the demonstrative *káa* ('that').

superscripted [n], but the actual affricate is still represented as voiceless (e.g., [^{n}ts] instead of [^{n}dz]) because the affricate portion of pre-nasalized affricates is usually, but not always, phonetically voiceless. This is different from pre-nasalized stops, which are almost always phonetically voiced.

	Bilabial	Alveolar	Palatalized alveolar	Palatal/ palato-alveolar	Velar	Labio-velar	Palatalized velar
Stop	p/^{m}b	t/^{n}d	t^{j}/^{n}d^{j}		k	k^{w}	k^{j}
Fricative		s		ʃ			
Affricate		ts$^{(j)}$/^{n}ts$^{(j)}$		tʃ/^{n}tʃ			
Lateral		l					
Nasal	m	n	ɲ				
Approximant	β			j			

Table 2.1: Consonants in SMPM

Pre-nasalization is listed as a feature of the following consonant in Table 2.1, contra Peters (2018:12). The reason for this is that there is a general ban on consonant clusters in SMPM, with pre-nasalized consonants like [^{n}d] presenting the only potential exception.[2] Additionally, pre-nasalized obstruents can occur root-initially, violating the sonority sequencing principle when they do. For example, in the word [^{n}dò^{ʔ}o] (‘adobe’), the period of nasalization is of higher sonority than the following stop. However, the opposite configuration (for example, a hypothetical root like *[sno’o]) is unattested. Because it is unlikely that a language would permit a cluster that violates sonority

---

[2] There is evidence against treating post-palatalization and post-labialization as involving consonant clusters: While onset [j] is allowed before high front vowels (e.g., [jìβì] ‘person/people’), I have not encountered any post-palatalized consonants before the high front vowel (e.g., the hypothetical *[k^{j}ìβì]). This distributional restriction is not expected if both involve the consonant [j]. Additionally, post-labialized stops cannot represent a cluster, since [w] is not a consonant in and of itself—there are no cases of simple onset [w].

sequencing but ban a cluster with the same component parts that obeys it, I analyze pre-nasalized stops and affricates not as consonant clusters but rather as internally-complex segments.

In addition to its consonantal inventory, SMPM also contrasts five oral vowels and three nasal vowels,[3] which are a subset a of the oral vowels, as shown in Table 2.2. I know of no phonological evidence for the inclusion or exclusion of [a] from the set of back vowels, but I tentatively list it here as a mid low vowel.

	Oral vowels		
	Front	Mid	Back
High	i		u
Mid	e		o
Low		a	

	Nasal vowels		
	Front	Mid	Back
High	ĩ		ũ
Mid			
Low		ã	

Table 2.2: Oral and nasal vowels in SMPM

SMPM's relatively simple vowel system is accompanied by a rich system of tonal contrasts. As noted in Peters (2018), there are five phonemic tones in SMPM. There are three level tones, namely high (marked with an acute diacritic [v́]), mid (no diacritic [v]), and low (marked with a grave accent [v̀]), and at least two contour tones, namely a low-to-high rising tone (marked with a hacek [v̌]) and multiple falling tones (marked with a circumflex accent [v̂]). Though falling tones may start and end at different tonal levels (i.e., high-to-mid, mid-to-low, and high-to-low), it is unlikely that these different falling tones are contrastive. Instead, at least some of them are likely

---

[3] Phonetically, [ũ] is sometimes realized as [õ], but the distinction is non-contrastive. I do not know if the distinction is environmentally conditioned in any way.

allophones of level tones. Because of this, falling tones in Table 2.3 are represented as F instead of a specific contour, unlike LH rises. Any one of SMPM's five tones may be hosted on a single, mono-moraic short vowel. The tones may combine relatively freely, though rising and falling tones are more restricted: Rises occur most often root-finally, and I know of no roots with a final underived falling tone, though word-final falls can be derived through pronominal enclisis (Peters and Mendoza, 2020). The list of possible tonal melodies on CVCV roots is listed in Table 2.3.

	H	M	L	LH
H	léló 'skunk'	tʃú^{h}tu 'cat'	tá^{h}tà 'father'	—
M	ijá 'sour'	le^{h}so 'rabbit'	ja^{h}k^{w}à 'dirty'	i^{h}kĩ̌ 'gourd/pumpkin'
L	tʃĩ̀^{h}tʃí 'avocado'	^{n}dà^{h}ʃĩ 'wet'	tsìnà 'dog'	sà^{h}tǎ 'back'
LH	—	—	ʃĩ̌lì 'woodpecker'	—
F	—	tʃêle 'rooster'	^{n}tsîkà 'wide'	îtũ̌ 'tree'

Table 2.3: Tonal melodies on CVCV roots

The tonal melodies on mono-syllabic, long vowel roots are somewhat more restricted, as can be seen in Table 2.4. For example, I know of no roots of this shape with a final LH rising tone. It is also worth noting that only CVV roots may host two phonemic tones on one vowel, since the long vowel is made up of two moras. This can be seen in that CVV roots may host contours made up of three pitch targets (i.e., LH-L in (1)), which are best thought of as tonal melodies, with each tone associated to an individual mora.

	H	M	L	LH
H	ĩ-ⁿtsʲéé-ni ‘fragile’	ⁿdúu ‘black bug’	kʷã́ã̀ ‘yellow’	—
M	kʷií ‘hardworking’	ĩĩ ‘one’	ñũũ̀ ‘town’	—
L	tsĩ̀ĩ́ ‘rat’	ⁿtsìi ‘sad’	ʃàà ‘chin’	—
LH	—	—	lěè ‘baby’	—

Table 2.4: tonal melodies on CVV roots

Aside from marking lexical distinctions, tone is involved in expressing aspect and negation (Ostrove, 2018; Eischens, 2020). In general, Continuative aspect is marked by a High tone (1), Potential aspect is marked by a Mid tone (2),[4] Completive aspect is marked by a Low tone (3), and negation in the Potential aspect is marked by a Low-High Rising tone (4)

(1) ⁿdáʰʧí saà
fly.CONT bird
‘The bird flies.’

(2) ⁿdaʰʧí saà
fly.POT bird
‘The bird will fly.’

(3) ⁿdàʰʧí saà
fly.COMPL bird
‘The bird flew.’

(4) ⁿdǎʰʧí saà
NEG.fly.POT bird
‘The bird won’t fly.’

In addition to its consonantal, vocalic, and tonal inventory, another defining characteristics of SMPM’s phonological systems is its syllable and root structure, as well as

---

[4] Potential aspect is more accurately characterized as the absence of grammatical tone on the initial vowel, so the tone seen in this example is actually the lexical tone of the first vowel of the verb.

its use of contrastive non-modal phonation. Because the analyses of root shape and non-modal phonation depend on each other, I walk through them together below.

2.5 Root shape and laryngealization

Syllable and root templates across Mixtec languages are relatively uniform: In general, coda consonants and consonant clusters are disallowed, so the syllable template is (C)V. Additionally, lexical roots must meet a bi-moraic minimal word template (See Penner, 2019 for a comprehensive overview), and many Mixtec languages also show a preference for roots to be maximally bi-moraic (e.g., Alcozauca Mixtec; Uchihara and Mendoza Ruiz, 2021). The conglomeration of these properties—a ban on codas, a bi-moraic minimal word requirement, and a preference for maximally bi-moraic roots—results in the canonical root shapes CVCV, CVV, VCV, and VV, where 'V' represents a mono-moraic short vowel, and 'VV' represents a bi-moraic long vowel.

Another shared feature across Mixtec languages is that laryngealization, which is usually transcribed as a glottal stop [ʔ], patterns differently from other consonants. Laryngealization is restricted to root-medial positions in most Mixtec languages,[5] there may only be one per root, and it seems to act as the only licit coda. The result of these characteristics is that laryngealized roots in Mixtec languages tend to be of the shapes $\mathrm{CV^{ʔ}CV}$, $\mathrm{CV^{ʔ}V}$, $\mathrm{V^{ʔ}CV}$, or $\mathrm{V^{ʔ}V}$. The general shape of roots with and without [ʔ] are shown in Table 2.5 below.

---

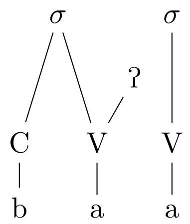
[5] Root-final glottalization is found in Ayutla Mixtec (Pankratz and Pike, 1967) and Zacatepec Mixtec (Towne, 2011)

Modal	CVCV	CVV	VCV	VV
Laryngealized	CV^{ʔ}CV	CV^{ʔ}V	V^{ʔ}CV	V^{ʔ}V

Table 2.5: General modal and laryngealized root shapes in Mixtec languages

Yet another characteristic that differentiates [ʔ] from other consonants is that, in some Mixtec languages, it is transparent for the purposes of tone sandhi. That is, words of the shape CVV and CV^{ʔ}V pattern as a natural class with regard to tone sandhi, to the exclusion of CVCV roots (Macaulay and Salmons 1995:58). Finally, the vowels on each side of the [ʔ] in CV^{ʔ}V words always match in vowel quality and nasality, while vowels in CVCV words do not always match. Because of these considerations, many researchers have advanced the hypothesis that laryngealization in Mixtec languages does not represent a consonant proper, instead positing that it is a supra-segmental feature of either vowels/moras or roots (Macaulay and Salmons, 1995; Gerfen, 2013; McKendry, 2013; Becerra Roldán, 2015; Mendoza Ruiz, 2016; León Vásquez, 2017; Penner, 2019, a.o.). For illustration, the representation of [βa^{ʔ}a] (‘good’) in Chalcatongo Mixtec is reproduced below from Macaulay and Salmons (1995).

(5)



The facts in SMPM basically mirror the broader Mixtec trend. First, there is a bi-moraic minimal word requirement, as the only words that are mono-moraic are functional items like weak pronouns and verbal morphology, which are likely prosodically dependent. This fact, coupled with the usual strong ban on coda consonants, results in the canonical root shapes in SMPM being those shown in Table 2.6:[6]

CVCV	CVV	VCV	VV
léló ‘skunk’	lěè ‘baby’	ámá ‘when?’	ĩ̃ĩ ‘one’

Table 2.6: Modal root shapes in SMPM

Like in other Mixtec languages, contrastive laryngealization in SMPM surfaces only once per root, appearing root-medially, either before a voiced consonant or a vowel. It never occurs root-finally or root-initially.[7] When laryngealization occurs in between two vowels in a mono-morphemic word, the two vowels obligatorily match in quality and nasality. This results in the laryngealized root shapes in Table 2.7.

CVˀCV	CVˀV	VˀCV	VˀV
sìˀβà ‘seed’	kʷìˀi ‘fruit’	iˀní ‘hot’	íˀi ‘raw’

Table 2.7: Laryngealized root shapes in SMPM

---

[6] Mono-morphemic words of the shape CVCVV and CVCVCV are rare, and most examples contain what is likely a fossilized noun-class prefix [tsi-/ʧi-] associated with the merger of the animal and historical round noun classes (Peters, 2018).

[7] Like in other varieties of Mixtec (i.e., San Pedro Tulixtlahuaca; Becerra-Roldán, 2019:112), [ʔ] sometimes occurs epenthetically at the beginning of vowel-initial roots, but in SMPM this only happens to resolve vowel hiatus. As a result, I do not analyze these cases of laryngealization as underlying. Additionally, utterance-final (and therefore root-final) [ʔ] is a marker of polar questions in SMPM (Eischens, to appear), but this is a prosodic feature and not a lexical feature.

Because of the phonotactic restrictions on laryngealization in SMPM, I follow the Mixtec literature in analyzing [ʔ] as a suprasegmental feature associated to a vowel, not as a consonant proper. Throughout the dissertation, I follow Penner (2019) in analyzing CVCV and CV^{ʔ}CV roots as bi-syllabic and CVV roots as monosyllabic. However, given the debate about the syllabification of CV^{ʔ}V roots (Penner, 2019:87) and the lack of conclusive evidence for their syllabic status in SMPM, I do not rely on syllables in the analysis of CV^{ʔ}V roots in Chapter 3.

Another important characteristic of laryngealization in SMPM is its interaction with tone. SMPM appears to be a laryngeally-complex language in the sense of Silverman (1997), as tone and phonation are fully cross-classified. That is, it is not the case that laryngealized vowels may host only a subset of the tones hosted on modal vowels, or vice versa. This can be seen in Table 2.8, which shows possible tonal melodies on laryngealized roots. As can be seen below, the range of attested tonal melodies on laryngealized roots is roughly identical to the range of possible melodies on modal roots.[8]

---

[8] It should be noted that the distinction between H-L and F-L is not clear—laryngealization causes pitch lowering, which makes it difficult to tell the difference between a phonological falling tone and a phonological high tone that falls due to coarticulation with laryngealization. The potential for coarticulation between an H and a following L exacerbates the problem. The same problems exist to some extent for the difference between H-M and F-M.

	H	M	L	LH
H	tsʲéˀé ‘hard’	jéˀe ‘door’	jéˀè ‘bright’	—
M	iˀní ‘hot’	βeˀe ‘house’	nãˀã̀ ‘early’	jaˀǎ ‘chile’
L	tsʲòˀó ‘flea’	tsʲòˀo ‘root’	sìˀβà ‘seed’	ʃìˀǐ ‘mushroom’
LH	—	—	mã̌ˀnã̀ ‘sleepless’	—
F	—	—	kũ̂ˀũ̀ ‘sick’	ñũ̂ˀũ̌ ‘dirt’

Table 2.8: Tonal melodies on laryngealized roots

Another point of interest in the phonological inventory of SMPM is the distribution of [h], detailed below.

2.6 Breathiness

SMPM has developed a laryngeal distinction between [ʔ] and [h], which is rare in Mixtec languages and is potentially an innovation (Peters, 2018). Contrastive [h] has the same phonotactic distribution as [ʔ], occurring only root-medially before voiced consonants or vowels (Table 2.9). Because they share a phonotactic distribution, it is likely that [h], too, is the realization of a pattern of non-modal phonation. However, unlike [ʔ], [h] does not occur root-initially even to resolve hiatus.[9]

---

[9] There are certain Spanish loanwords that have initial [h], such as ɾ*à hẽ́ⁿtè*, a loan from the Spanish ‘agente,’ which is a term used in municipal governments. I know of no Spanish loanwords with laryngealization.

There is also a distinction between contrastive and non-contrastive [h], similar to the distinction in Coatzospan Mixtec between contrastive and non-contrastive [ʔ] (Gerfen, 2013). In SMPM, every root-medial, voiceless consonant is preceded by [h]. It is unclear at present whether this is best analyzed as predictable breathy phonation, or as pre-aspiration of root-medial voiceless consonants. Because predictable [h] only occurs before root-medial consonants, the root shapes with predictable [h] in Table 2.9 are a subset of those that contain contrastive (non-predictable) [h].

CV^{h}CV	CV^{h}V	V^{h}CV	V^{h}V
sà^{h}βĭ	sâ^{h}ă	ĩ̀^{h}mã̌	ĩ̂^{h}ĩ̌
‘rain’	‘is angry’	‘wax’	‘skin’

CV^{h}CV	V^{h}CV
tá^{h}tà	i^{h}kĭ
‘father’	‘bone’

Table 2.9: Contrastively (left) and non-constrastively (right) breathy roots in SMPM

There are also several allophonic realizations of [h]. First, when occurring after the high front vowel [i], [h] is realized as the palatal fricative [ç], as indicated by the narrow transcription in (6). Additionally, when hosted on a nasal vowel, [h] is nasalized, as shown in the narrow transcription in (7).[10] Though I do not notate these realizations throughout the dissertation, they are consistent.

(6) k^{w}îçĭ

‘Green’

(7) tã́h̃ã

‘Earthquake’

Contrastively breathy roots have a much more restricted tonal distribution, as can be seen below in Table 2.9. As far as I have been able to tell (and as is evident in

---

[10] In SMPM, V1 is only nasalized if V2 is nasal, so this only happens when nasalization spreads from V2 to V1. It is possible for V1 to be oral while V2 is nasal, as in [βitsĩ (‘now/today’).

Table 2.3), the melodies of roots with non-contrastive [h] are much freer. Note that I have not included F as an initial tone for breathy roots because in all sequences of T1 + T2 where T2 starts with an L (i.e., H-LH, M-LH), any T1 that starts higher than it is realized as a fall. So, there appears to be no melody for which there is a clear contrast between a level tone and a falling tone (i.e., no H-M and F-M distinction).

	H	M	L	LH
H	—	tã́^{h}ã ‘earthquake’	—	nĩ́^{h}ĩ̌ ‘skinny’
M	i^{h}í ‘husband’	—	—	nĩ^{h}ĩ̌ ‘corn cob’
L	ts^{j}à^{h}á ‘man’	βè^{h}e ‘heavy’	—	nĩ̀^{h}ĩ̌ ‘blood’
LH	—	—	—	—

Table 2.10: Tonal melodies on contrastively breathy roots

Having detailed the phonological inventory of SMPM and argued for the supra-segmental status of laryngealization and breathiness, I would like to devote some more space to further motivating a supra-segmental analysis of laryngealization in SMPM.

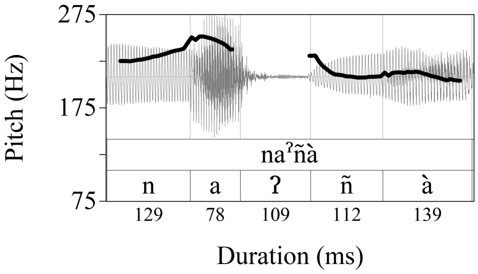
2.7 Using rate to probe supra-segmental distinctions

Cross-linguistically, the distinction between consonantal [ʔ] and creaky phonation, as well as between consonantal [h] and breathy phonation, is not always apparent from investigation of their acoustic correlates. This runs somewhat counter to the intuition that creaky vowels should be phonetically vowel-like, while glottal stops should be phonetically consonant-like. For example, if creaky voice is a part of a vowel, and vowels involve relatively unobstructed airflow cross-linguistically, then one might expect creaky vowels not to involve total occlusion of the vocal folds. On the other hand, if a consonantal [ʔ] is defined as a stop, and stops involve total blockage of airflow, one might expect a glottal stop not to involve voicing, since this would indicate incomplete or interrupted vocal fold closure. However, it is often the case that consonantal [ʔ] is produced with voicing and without total closure, and that supra-segmental [ʔ] is produced with glottal closure. For example, in an examination of 201 JIPA illustrations, Garellek et al. (2021) found that supra-segmental [ʔ/h] and consonantal [ʔ/h] displayed very similar amounts of voicing, especially in intervocalic position. They argue that the amount of voicing during a [ʔ] or [h] should not be used to argue for a consonantal or supra-segmental analysis. This point highlights that whether a laryngeal gesture has phonetic properties often associated with an obstruent (e.g., full or near-full closure, voicelessness) or a vowel (e.g., unobstructed vocal tract, voicing) is not always a reliable cue to its phonological representation as either consonantal or vocalic. So, instead of arguments from acoustic correlates of [ʔ] and [h], arguments for a segmental or supra-segmental analysis of laryngeal gestures tend to be phonological in nature, such as the phonotactic arguments for the supra-segmental nature of laryngealization in SMPM given earlier.

The lack of a consistent cross-linguistic manifestation of non-modal phonation, as well as the ambiguity between a segmental and supra-segmental analysis, can be seen in laryngealized vowels in SMPM. In (8) the initial syllable of the word [naˀñà] ('lizard') has a modally-voiced vowel, on which the pitch associated with the mid tone is realized, followed by a period of glottal closure.

(8) naˀñà

'Lizard'



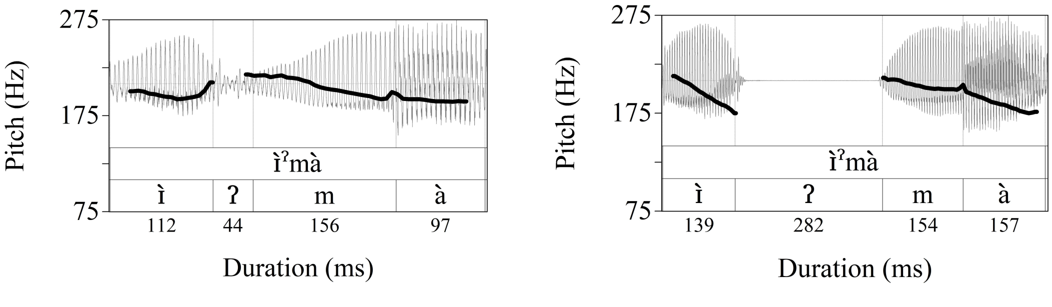
The glottal closure in (8) could, in principal, be a coda consonant of the initial syllable (a segmental [ʔ]), or it could be the realization of laryngealized phonation (a vowel Vˀ), with its occurrence after the modal portion of the vowel being due to the fact that the same vowel needs to convey both tone and phonation contrasts, so the two are phased relative to each other in order to provide adequate auditory cues for both tone and phonation (Silverman, 1997). The fact that [ʔ] is realized as glottal closure in this example does not by itself determine which of the two possible hypotheses is correct. Instead, evidence from phonotactic patterns has been used to argue that the glottal closure in (8) is the realization of non-modal phonation phased relative to the modal portion of the vowel (§§5-6). This evidence is found in other Mixtec languages as well, and has largely led the Mixtecanist literature to analyze [ʔ] as the realization of non-modal phonation.

In this section, I provide some more preliminary evidence in support of the supra-segmental analysis using a different line argumentation than the usual phonotactic considerations. This additional evidence is important for several reasons. First, much of the argumentation and analysis in this dissertation hinges on the supra-segmental nature of laryngealization. For example, generalizations about the application of low tone spread in Chapter 2 necessarily reference laryngealized vowels, and the phonological analysis of mora deletion in Chapters 3 and 4 also hinges on laryngealization's vocalic association. Because a supra-segmental analysis of laryngealization in SMPM is extremely important to the argumentation throughout, it is worth shoring up this type of analysis. The second reason this type of convergent evidence is important is because it has the potential to be used in at least some other languages (most likely other laryngeally-complex languages) to help differentiate between a segmental or supra-segmental analysis of laryngeal contrasts. This is important because, as discussed earlier, this distinction is often not clear-cut, and any piece of evidence that can be used to distinguish between the two possibilities is welcome. So, in order to provide more evidence for a supra-segmental analysis of laryngealization (and, by extension, breathiness) in SMPM, and to introduce a novel technique for probing the distinction between the two potential representations of laryngeals in other languages, this section presents preliminary evidence from a speech rate manipulation task with Consultant 1 that suggests that laryngealization in SMPM is analyzed by speakers as a part of the vowel, and not as a consonant in and of itself.

In SMPM, as mentioned above, laryngealized vowels are phased as modal-then-laryngealized. As speech rate slows down and segments lengthen, the modal and laryngealized portions of the vowel do not lengthen to an equal extent: The period of laryngealization increases in duration much more than the period of modal voicing. Take, for example, the initial vowel of the word [ìˀmã̀] (‘smoke’) in (9). In a production at a normal rate of speech, shown on the left, the modal portion of the vowel lasts 112 ms, and the laryngealized portion of the vowel, realized as creaky voice, lasts 44 ms. In a production in very slow speech, the modal portion of the vowel lasts 139 ms, representing an increase of 27 ms (24%) from normal speech. However, the period of laryngealization, realized as glottal closure, lasts 282 ms in very slow speech, representing an increase of 238 ms (541%). This difference can also be stated in terms of proportions: 72% of the modal + nonmodal sequence is occupied by the modal vowel in normal speech, but only 33% of that same sequence is occupied by the modal vowel in slow speech. In other words, the proportion of the vowel that is modal decreases as speech rate decreases.

(9) ìˀmã̀

‘Smoke’

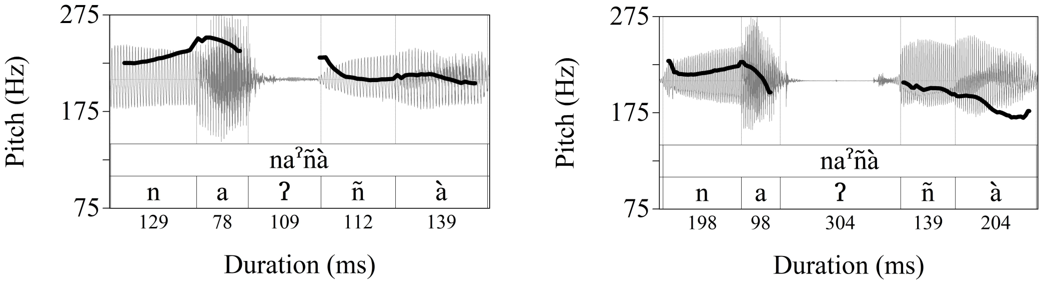


Now, this example is somewhat more drastic than most, but it is by no means misleading.

A similar pattern can be seen for normal and very slow productions of the word [na^{ʔ}ñà] (‘lizard’) in (10). The modal portion of the first vowel increases in duration by 20 ms (26%) in slow speech relative to normal speech, while the laryngealized portion increases by 195 ms (179%). In terms of the ratio of modal to nonmodal voicing, 69% of the vowel is modal in normal speech, but only 24% of the vowel is modal in very slow speech.

(10) na^{ʔ}ñà

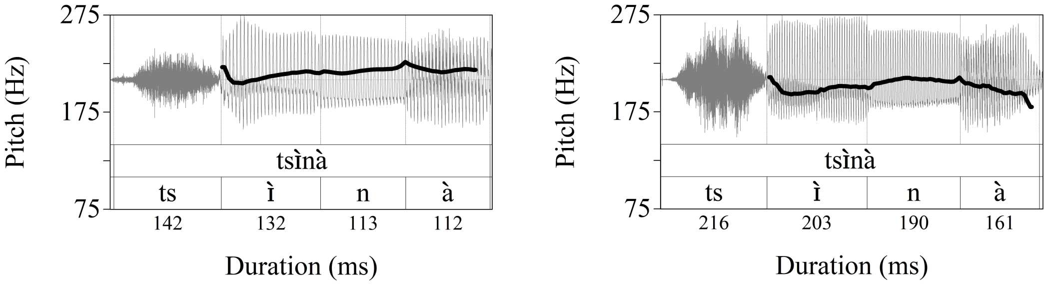
‘Lizard’



Crucially, it is not the case that modal vowels simply do not lengthen much in slow speech. This can be seen in the following two pairs of productions which involve a non-laryngealized, modal vowel in the first syllable of the word. In (11), the first vowel increases by 71 ms (54%) in slow speech, and in (12) the first vowel increases by 79 ms (68%) from normal to slow speech.

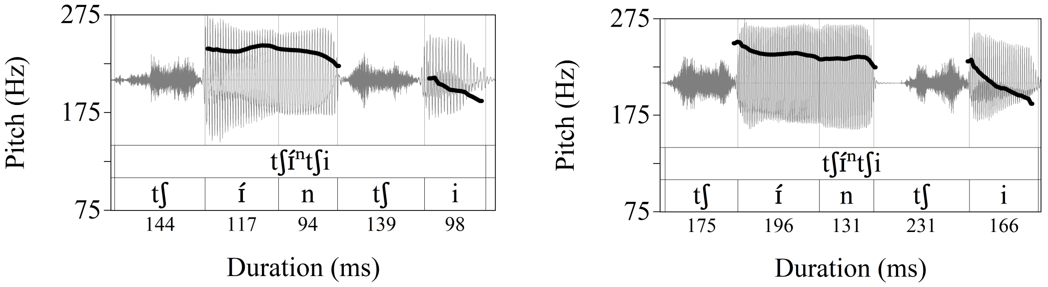
(11) tsìnà

‘Dog’



(12) tʃí^{n}tʃi

‘Cricket.’



Importantly, the example in (12) is similar to laryngealized vowels in that it involves a sequence of a modal vowel and a portion of the signal between the modal vowel and following consonant (in this case, a period of pre-nasalization). This sequence is similar to the modal-nonmodal sequencing of laryngealized vowels, but distinct phonologically in that pre-nasalization is property of the following consonant. For the purposes of comparison with laryngealized vowels, this example can also be stated in terms of proportions: 55% of the vowel + pre-nasalization sequence is occupied by the modal vowel in normal speech, and 60% of that same sequence is occupied by the modal vowel in slow speech. In other words, modal vowels do not decrease in proportion to the following portion of the signal in slow speech.

There appears to be a difference, then, between the lengthening patterns of laryngealized vowels and modal vowels—in slow speech, modal vowels seem to lengthen by a larger proportion than does the modal portion of a laryngealized vowel. There are, in principal, two ways to interpret this: Under a consonantal analysis of [ʔ], the first syllable of [na^{ʔ}ñà] (‘lizard’) has an onset, a nucleus, and a coda ([ʔ]). The disproportionate lengthening of the [ʔ] might represent a general lengthening strategy in slow speech whereby codas lengthen more than the nucleus of a syllable. In words like [tsìnà] (‘dog’), there is no coda consonant, and because of this the nucleus of the initial syllable lengthens in slow speech. Under a supra-segmental analysis of laryngealization, the laryngealized vowel in [na^{ʔ}ñà] (‘lizard’) has an onset and a nucleus (V^{ʔ}), but not a coda. In slow speech, the laryngealized vowel lengthens, and it might be a general lengthening strategy in slow speech to lengthen the non-modal portion of the vowel more than the modal portion of the vowel. In a word like [tsìnà] (‘dog’), the same type of lengthening of the vowel of the first syllable happens, but because it is a modal vowel, what lengthens is, by definition, the modal vowel.

The segmental and supra-segmental hypotheses are both able to account for the disproportionate lengthening of laryngealization in slow speech, but they make distinct predictions about what should happen in laryngealized roots without a medial consonant (CV^{ʔ}V). In a root of this shape, if the [ʔ] is a consonant, then it should syllabify as the onset of the second syllable.[11] If the lengthening strategy of an open syllable is to lengthen the nucleus, then the vowel of the first syllable should lengthen more in CV^{ʔ}V roots than it does in CV^{ʔ}CV roots, since in a CV^{ʔ}V root the first vowel is in an open syllable. The supra-segmental analysis, on the other hand, states that the disproportionate lengthening of [ʔ] in slow speech is a general lengthening strategy for laryngealized vowels. Because of this, no qualitative difference is predicted between CV^{ʔ}CV and CV^{ʔ}V roots—because both involve laryngealized vowels, both should be subject to the same lengthening strategy.

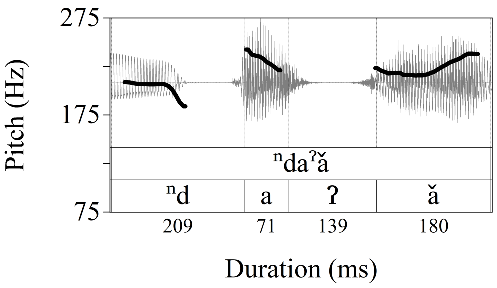
It appears that the predictions of the segmental hypothesis are incorrect, and that the predictions of the supra-segmental hypothesis are correct. Take, for example the following productions of the word ^{n}*da*^{ʔ}*ǎ* (‘hand’). In the normal-rate production on the left, the initial modal vowel lasts 71 ms, while laryngealization lasts 139 ms. In the very slow production, the length of the modal vowel is actually shorter (52 ms, a 13% decrease), while laryngealization is, as usual, significantly longer (252 ms, a 81% increase). Stated again in terms of proportions, the proportion of the V^{ʔ} sequence that is modal is 34% in normal speech and 17% in slow speech.

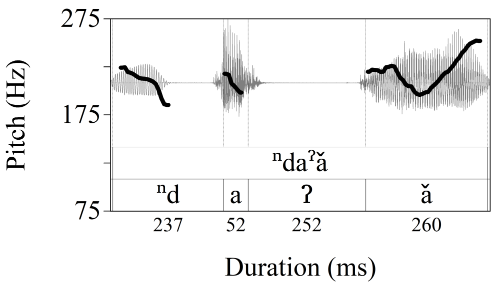
---

[11] One might argue that it syllabifies exceptionally as a coda, but there is no empirical evidence for this in SMPM. It is also theoretically undesireable, since it would treat CV^{ʔ}V roots differently from CVCV roots, which are unambiguously syllabified [CV.CV].

(13) ⁿdaˀă

‘Hand.’





So, the period of modal voicing preceding laryngealization does not lengthen very much (or, in the previous case, at all), whether the [ʔ] is followed by another consonant or not. This is inconsistent with the predictions of the consonantal analysis, but consistent with the a supra-segmental analysis, which holds that the laryngealized portion of a laryngealized vowel increases more in slow speech than the modal portion. Crucially, this should not be affected (or at least, not greatly affected) by whether or not there is a consonant following laryngealization. In order to test the robustness of this finding, and also to examine whether breathy vowels (both contrastive and non-contrastive) pattern in a similar way, I carried out a production task with Consultant 1. The goal of this task was to investigate how speech rate affects the relative durations of the modal and non-modal portions of laryngealized and breathy vowels in comparison to modal vowels.

2.7.1 Roots with a medial C: Methods

In this task, Consultant 1 produced target words in the carrier phrase in (14) at three different rates of speech. They were asked to produce the first utterance at a normal rate, the second more slowly, and the third very slowly.

(14) kã̂ˀ=ĩ̀ ___ βiʰtsĩ
POT.say=1SG ___ now
'I will say ___ now.'

The target word varied in whether the initial vowel was laryngealized, breathy, pre-aspirated,[12] modal, or pre-nasalized. Laryngealized and breathy roots varied in whether they included a medial consonant or not. All elicited root types are shown in Table 2.11. Roots with a medial, pre-nasalized consonant were included for purposes of comparison because pre-nasalization is unambiguously a property of the following consonant, and because pre-nasalized roots do not involve disproportionate lengthening in slow speech of the portion of the signal between the initial modal vowel and the following consonant (in this case, the period of pre-nasalization as in (12)).

Consultant 1 produced this three-sentence sequence for 41 laryngealized target words (21 with medial C, 20 without), 40 breathy roots (20 with medial C, 20 without), 21 modal roots, 28 preaspirated roots, and 21 pre-nasalized roots. This means that they produced 151 target words (41+40+21+28+21) at three different speech rates for a total of 453 productions. There were 38 distinct lexical items for laryngealized roots (3 used twice),

---

[12] Recall that all root-medial voiceless consonants are predictably preceded by [h]. To clearly distinguish this [h] from contrastive breathiness, which unpredictably precedes voiced consonants or vowels, I refer to it here as pre-aspiration, though the results of this task cast some doubt on this characterization.

	Example	
Root type	**With medial C**	**Without medial C**
Laryngealized	na^{ʔ}ñà ‘Lizard’	^{n}dò^{ʔ}o ‘Adobe’
Breathy	nũ̀^{h}nĩ̌ ‘Corn’	nĩ̀^{h}ĩ̌ ‘Blood’
Pre-aspirated	tsì^{h}kà ‘Grasshopper’	
Modal	tsìnà ‘Dog’	
Pre-nasalized	tʃí^{n}tʃi ‘Cricket’	

Table 2.11: Target root types

33 distinct items for breathy roots (7 used twice), 28 distinct pre-aspirated items (none used twice), and 21 distinct modal items (none used twice). There were only 5 distinct items for pre-nasalizated roots (all used more than once) because pre-nasalization of root-medial consonants is rather uncommon in SMPM.

For the first comparison, only roots with a medial consonant were analyzed, since the goal of the first analysis is to verify that the modal portion of the first vowel in CV^{ʔ}CV and CV^{h}CV roots lengthens less than the nonmodal portion in slow speech. CV^{n}CV roots were used as a baseline for comparison because, as noted before, they do not appear to involve the same pattern of disproporionate lengthening. Because the comparison was crucially between CV$^{ʔ/h}$CV roots and CV^{n}CV roots, modal roots (CVCV) and laryngealized and breathy roots with no medial consonant (CV^{ʔ}V and CV^{h}V) were excluded from the first analysis. This subsetting left 269 analyzable tokens

(21 laryngealized items, 20 breathy items, 28 preaspirated items, and 21 pre-nasalized items, each produced at three different rates of speech).

Each token was spliced and annotated in Praat, and a Praat script extracted the total duration of each token, plus the durations of each segment. The relevant portions of the tokens whose duration was measured, as well as the labels used, is given for each phonation type below in Figure 2.12. For each token, the ratio of the duration of V1 to the duration of V1 + Pre-C2 was measured. For laryngealized and breathy roots, this constituted the proportion of the V$^{ʔ/h}$ sequence taken up by the modal vowel. For pre-nasalized roots, this constituted the proportion of the the V^{n} sequence taken up by the modal vowel.

		V1 + Pre-C2			
	C1	**V1**	**Pre-C2**	**C2**	**V2**
Laryngealized na^{ʔ}ñà ‘Lizard’	n	a	ʔ	ñ	à
Breathy nũ̀^{h}nĩ̌ ‘Corn’	n	ũ̀	h	n	ĩ̌
Pre-aspirated tsì^{h}kà ‘Grasshopper’	ts	ì	h	k	à
Pre-nasalized tʃí^{n}tʃi ‘Cricket’	tʃ	í	n	tʃ	i

Table 2.12: Relevant portions of target words for Comparison 1

Two-tailed t-tests were run on the ratio of V1 to the V1 + Pre-C2 sequence for pre-nasalized roots across prompted speech rate to verify that this measure was not significantly affected by speech rate for these roots. This was done to ensure that pre-nasalized roots could be used as a baseline of comparison for the effect of speech rate on the laryngealized, breathy, and pre-aspirated roots. The result of the comparison between the normal and slower productions ($t = -1.45$, $p = 0.15$), between the slower and slowest productions ($t = 1.67$, $p = 0.1$), and between the normal and slowest productions ($t = 0.8$, $p = 0.93$) all did not reach significance. This means that the null hypothesis that the relevant measurements across each distribution come from the same underlying distribution cannot be rejected. Because the ratio of V1 to V1 + Pre-C2 did not reliably vary across prompted speech rates, the pre-nasalized condition was used as the baseline condition to which the other phonation types were compared.

In order to obtain a more accurate measure of speech rate than prompted rate, a continuous measure of speech rate was obtained by calculating speed as moras per second for each token.[13] Because each target word was bi-moraic, the total duration of each token could be taken as the amount of time it took to produce two moras. As a result, the measure of moras per second was directly derivable from the total duration of each token. The resulting values are shown in Figure 2.1 below.

The dependent variable analyzed was the ratio of V1 to V1 + Pre-C2. This was used instead of a raw value by V1 duration because the measure of interest is how much the modal vowel lengthens in proportion to what follows. However, a linear regression assumes unbounded variables, and the ratios I analyze are necessarily bounded by 0 and 1. To remedy this, I centered and scaled each token's value for the ratio of V1 to V1

---

[13] The measure of moras per second is used instead of syllables per second because of the ambiguity between a mono-syllabic or bi-syllabic analysis of $CV^{ʔ}V$ roots.

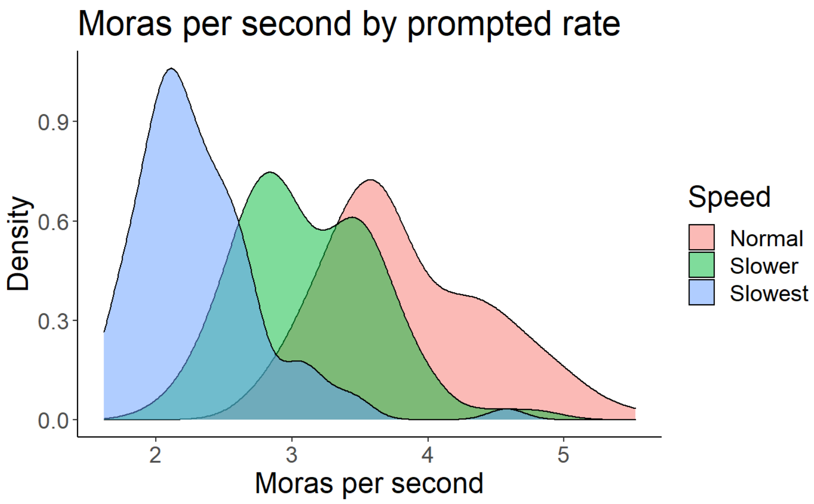


Figure 2.1: Distribution of moras per second across prompted speech rates for roots used in comparison 1.

+ Pre-C2 by taking each data point, subtracting the mean ratio of V1 to V1 + Pre-C2 from it, and dividing the resulting value by the standard deviation of the ratio of V1 to V1 + Pre-C2. Additionally, the measure of moras per second (MPS) is a variable of interest, but it is bounded at 0 (no actual speech can have a negative value for MPS). To correct for this, I center and scaled each data point's MPS measure using the same method. This resulted in data that satisfies the assumption of unboundedness in a linear regression model, since the data in z-space is technically unbounded on either side.

After these transformations, a linear mixed effects model was run on the dataset in R (R Core Team, 2013) using the lme4() package (Bates et al., 2015). The dependent variable was the centered and scaled value for each token's ratio of V1 to V1 + Pre-C2, the independent variables were speech rate (centered and scaled MPS), phonation type, and their interaction, and 'Item' was set as a random effect. Data were illustrated using ggplot (Wickham, 2016) in R.

2.7.2 Roots with a medial C: Results

The residuals of the model were normally distributed (R = 0.989). Model criticism was carried out using the drop1() function in lmerTest package (Kuznetsova et al., 2017). The full model was compared with a simpler model omitting the Phonation by Speech Rate interaction. This comparison came out as significant ($p < 0.001$) using Satterthewaite's method, indicating that this interaction should not be excluded. No further simplification of the model was possible. The results of the analysis are given in Table 2.13.

Predictor	β	**SE(β)**	**|t|**	**p-value**
Intercept	0.68	0.21	3.28	< 0.01
Moras per second (MPS)	-0.03	0.08	-0.39	$= 0.7$
Laryngealized	-1.53	0.24	-6.36	< 0.001
Breathy	-1.42	0.25	-5.61	< 0.001
Pre-aspirated	-0.12	0.23	-0.54	$= 0.59$
MPS * Laryngealized	0.6	0.10	5.7	< 0.001
MPS * Breathy	0.34	0.12	2.92	< 0.01
MPS * Pre-aspirated	0.38	0.1	3.8	< 0.001

Table 2.13: Results of mixed effects model

There was no main effect of mora per second (MPS; β=-0.03, p = 0.7), meaning that speech rate was not a general predictor of the ratio of V1 to V1 + Pre-C2 across the dataset. There were main effects for Laryngealized and Breathy phonation types (Laryngealized: $\beta = -1.53$, $p < 0.001$; Breathy: $\beta = -1.42$, $p < 0.001$), meaning that the ratio of V1 to V1 + Pre-C2 was lower in laryngealized and breathy roots than it was in pre-nasalized vowels, with a similar effect size for both laryngealized roots and breathy roots. There was no main effect for the Preaspirated phonation type (β = -0.12, p = 0.59), meaning that the ratio of V1 to V1 + Pre-C2 was not significantly different between preaspirated and pre-nasalized roots.

Though the main effects varied by phonation type, there were significant interactions between speech rate (MPS) and phonation type for each phonation type (MPS * Laryngealized: $\beta = 0.6$, $p < 0.001$; Breathy: $\beta = 0.34$, $p < 0.01$; Preaspirated: $\beta = 0.38$, $p < 0.001$). For all of these phonation types, this means that at fast speech rates (more moras per second), the ratio of V1 to V1 + Pre-C2 was higher, and at slower speech rates (fewer moras per second), the ratio of V1 to V1 + Pre-C2 was lower, and that these effects were significantly different from the baseline case of pre-nasalized roots. This effect was larger for laryngealized roots than breathy or pre-aspirated roots.

These results can be visualized in Figure 2.2, which shows that the V1 proportion of the V1 + Pre-C2 sequence stayed relatively constant for roots with a medial pre-nasalized consonant, but decreased with speech rate for breathy and laryngealized roots, as well as for roots with a medial pre-aspirated consonant. Additionally, the lack of a main effect for preaspirated roots can be seen in that the ‘Proportion V1’ values for preaspirated roots, though they changed across speech rates, overlapped with the ‘Proportion V1’ values for prenasalized roots. This value was consistently lower for both Laryngealized and Breathy roots.

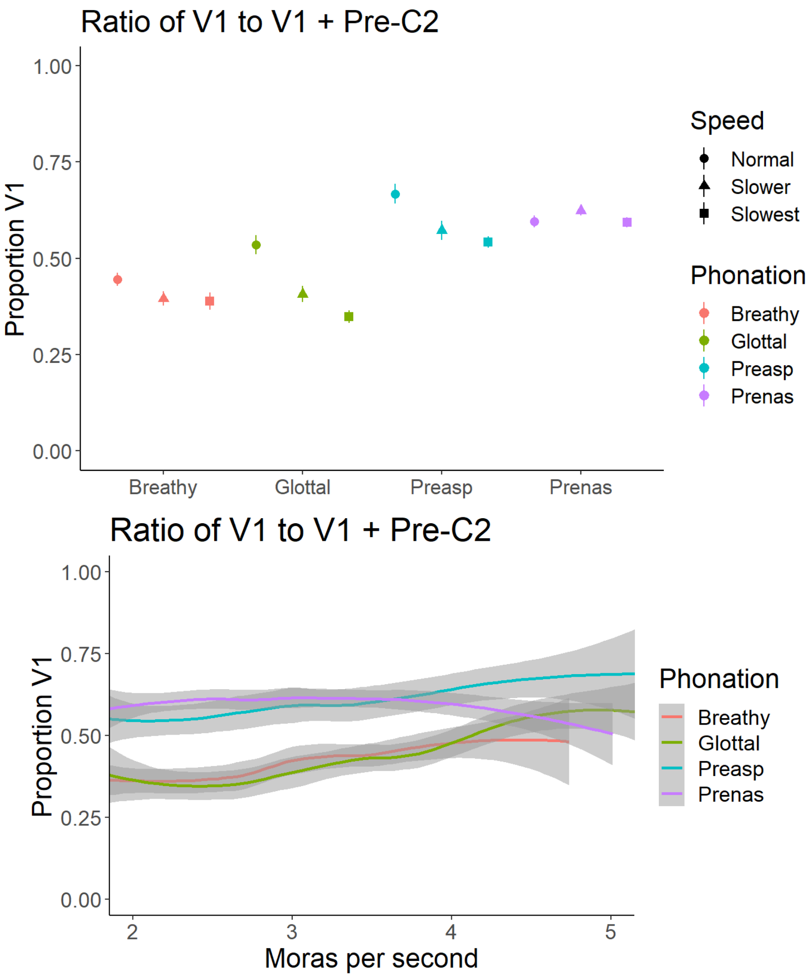


Figure 2.2: Mean ratio of V1 to V1 + Pre-C2 by prompted rate and phonation type (top) and by moras per second and phonation type (bottom)

2.7.3 Roots with no medial C: Methods

In order to examine whether the speech-rate-conditioned effect on the ratio of V1 to V1 + Pre-C2 has an analogue in laryngealized and breathy roots with no medial consonant (CV$^{\text{ʔ/h}}$V), a similar analysis was performed on these roots. Because the consonantal analysis states that the [ʔ/h] in these roots should be considered an onset consonant, this analysis compared CV$^{\text{ʔ/h}}$V roots to modal, CVCV roots, which have a modal vowel followed by an onset consonant of $\sigma 2$. If [ʔ] and [h] are considered to be C2 in this case, then this allows for a straightforward comparison of the ratio of V1 to V1 + C2.[14] This subsetting left 183 analyzable tokens (20 Laryngealized, 20 Breathy, 21 Modal, each at three rates of speech).

Once again, each token was spliced and annotated in Praat, and a Praat script extracted the total duration of each token, plus the durations of each segment. For each token, the duration of V1 to V1 + C2 was extracted. The relevant portions of the tokens for this comparison are shown in Table 2.14.

		V1 + C2		
	C1	**V1**	**C2**	**V2**
Laryngealized ⁿdaˀă ‘Hand’	ⁿd	a	ʔ	ă
Breathy nĩ̀ʰĩ̌ ‘Blood’	n	ĩ̀	h	ĩ̌
Modal tsìnà ‘Dog’	ts	ì	n	à

Table 2.14: Relevant portions of target words for Comparison 2

The measure of moras per second was once again calculated as a continuous proxy for speech rate, and the distribution of moras per second values by prompted speech rate are given in Figure 2.3 below. To satisfy the assumptions of a linear model, the measures of MPS and the ratio of V1 to V1 + C2 were centered and scaled using the same method as in the previous task.

---

[14] This was a more straightforward baseline case than pre-nasalized roots, since pre-nasalized roots have a C2 made up of two distinct parts.

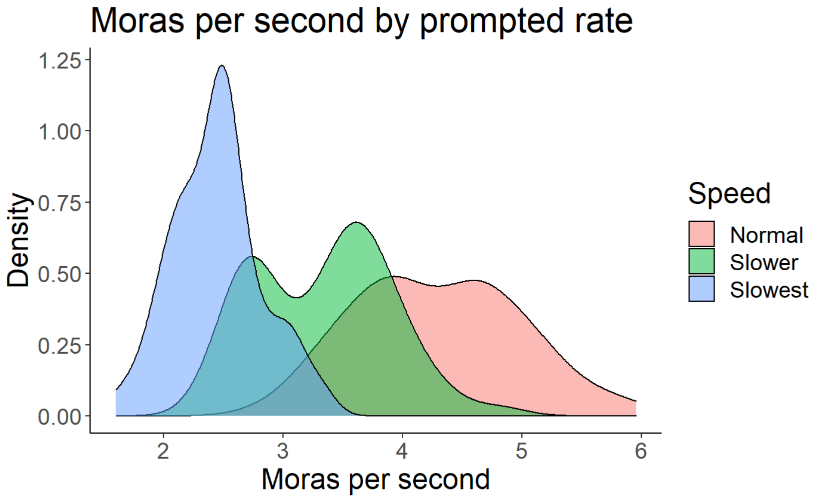


Figure 2.3: Distribution of moras per second across prompted speech rates for roots used in comparison 2.

Before running a linear regression, the ratio of V1 to V1 + C2 for modal roots across prompted speech rate was submitted to two-tailed t-tests to verify that this measure did not to vary significantly by speech rate. The t-tests comparing this measure between the normal and slower productions ($t = 0.13$, $p = 0.9$), between the slower and slowest productions ($t = 0.12$, $p = 0.91$), and between the normal and slowest productions ($t = 0.27$, $p = 0.79$) all failed to reach significance, meaning that the null hypothesis that the values for this measure came from the same underlying distribution across all three speech rates could not be rejected. After establishing the adequacy of a modal baseline, a linear regression model was run in R, using the lm() function, with the dependent variable being the centered and scaled values for the ratio of V1 to V1 + C2, and the independent variables being speech rate (centered and scaled moras per second) and phonation type.

2.7.4 Roots with no medial C: Results

As before, the residuals were normally distributed (R = 0.99). Model criticism was carried out again using the drop1() function, and the full model was compared with a simpler model omitting the Phonation by Speech Rate interaction. The comparison came out as significant ($p < 0.001$) using Satterthewaite's method, indicating that this interaction should not be excluded. No further simplification of the modal was possible. The results of the analysis are given in Table 2.15.

Predictor	β	**SE(β)**	**|t|**	**p-value**
Intercept	1.16	0.09	13.23	< 0.001
Moras per second (MPS)	0.03	0.05	0.52	$= 0.61$
Laryngealized	-1.81	0.13	-14.42	< 0.001
Breathy	-1.73	0.13	-13.8	< 0.001
MPS * Laryngealized	0.36	0.07	5.14	< 0.001
MPS * Breathy	0.23	0.07	3.23	< 0.01

Table 2.15: Results of mixed effects model

Once again, there was no main effect of MPS (β= 0.03, p = 0.61), meaning that speech rate was not a general predictor of the ratio of V1 to V1 + C2 across the dataset. There were main effects for each phonation type (Laryngealized: β = -1.81, $p < 0.001$; Breathy: β = -1.73, $p < 0.001$), meaning that the ratio of V1 to V1 + C2 was smaller in laryngealized and breathy roots than in modal roots. There were also significant interactions between speech rate (moras per second) and each phonation type (MPS * Laryngealized: β = 0.36, $p < 0.001$; MPS * Breathy: β = 0.23, $p < 0.01$). This suggests that, as speech rate decreased, so did the ratio of V1 to V1 + C2 in laryngealized and breathy roots. These effects can be visualized in Figure 2.4, which shows that the ratio of V1 to V1 + C2 is higher and does not change for modal roots by prompted speech rate, but is lower and does change for breathy and laryngealized roots by speech rate.

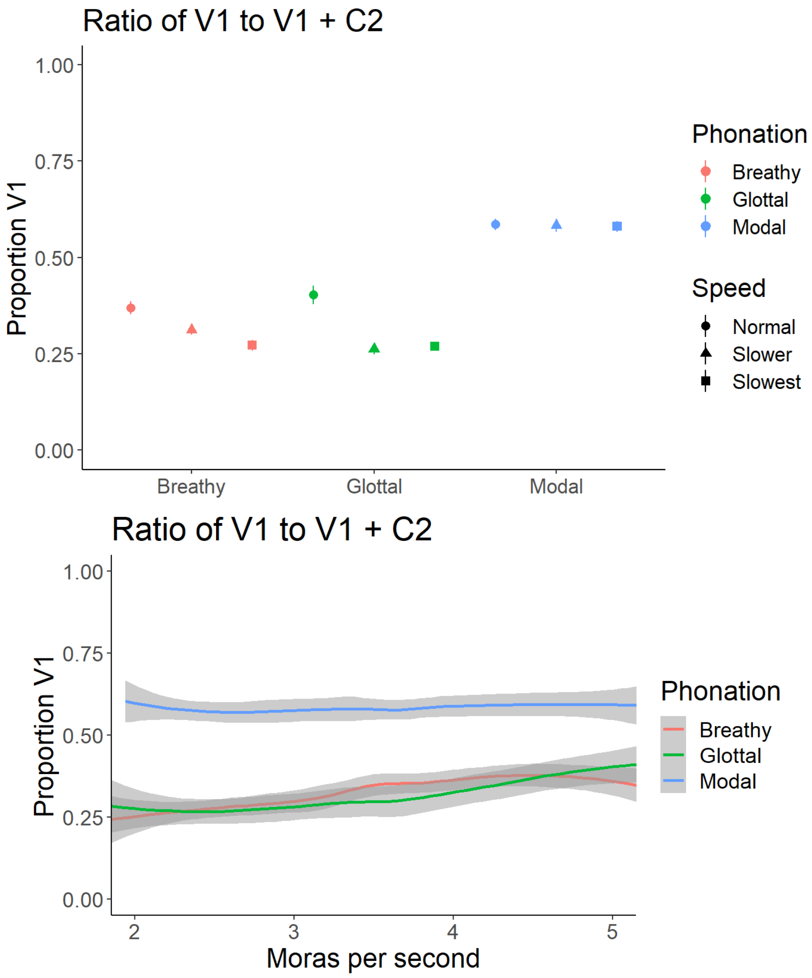


Figure 2.4: Ratio of V1 to V1 + C2 for modal, laryngealized, and breathy roots by prompted speech rate (top) and moras per second (bottom).

2.7.5 Discussion

Throughout all comparisons, the proportion of the analyzed sequence (V1 to V1 + Pre-C2, or V1 to V1 + C2, referred to hereafter as V1 to V1 + (Pre-)C2) that was made up of the modal vowel was lower in laryngealized and breathy roots relative to the baseline of pre-nasalized or modal roots, both of which involved a modal vowel not followed by nonmodal phonation. The proportion of the analyzed sequence made up of the modal vowel in pre-aspirated roots was not significantly different from pre-nasalized roots, though. This means that modal vowels are relatively shorter before [ʔ] and contrastive [h] than before pre-nasalization or medial voiced consonant, but modal vowels are *not* relatively shorter before non-contrastive [h] (that is, pre-aspiration) than they are before pre-nasalization.

Though they behave somewhat differently in their baseline ratio of V1 to V1 + (Pre-)C2, laryngealized, breathy, and pre-aspirated roots are all affected by speech rate in a way that pre-nasalized and modal roots are not. Specifically, as speech rate decreases, so does the ratio of V1 to V1 + (Pre-)C2. This is the case for laryngealized, breathy, and pre-aspirated roots with a medial consonant, as well as for laryngealized and breathy roots without a medial consonant. This result goes against the predictions of a segmental analysis of [ʔ] and [h], since this analysis predicts that the effect of speech rate should only be seen (or, should be most strongly seen) in roots with a medial consonant. The reason for this is that the segmental analysis explains the disproportionate lengthening of [ʔ/h] in $\mathrm{CV^{ʔ/h}CV}$ roots as being the result of a lengthening of the rime of $\sigma 1$, of which [ʔ/h] is the coda. In roots of the shape $\mathrm{CV}^{ʔ/\mathrm{h}}\mathrm{V}$, where [ʔ/h] would be syllabified as the onset, then a different lengthening pattern is expected. Specifically, because V1 is the only member of the rime in these cases, it should lengthen more. However, even in these cases, the proportion of the modal vowel relative to the following [ʔ/h] also decreases, suggesting that the same lengthening strategy is used for vowels followed by [ʔ/h] regardless of whether the [ʔ] or [h] would be syllabified as a coda or as an onset.[15]

Though the disproportionate lengthening of [ʔ] and [h] in slow speech is inconsistent with the predictions of the segmental analysis, it is consistent with an analysis of [ʔ] and [h] as the expression of a supra-segmental phonation contrast on the vowel. Under this hypothesis, the lengthening strategy for non-modal vowels in slow speech is to lengthen the non-modal portion of the vowel. Because of this, whether or not there is a following, root-medial consonant should have little to no effect; in both cases, the non-modal phonation should lengthen more than the modal portion of the vowel. This is the pattern seen above, and from it I conclude that the disproportionate lengthening of [ʔ] and [h] in slow speech can be taken as suggestive evidence in support of a supra-segmental analysis of these laryngeal gestures.

---

[15] It is still in principle possible that a decrease in the proportion of modal voicing in slow speech is a general slow speech strategy for vowels followed by voiceless consonants. Under this alternative hypothesis, [ʔ] and [h] would be voiceless consonants, and all voiceless consonants would lengthen disproportionately in slow speech, causing the preceding vowel to lengthen less. This hypothesis cannot be tested with the current data, since all root-medial voiceless consonants are preceded by [h], so the effect of the consonant cannot be teased apart for the effect of [h]. However, this hypothesis could be tested with morphologically-complex or fossilized tri-moraic roots of the shape CV-CVCV, in which the root (CVCV) starts with a voiceless consonant. These consonants are not preceded by [h] because they are not root-medial. If this alternative hypothesis is correct, we would expect the duration of the initial vowel (the V in CV-) in these cases to decrease relative to the duration of the following voiceless consonant as speech rate slows down.

One curious point that merits further consideration is the effect of speech rate on the ratio of V1 to V1 + C2 in laryngealized roots (where [ʔ] is classified as C2). As seen in Figure 2.4, the modal proportion of that sequence decreases from normal to slower prompted speech rate, but stays relatively consistent from the slower to slowest prompted speech rate. This lack of difference between slower and slowest prompted speech rate is also seen for the ratio of V1 to V1 + Pre-C2 for breathy roots in Figure 2.2. Though I am not sure what the cause of this lack of decrease in these conditions is, one point is clear: The lack of difference in proportion is not due to a lack of difference in speech rate. This can be seen in Figure 2.5, which shows the mean and standard error for speech rate (mora per second) for each phonation type and prompted speech rate from Comparison 2. As can be clearly seen, speech rate as measured by moras per second decreased consistently across prompted speech rates.

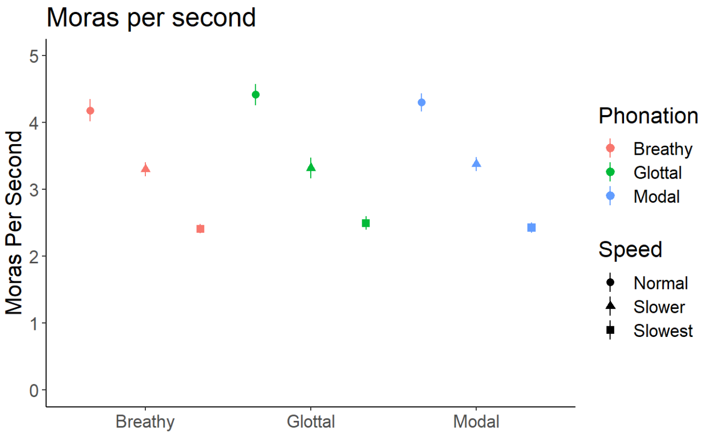


Figure 2.5: Moras per second by prompted speech rate for modal roots roots, and for laryngealized and breathy roots with no medial C.

So, though speech rate reliably slows down, the ratio of V1 to V1 + (Pre-)C2 does not lower from the slower to slowest prompted speech rate for breathy roots with a medial C and laryngealized roots with no medial C. It is possible that this lack of difference represents some kind of floor effect. For example, it might be the case that, from a normal speech rate to a slower speech rate, the ratio of V1 to V1 + (Pre-)C2 decreased to a minimal allowable amount for these configurations. Because it could go any lower, it did not decrease even as speech rate decreased further. At this point, this possibility amounts to speculation, but it is worth keeping in mind.

In any case, the results of this task clearly contradict the consonantal analysis, which predicts that coda [ʔ/h] should behave differently than onset [ʔ/h]. However, the results are perfectly consistent with the predictions of a supra-segmental analysis, which predicts no qualitative distinction between CV$^{\text{ʔ/h}}$CV and CV$^{\text{ʔ/h}}$V roots. In addition, they suggest that, since non-contrastive [h] before voiceless consonants patterns in the same way as contrastive [ʔ] and [h], it might be best analyzed not as pre-aspiration of a root-medial voiceless consonant, but rather as non-contrastive breathy phonation that is conditioned by the voicing of the following consonant. This analysis is bolstered by the fact that only root-medial voiceless consonants (not, for example, root-initial voiceless consonants) are pre-aspirated, which is consistent with the licensing of non-modal phonation only in root-medial positions.

In addition to bearing on the segmental/supra-segmental status of [ʔ] and [h] in SMPM, this study also contributes an interesting empirical observation: In slow speech in SMPM, the period of the vowel conveying tone—the modal portion—and the period of the vowel conveying the phonation contrast—the non-modal portion—do not lengthen equally in slow speech. This pattern naturally raises the questions of why the non-modal portion of the vowel is lengthened disproportionately to the modal portion of the vowel in slow speech. Though I can do no more at present than speculate about the potential reasons, it is worth considering some possibilities.

The first potential reason for this specific lengthening pattern might be tied to SMPM's status as a laryngeally-complex language in the sense of Silverman (1997). Under Silverman's proposal, languages like SMPM that independently contrast both tone and phonation on the same vowel often phase the two, since the realization of non-modal phonation can interfere with the realization of tone (for example, there is necessarily no pitch during glottal closure). In SMPM, tone is realized at the beginning of the non-modal vowel during a phase of periodic vocal fold vibration, and non-modal phonation is realized afterward. Now, if vowels are phonologically specified as non-modal but must nonetheless realize a tonal contrast, it is possible that the periodic portion of the vowel is not phonologically specified but rather a necessary phonetic byproduct of the need to realized a tonal contrast. It is possible that, because this period of periodic vocal fold vibration is not phonologically specified, it lasts only as long as is necessary to realize the tonal contrast, and no longer. So, it might be the case non-modal phonation is given 'priority' in some way, and tone is given only as much time as is necessary to realize the tonal contrast, but not more, with the reason being tied to the vowel's phonological representation in some way. Another possibility is that this lengthening effect has some functional motivation. For example, non-modal phonation might require less airflow than modal phonation, and pressure to maintain sufficient sub-glottal air pressure might be exaggerated in slow speech, leading to the selective lengthening of portions of the utterance that do not decrease sub-glottal air pressure as drastically. If it is the case that listeners do not necessarily benefit in comprehension from increased duration of the period of the vowel on which tone is realized (that is, they do not *need* more time to accurately identify the tone), then this pressure to decrease airflow in slow speech could give rise to this lengthening pattern. Another possibility about is that some auditory cues to tone in the context of a laryngealized vowel are best expressed when the period of modal voicing is relatively short, though I do not know what these cues would be. Whatever the explanation for this lengthening patter may be, it is clear that it has potential to further our understanding of the phonology and phonetics of non-modal phonation in laryngeally-complex languages.

Finally, two points are important to make: Though the high number of observations makes it likely that these results accurately convey the slow-speech patterns of Consultant 1, it is possible that these patterns represent an idiosyncratic lengthening strategy of this consultant only. Impressionistic observations of limited work with another consultant appear to show similar patterns, though, which would point toward this lengthening strategy being a characteristic shared by multiple members of the SMPM speech community. However, the conclusions of this analysis would be best supported if this lengthening pattern in slow speech holds of the general SMPM speech community. If this is the case, it will be worth exploring the effect of speech rate on the lengthening of laryngeal gestures in other languages, with the ultimate goal of determining whether speech rate effects of this kind can be broadly useful in teasing apart a segmental and supra-segmental analysis of [ʔ] and/or [h].

2.8 Conclusion

In this chapter, I have provided a brief overview of SMPM's sound system, detailing the consonantal, vocalic, and tonal contrasts, as well as a more extended discussion of root shapes and phonation types. The claim of greatest relevance for the following chapters of the dissertation is that laryngealization is best understood as a supra-segmental feature, not as a consonant proper. This claim is supported primarily by phonotactic evidence which suggests that laryngealization (and, by extension, breathiness) do not act like consonants, but rather like a part of the initial vowel of the root. Additional suggestive evidence can be taken from the results of the speech rate manipulation task detailed above, which shows that laryngealized vowels undergo a rather peculiar lengthening in slow speech regardless of whether or not there is a following consonant. With this information in hand, we may now turn to the first of several empirical investigations that have consequences for our understanding of the phonology-phonetics interface, namely the highly-specific process of low tone spread in SMPM.

Chapter 3

Differentiating phonological and phonetic levels of representation

3.1 Overview

A central question in literature on the phonology-phonetics interface is the nature of phonological units of representation: To what extent are they defined by their physical realization in time and space, and to what extent are they independent? As mentioned earlier, different phonological frameworks answer this question in different ways. For example, substance-free phonology (e.g., Reiss, 2017) defines phonological units purely as abstract features, with the phonetic grounding of phonological inventories and alternations being a result of phonetic pressures in diachronic sound change. On the other hand, a framework like Flemming's (2001) defines phonological units in terms of fine-grained phonetic detail, such that what it means to be a low vowel in a given language, for instance, is to have a specific, high F1 target. In between these two extremes are a myriad of frameworks of the interface that allow differing levels phonetic detail into phonological representations. This is schematized in Figure 3.1, which shows various frameworks of the phonology-phonetics interface along a continuum in terms of the degree of separation or integrality they propose between phonology and phonetics.[1]

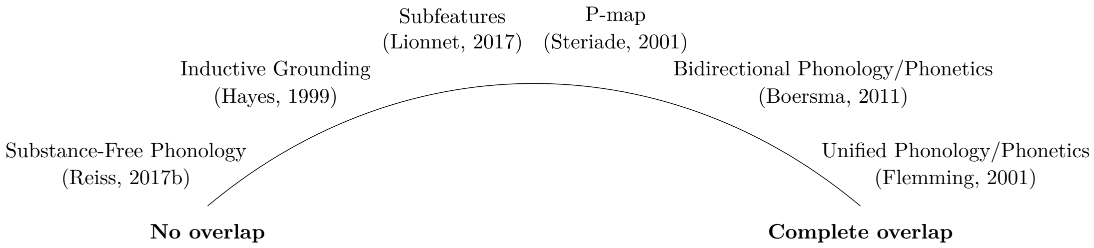


Figure 3.1: Continuum of frameworks of the phonology-phonetics interface

As mentioned earlier, the claims of this dissertation touch most directly on the two ends of this continuum, with this chapter mostly addressing frameworks on the right end of the continuum. These types of frameworks can be seen at work in the analysis of a multiply-triggered process of vowel fronting in Cantonese (Cheng, 1991). In Cantonese, the high front round vowel [y] contrasts with the high back round vowel [u] in most phonological environments. However, the two vowels do not contrast between coronal consonants. The front vowel [y] may occur before or after a coronal consonant, and also if flanked on either side by coronal consonants. The back vowel [u] may occur before or after a coronal consonant, but it never surfaces between two coronal consonants (Table 3.1).

---

[1] Again, this figure collapses a multi-dimensional space and, as a result, should not be taken as a literal claim about monotonic differences between the frameworks listed.

$\mathrm{k^hyt}$ ‘Decide’	$\mathrm{t^huk}$ ‘Bald head’
$\mathrm{t^hyt}$ ‘to take off’	$\mathrm{k^hut}$ ‘bracket’
	*$\mathrm{t^hut}$

Table 3.1: Cantonese vowel fronting (Cheng, 1991, as analyzed in Flemming, 2001)

The lack of contrast between [y] and [u] in this environment has a coarticulatory basis, since [u] requires more articulatory effort and likely is more perceptually confusable with [y] when it occurs between two coronals. Coronal consonants have a high F2 value corresponding to their place of articulation, as do front vowels, so the coronal-vowel-coronal transition is easy to achieve in t^hyt (‘to take off’). However, back vowels have a low F2 value and a farther-back place of articulation, so a coronal-back vowel-coronal sequence like the hypothetical t^hut in Table 3.1 requires a quick transition from high F2 to low F2, and then back to high F2 (a transition from a coronal place of articulation to a back vowel and then back again to coronal). This results in greater articulatory effort and likely greater perceptual confusability between [u] and [y], since their F2 values are are less distinct in this environment, so the fact that [u] does not occur in this specific environment can be thought as being motivated by the avoidance of articulatory effort and/or perceptual confusability between contrastive segments.

An analysis of this neutralization as being directly triggered by a combination of coarticulatory pressures is exactly the approach taken by an account like Flemming’s (2001) unified model of phonetics and phonology. In this account, there are target constraints requiring [y] and [u] to have high and low F2 values, respectively, which must be a certain acoustic distance apart. When [u] is next to one coronal consonant, as in $t^h uk$ ('bald head'), [u] coarticulates with the preceding [t] and, as a result, has a slightly higher F2 value, which is closer to the target value for [y], but not too close. However, when [u] is surrounded by coronals, as in the hypothetical $t^h ut$, coarticulation with the consonants raises its F2 significantly, bringing it much closer to the target value for [y]. Because the F2 values for [y] and [u] are too close close together in this case for them to be reliably distinguished from each other, and because realizing a lower F2 value for [u] would require too great of articulatory effort, the contrast between [y] and [u] is neutralized. In this way, frameworks like Flemming's (2001) can model phonological patterns by defining phonological units in terms of fine-grained phonetic detail.

In this chapter, I outline a process of tone sandhi, which I term **low tone spread**, in SMPM that presents significant problems for the line of analysis outlined above. This process, like Cantonese vowel fronting, is a multiply-triggered alternation that appears to be conditioned by a combination of coarticulatory pressures. However, I argue that an account defining phonological units in terms of fine-grained phonetic detail is unable to model the alternation. The reason for this is that low tone spread interacts opaquely with another tone sandhi process in SMPM, and the phonetically-detailed account cannot model this. On the contrary, a phonological analysis that makes use of units of representation and constraints defined at a more coarse-grained level than phonetic detail is able to both model the alternation and account for its opaque interaction with the other sandhi process. I conclude that phonological units of representation are best thought of as being defined at a more coarse-grained level than that of fine-grained phonetic detail (Keating, 1996; Hayes, 1999; Zsiga, 2000; Smith, 2005; Kingston, 2007; Cohn, 2007; Bermúdez-Otero, 2011). In addition, I argue that these coarse-grained phonological constraints are still broadly phonetically grounded, and that their phonetic grounding can be modeled through a process of constraint induction like that outlined in Hayes (1999).

3.2 Low tone spread

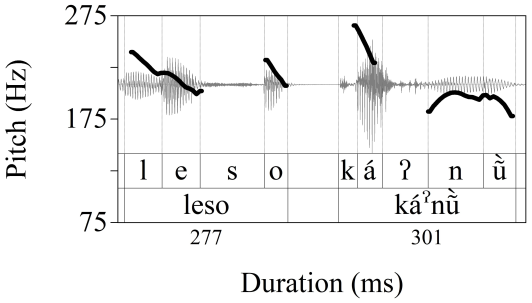
3.2.1 The process

The process of low tone spread in SMPM was originally described in Hedding (2019b). It is a process by which word-final Low tones spread to some adjectives that begin with a High tone.

(1) **No low tone spread**

le^{h}so ká^{ʔ}nũ̀
rabbit big

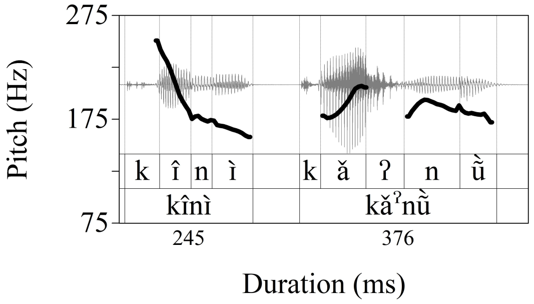
'A big rabbit.'



(2) **Low tone spread**

kînì kǎ^{ʔ}nũ̀
pig big

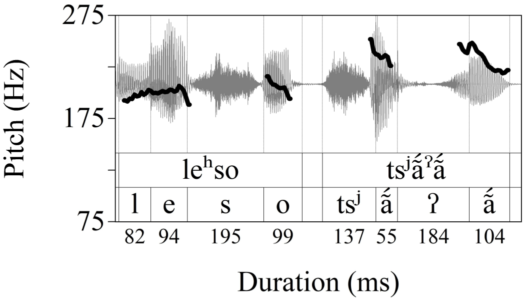
'A big pig.'



One puzzling fact about this process is that it applies to some adjectives, but not to others. For example, the examples in (3)-(4) show that low tone spread applies to the adjective *tsʲá̃ˀá̃* ('dirty'). However, (5)-(6) show another High-initial adjective *íʰtʃi* ('dry') surfacing with an initial High tone in all phonological environments, even if the preceding tone is Low.

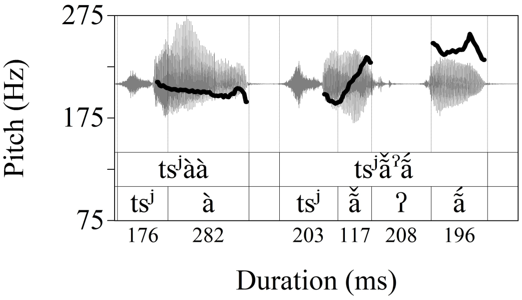
(3) **No low tone spread**

leʰso tsʲá̃ˀá̃
rabbit dirty
'A dirty rabbit.'



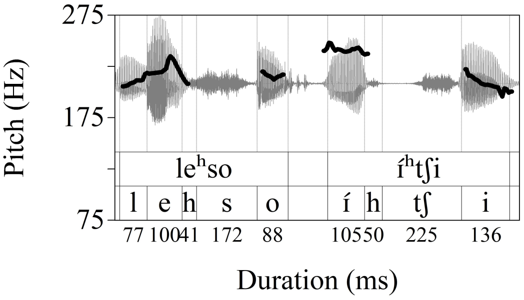
(4) **Low tone spread**

tsʲàà tsʲǎ̃ˀá̃
clothes dirty
'Dirty clothes.'



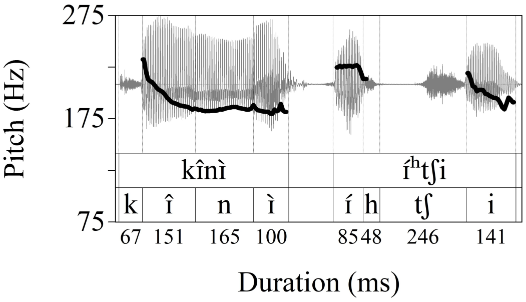
(5) **No low tone spread**

leʰso íʰtʃi
rabbit dry
'A dry rabbit.'



(6) **No low tone spread**

kînì íʰtʃi
pig dry
'A dry pig.'



Examples (1)-(6) show that low tone spread applies to some High-initial adjectives, but not all. However, the lack of application of low tone spread in (5)-(6) does not reflect lexical exceptionality, but is rather due to the fact that low tone spread has a highly-specific phonological conditioning environment. Specifically, low tone spreads only to adjectives that begin with a High-toned, laryngealized low vowel [áˀ]. This can be seen in the following table, which shows all of the adjectives that I know of to which low tone spread applies. As can be seen, the initial vowel of each is a High-toned, laryngealized, low vowel.

káˀnũ̀ ~ kăˀnũ̀ ‘Big’	jáˀà ~ jăˀà ‘Brown’	láˀnũ̀ ~ lăˀnũ̀ ‘Old’	sã́ˀã́ ~ sã̆ˀã́ ‘Spanish	tsʲáˀjì ~ tsʲăˀjì ‘Rotten’
kʷáˀà ~ kʷăˀà ‘Red’	tʃáˀmã̀ ~ tʃăˀmã̀ ‘Crushed’	táˀβì ~ tăˀβì ‘Broken’	ⁿdáˀβì ~ ⁿdăˀβì ‘Poor’	tsʲã́ˀã́ ~ tsʲã̆ˀã́ ‘Dirty’

Table 3.2: Adjectives that undergo low tone spread

Low tone spread, then, can be understood as a general process by which a Low tone spreads to a high-toned, laryngealized low vowel across a word boundary. This is schematized in rule form below:

(7) **Low tone spread**

/v̀ + áˀ/ → [v̀ + ăˀ]

3.2.2 The ingredients

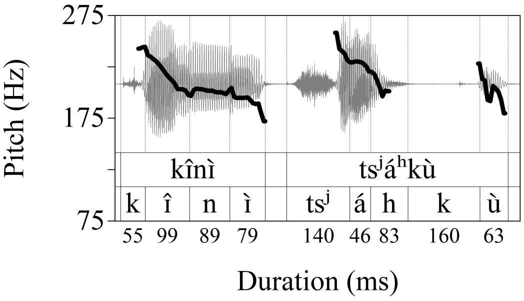
An important characteristic of the conditioning environment of low tone spread is that all three characteristics—laryngealization, low vowel quality, and High tone—are necessary in order for the process to apply. If an adjective's initial vowel has two of the three characteristics, but not the third, then the process does not occur.

The first of these necessary ingredients is laryngealization. If an adjective begins with a vowel that has the other required characteristics, namely low vowel quality and High tone, but is not laryngealized, then Low tone does not spread. This is seen in (8)-(10), where adjectives that begin with a High-toned, non-laryngealized low vowel do not undergo low tone spread.

(8) **No low tone spread**

kînì tsʲáʰku
pig alive

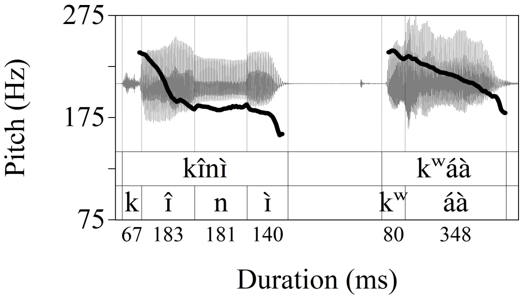
‘A live pig.’



(9) **No low tone spread**

kînì kʷáà
pig blind

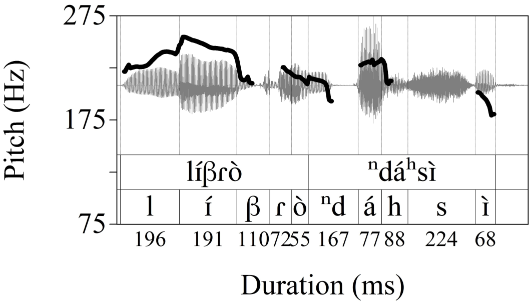
‘A blind pig.’



(10) **No low tone spread**

líβɾò ⁿdáʰsi
book closed

‘A closed book.’

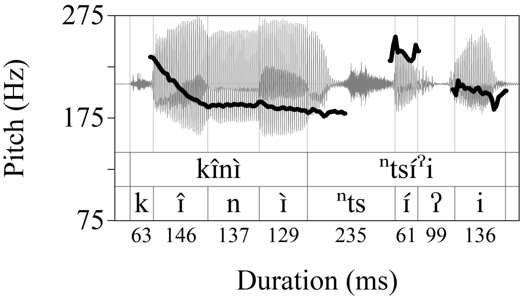


Because High-toned, non-laryngealized low vowels do not undergo low tone spread, laryngealization is a necessary component in the conditioning of the alternation. Another crucial component is low vowel quality—if an adjective begins with a laryngealized, High-toned vowel, but that vowel is not low, then low tone spread does not apply. This can be seen in (11) and (12), where the adjectives *ⁿtsíˀi* ('blue') and *sóˀo* ('deaf') begin with High-toned laryngealized vowels that are not low (they are high and mid, respectively). When they occur after a Low tone, they do not undergo low tone spread.

(11) **No low tone spread**

kînì ⁿtsíˀi
pig blue

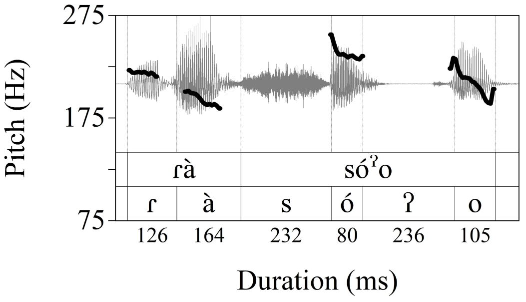
'A blue pig.'



(12) **No low tone spread**

ɾà sóˀo
3M deaf

'The deaf man.'



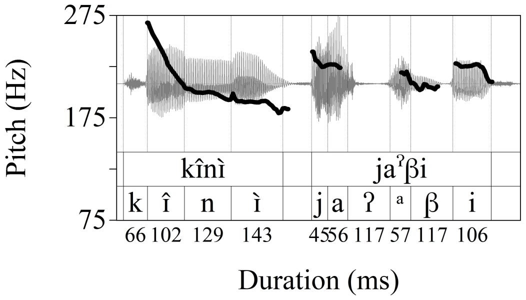
Because High-toned, laryngealized non-low vowels do not undergo low tone spread, vowel quality is also a necessary component in the conditioning environment of the process. And in the same way, High tone is also necessary. This can be seen in the following examples, where the adjectives *jaˀβi* ('expensive') and *tãˀmã̀* ('flat') begin with laryngealized low vowels that bear a Mid tone. When preceded by a Low tone, these adjectives are not the target of spread.[2]

---

[2] It is likely relevant here that mono-moraic Low-Mid contours are not attested in SMPM's tonal inventory. If low tone spread is a structure-preserving process, then the derivation of a Low-Mid contour might be blocked by phonotactic constraints.

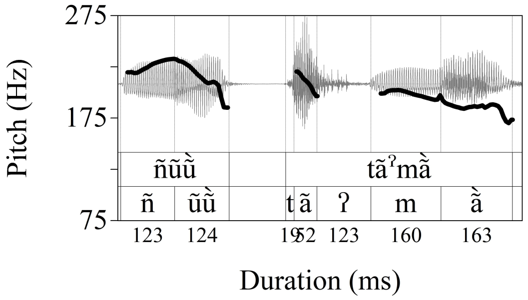
(13) **No low tone spread**

kîǹı ja^{ʔ}βi
pig expensive
‘An expensive pig.’



(14) **No low tone spread**

ñũũ̀ tã^{ʔ}mã̀
town flat
‘A flat town.’



The above examples show that each of the three characteristics of the initial vowel of the adjective are necessary for the conditioning of low tone spread. The adjective’s initial vowel must be laryngealized, must be the low vowel [a], and must bear a High tone. If any of these three characteristics is missing, then low tone spread does not apply. In the following section, I will show that this high degree of specificity is phonetically grounded.

3.2.3 Phonetic grounding

Each of the three necessary characteristics for the application of low tone spread reflect this highly-specific process’s phonetic grounding. That is, the application of low tone spread in this environment makes sense when one examines it from the perspective of coarticulatory pressures on the realization of pitch. Specifically, laryngealization lowers F0, and low vowels have a lower intrinsic F0 than high vowels in SMPM and cross-linguistically (Whalen and Levitt, 1995). Additionally, Whalen & Levitt note that intrinsic F0 effects are greatest at the high end of the pitch range, so it is exactly on High-toned vowels that we would expect to see the greatest intrinsic F0 effect. Finally, a High tone preceded by a Low tone might be expected to lower somewhat via tonal coarticulation. I will walk through the phonetic grounding of each of these characteristics in turn.

The first point is that laryngealization lowers pitch in SMPM, as it does in many languages (Keating et al., 2015). The lowering effect of laryngealization on pitch in SMPM can be seen in some of the preceding examples, but is more robustly shown in Figure 3.2, which compares High- and Low-toned laryngealized vowels with High- and Low-toned modal vowels.[3] Pitch drops throughout the timecourse of the vowel in laryngealized vowels, but stays relatively steady in modal vowels.

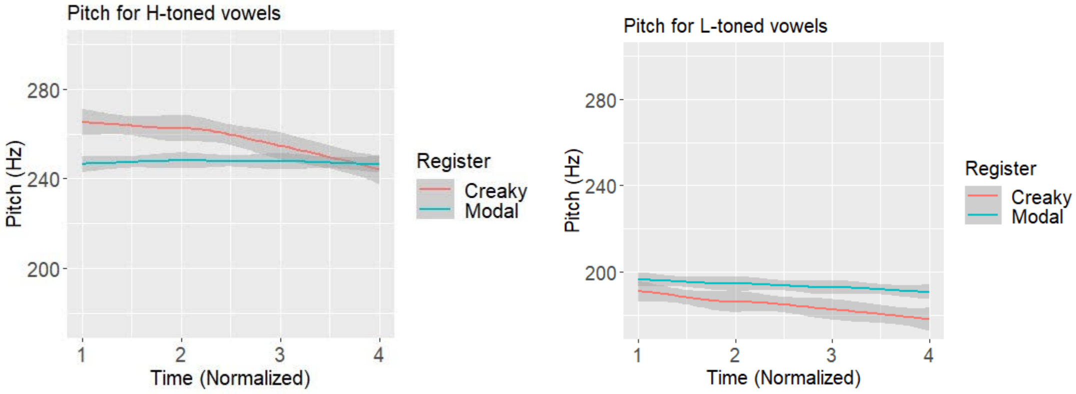


Figure 3.2: Pitch (Hz) for laryngealized and modal vowels with an H tone (left; 26 Creaky, 31 Modal) and L tone (right; 27 Creaky, 24 Modal)

The second and third pieces of the phonetic grounding of low tone spread have to do with intrinsic F0. It is arguably a cross-linguistic universal that, all else being

---

[3] Note that these were not controlled for vowel height or onset voicing. Many of the H-initial modal words begin with voiced consonants, which might have had a lowering effect on their initial F0 (e.g., Kingston, 2011).

equal, low vowels have a lower intrinsic F0 than high vowels (Whalen and Levitt, 1995), and SMPM is no exception. Figure 3.3 shows that High-toned high vowels ([í] and [ú]) have a higher pitch than High-toned low vowels ([á]).

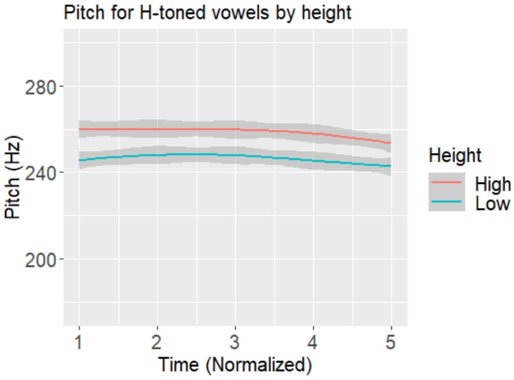


Figure 3.3: Pitch (Hz) for high and low vowels with a High tone (23 High, 20 Low).

Additionally, it is the case that intrinsic F0 differences are greatest in the high portion of a speaker's pitch range, meaning that high and low vowels tend to have similar F0 in the lower portion of a speaker's pitch range (Whalen and Levitt, 1995). This finding also holds for SMPM, as seen in Figure 3.4, which shows that Low-toned high and low vowels have roughly the same pitch. In this light, the fact that low tone spread applies only to High-toned vowels makes some sense—it is exactly in the high portion of the pitch range that we see the greatest differences in intrinsic F0.

Finally, the fact that a preceding Low tone triggers low tone spread makes sense from a coarticulatory standpoint: One one tone (T1) ends at a pitch level different from the beginning pitch level of the following tone (T2), then in the sequence T1 + T2, the beginning pitch of T2 might undergo a degree of assimilatory coarticulation to the ending pitch level of T1 (and vice versa). For example Xu (1994) showed that, when a rising

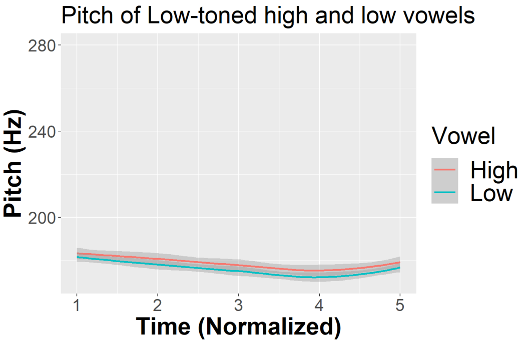


Figure 3.4: Pitch (Hz) for high and low vowels with a Low tone (24 High, 32 Low).

tone is preceded by a high tone and followed by a low tone in Mandarin Chinese, instead of the predicted high-low-high-low pitch sequence, the rising tone is often realized with falling pitch, creating a high-high-low-low pitch sequence. This suggests that a High tone in SMPM might similarly lower when preceded by a Low tone. However, this does not appear to be the case—the pitch of word-initial High tones appears not to be greatly affected by the preceding tone. This can be seen in the following aggregated pitch plot, which shows the pitch of root-initial, High-toned vowels based on the preceding tone. The measurements were taken from the initial vowel of adjectives in N-Adj sequences, and from the initial vowel of the second noun in N-N possessive constructions, each of which were embedded in a carrier sentence. Examples of each type of construction are given below, with the underlined portion showing the vowel from which the pitch readings were extracted. The tone preceding these vowels was varied between High, Mid, and Low.

(15)	kã̂ˀ=ĩ̀	ñũˀũ̀	íʰtʃî	βitsĩ
	POT.say=1SG	fire	dry	now

“I will say ‘dry fire’ now.”

(16)	kã̂ˀ=ĩ̀	ñũˀũ̀	léló	βitsĩ
	POT.say=1SG	fire	skunk	now

“I will say ‘the skunk’s fire’ now.”

As shown in Figure 3.5, the pitch of High tones in these contexts is relatively consistent regardless of the preceding tone. That is, the initial High tone of an adjective or noun in a N-Adj or N-N sequence is not greatly affected by the preceding pitch in this task. In this light, the coarticulatory pitch-lowering effect that a preceding Low tone has on an adjective-initial High tone appears to be very small.

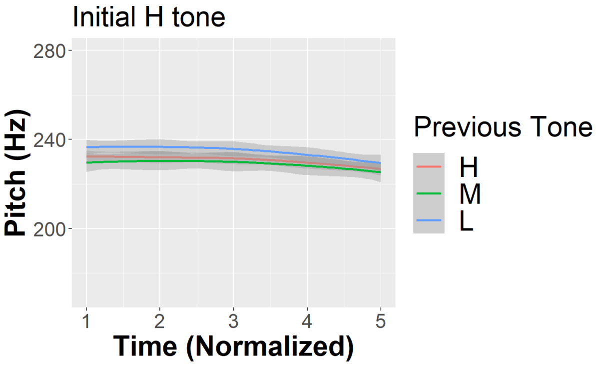


Figure 3.5: Pitch (Hz) for High-toned initial vowels of adjectives based on preceding tone (30 after High, 30 after Mid, 28 after Low)

In SMPM, then, a preceding Low tone might be expected to have a lowering effect on a following High tone, but this effect was not found in the task described above. However, it is the case that most of the required triggers of low tone spread have a pitch-lowering effect: Laryngealization lowers F0, and low vowel quality also lowers F0. Additionally, it is only in the high portion of the pitch range that large intrinsic F0 differences are expected. When one considers that both laryngealization and low vowel quality lower F0, and that the co-occurrence of these characteristics allows a preceding Low tone to spread to the vowel in question, the highly specific process of low tone spread makes some phonetic sense. That is, it appears that the pitch-lowering properties of laryngealization and low vowel quality combine to make the initial vowel of the adjective ‘compatible’ with low pitch, which then allows the preceding Low tone, whose pitch is necessarily low, to spread. That being said, it is not the case that low tone spread is simply a process of coarticulation. Instead, low tone spread is a phonological alternation, as argued in the following section.

3.2.4 Phonological status

Despite low tone spread’s highly specific and phonetically-grounded nature, it is not simply a process of coarticulation between a High tone, laryngealization, low vowel quality, and a preceding Low tone. Instead, it involves an alternation between tonal categories, and this alternation appears to be triggered when a threshold of coarticulatory pitch-lowering pressures is reached. There are several reasons to believe this, the first being that the pitch lowering seen in low tone spread is categorical in a way that is separate from its coarticulatory sub-parts.

This can be seen in that the pitch lowering in low tone spread is greater than the pitch-lowering pressures that trigger it. To show this, it is worth considering the effect that each of the component parts—laryngealization, low vowel quality, and a preceding Low tone—have on the pitch of an initial High tone. Figure 3.2 showed that pitch in laryngealized vowels with a High tone drops by as much as ∼20 Hz, and Figure 3.3 showed that the intrinsic F0 of low vowels with a High tone is ∼10 Hz lower than that of High-toned high vowels. It appears, then, that laryngealization and low vowel quality each have a substantial pitch-lowering effect on High tones—for reference, in many phonological environments, the difference between a High and Mid tone is about 25-30 Hz.

With the understanding that laryngealization and low vowel quality have substantial effects on pitch, but a preceding Low tone does not, it is useful to consider the following two examples in (17)-(18). The adjective *tsʲã́ˀã́* ('dirty') is the same in each case, but it is preceded by a Mid tone in (17) and by a Low tone in (17). As discussed earlier, the pitch of an initial High tone is not greatly affected by the level of the preceding tone. However, in (18), the pitch of the first vowel of the adjective begins at a level ∼45 Hz lower than in (17) as a result of sandhi.

(17) **No low tone spread**

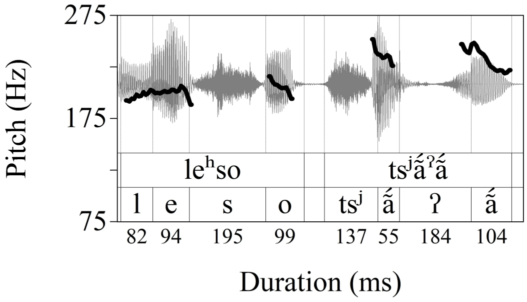
leʰso tsʲã́ˀã́
rabbit dirty

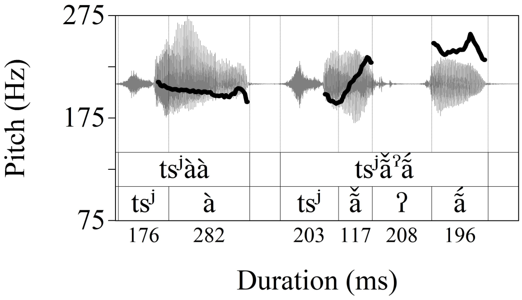
'A dirty rabbit.'

(18) **Low tone spread**

tsʲàà tsʲã̌ˀã́
clothes dirty

'Dirty clothes.'

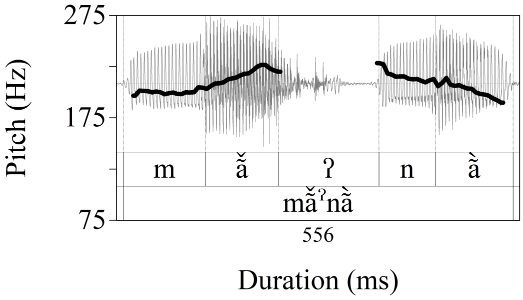




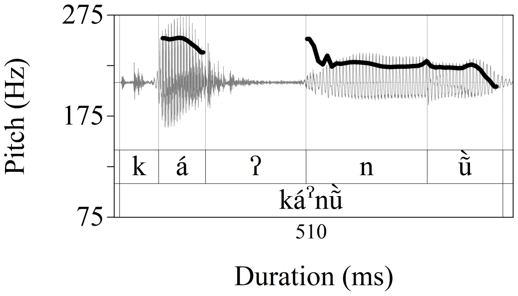
This difference is far greater than what is expected due to tonal coarticulation with the preceding Low tone. Instead, it appears that the preceding Low tone adds just enough coarticulatory pitch-lowering pressure that, when it acts in conjunction with laryngealization and low vowel quality, an alternation between High and Low-High Rise is triggered.

What is more, the alternation between High and Low-High Rise appears to be neutralizing. Crucially, it is the case that Low and Low-High rising tones *do* contrast on laryngealized, low vowels. For example, the adjective *mã̌ˀnã̀* (‘sleepless’) always surfaces with a Low-High rising tone, regardless of phonological context (19). This contrasts with *káˀnũ̀* (‘big’), which usually surfaces with an initial High tone (20). However, when low tone spread applies to an adjective with an initial High-toned, laryngealized low vowel, the resulting pitch contour is extremely similar to the pitch contour corresponding to an underlying Low-High rise in the same phonological environment. This can be seen by comparing the pitch contour of the first vowel in (19) with the that of the first vowel in (21).

(19) **Underlying Rise**
mã̌ˀnã̀
‘Sleepless’



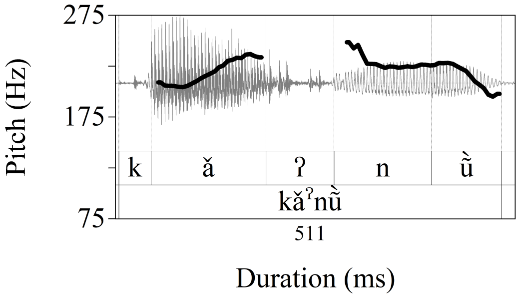
(20) **Underlying High**
káˀnũ̀
‘Big’



(21) **Derived Rise**

kăˀnũ̀

‘Big’



Of course, the preceding examples do not conclusively show that the process of low tone spread is phonetically complete. That is, it might be the case that Rises derived via low tone spread are acoustically distinct from underlying Rises. To convincingly show whether a neutralization is phonetically complete or incomplete, a large amount of data is required (Nicenboim et al., 2018). However, within-category variance for acoustic measures associated with phonological items is usually relatively large (e.g., Warner, 2011)(Warner and Tucker, 2011). Given the high degree of similarity between the rises in (19) and (21), it is likely that the two fall within the expected range of variability for underlying Rises. What is more, even if the process of low tone spread were shown to be phonetically ‘incomplete’ in the sense that derives Rises were different from underlying Rises, this would not necessarily be definitive evidence against the phonological nature of the alternation, since phonetically-incomplete neutralization can still reflect wholesale phonological change, as in the case of Mandarin Tone 1 sandhi (Du and Durvasula, 2020) or Uyghur backness harmony (McCollum, 2019).

Despite its highly-specific and phonetically-grounded nature, low tone spread appears to be a categorical, neutralizing (or near-neutralizing) alternation. In this sense, it appears similar to the Cantonese vowel fronting example discussed earlier. In Cantonese, the coarticulatory fronting effects of coronal consonants on the back vowel [u] appear to interact cumulatively to trigger neutralization between [u] and [y]—when only one coronal is present, neutralization isn't triggered, but when two are present, neutralization is triggered. In SMPM, the cumulative coarticulatory effects of laryngealization, low vowel quality, and a preceding Low tone might similarly be analyzed as directly triggering neutralization between High and Low-High rising tones. This is shown in Table 3.4, where the individual coarticulatory effects of coronal consonants are not enough to trigger vowel fronting in Cantonese, just as the individual coarticulatory effects of laryngealization and vowel height are not enough to trigger low tone spread in SMPM. In each case, the coarticulatory pressures only trigger an alternation when they are all present.

	Cantonese	**SMPM**
One trigger	✓ $t^h uk$	✓L + í^{ʔ}
	✓ $k^h ut$	✓L + á
Multiple triggers	✗ $t^h ut$	✗ L + á^{ʔ}

Table 3.3: Multiple triggering of vowel fronting in Cantonese and low tone spread in SMPM

Given the multiply-triggered nature of the alternation, and the fact that such alternations have been used to argue for the inclusion of various amounts of phonetic detail in phonological representations (Flemming, 2001; Lionnet, 2017), it is worth examining what sort of phonological analysis is best suited to the data. As I will argue in the following section, any phonological analysis of low tone spread must necessarily make use of phonological units that are defined at a more coarse-grained level than that of fine-grained, physical phonetic detail.

3.3 Phonological analysis

In the following sections of this chapter, I will argue that a phonological analysis that makes use of units of representation and constraints defined at the level of fine-grained phonetic detail is unable to adequately model the process of low tone spread because it cannot account for its opaque interaction with a separate sandhi process. However, a phonological analysis whose units of representation and constraints are defined at a more coarse-grained level than that of fine-grained phonetic detail is able to derive the alternation and its opaque interaction with the separate sandhi process, and to do so while reflecting the alternation's phonetic grounding. In general, I will refer to these two types of approaches as **direct phonetics** and **indirect phonetics**, as detailed below. In a direct phonetics framework, phonological units are defined in terms of their physical realization. This means that, for example, what it means for a vowel to have a high tone is for the fundamental frequency of that vowel to be a certain value, modulo speaker normalization.

Direct phonetics	**Indirect phonetics**
Phonological units = fine-grained phonetic detail (e.g., High tone = 245 Hz)	Phonological units = coarser-grained than phonetic detail (e.g., features, segments)
Phonological constraints operate at this level of granularity	Phonological constraints operate at this level of granularity

Table 3.4: Direct and Indirect Phonetics

3.3.1 Direct phonetics

The essence of a direct phonetics model is that neutralizations between two phonological categories can be directly triggered by a combination of physical coarticulatory pressures. In this type of analysis, the phonological grammar has access to the fine-grained phonetic effects that coarticulation of a sound with its surrounding environment will have. When these coarticulatory effects are strong enough that realizing a phonological unit in a certain context will (1) require too much articulatory effort and/or (2) result in insufficient perceptual distance between two contrastive phonological categories, then the grammar allows for neutralization of the contrast.

For example, in Cantonese, the vowels [u] and [y] contrast, but not in between coronal consonants, as shown below.

k^{h}yt 'Decide'	t^{h}uk 'Bald head'
t^{h}yt 'to take off'	k^{h}ut 'bracket'
	*t^{h}ut

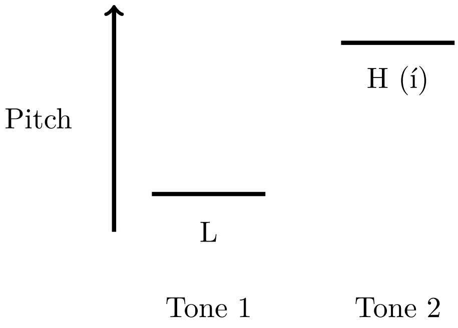
Table 3.5: Cantonese vowel fronting (Cheng, 1991, as analyzed in Flemming, 2001)

One of the main differences between [u] and [y] is their F2 profile: [u] generally has low F2, and [y] generally has high F2. When [u] is coarticulated with a preceding coronal consonant as in [t$^{\mathrm{h}}$uk] (‘decide’), there is an articulatory and acoustic transition from the [t]’s high F2 to the [u]’s low F2. This transition requires articulatory effort, and also results in a portion of the [u] vowel having a higher F2 than its target value, but the sequence is still allowed. However, when [u] is flanked on both sides as in the hypothetical and unattested form *t^hut, there is a transition from the initial consonant’s high F2 to the vowel’s low F2, and then back to a high F2 again. In this case, the articulatory effort required is higher relative to the effort required to produce a word like [t$^{\mathrm{h}}$uk] (‘decide’). Additionally, the F2 of the [u] is significantly raised, bringing its value much closer to the usual F2 value of [y]. Because [u] and [y] are too close to each other in acoustic space (which is assumed in this example to map straightforwardly onto perceptual space), and because too much articulatory effort would be required to further differentiate them in this environment, the grammar allows the two categories to be neutralized. That is, in the face of great articulatory effort and/or insufficient perceptual distance between two contrastive elements, the grammar allows these two categories to be neutralized to one. In this way, the pressure to maintain contrast can be overruled directly by coarticulatory and perceptual pressures.

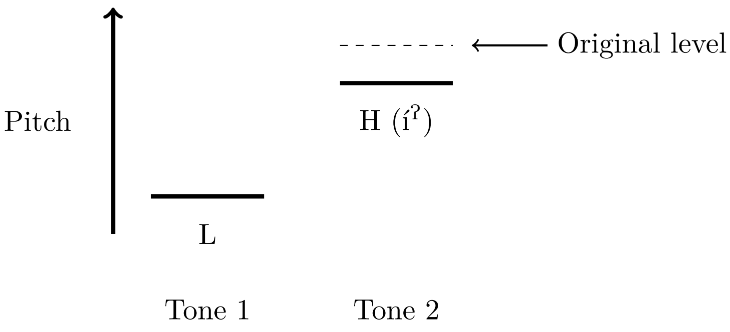
Given that low tone spread in SMPM is a multiply-triggered alternation that appears to be driven by a combination of coarticulatory pressures, it is a good candidate for analysis in a direct phonetics framework. That is, when a High tone on a laryngealized low vowel is preceded by a Low tone, the cumulative pitch-lowering pressures of the laryngealization, low vowel quality, and the preceding Low mean that a High tone either requires too much articulatory effort, is not perceptually distinct enough from another contrastive category (maybe a Mid tone, for example), or both. In this case, the grammar allows a neutralization of contrast, and low tone spread occurs. This section walks through the basics of this type of analysis.

3.3.1.1 A direct phonetics analysis of low tone spread

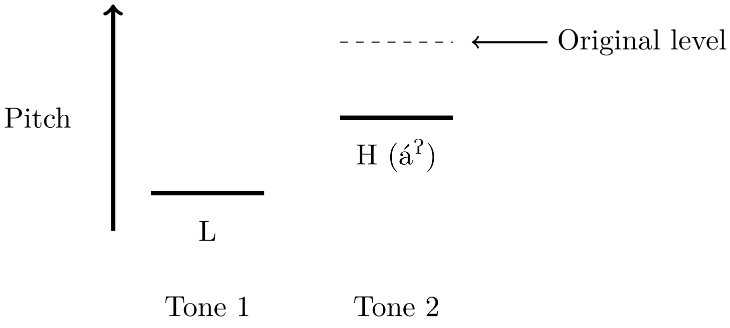
Under a direct phonetics analysis of low tone spread, tones are connected with pitch targets, defined either articulatorily or perceptually. In a sequence of Low and High tones, where the High is realized on a non-laryngealized, non-low vowel, the pitch of the High tone is consistent with its target realization.



However, when in a sequence of a Low and High tone, where the tone is realized on a laryngealized vowel, the pitch corresponding to the High tone is dragged down slightly by laryngealization. This means that the pitch of the High tone is slightly lower than its target realization.



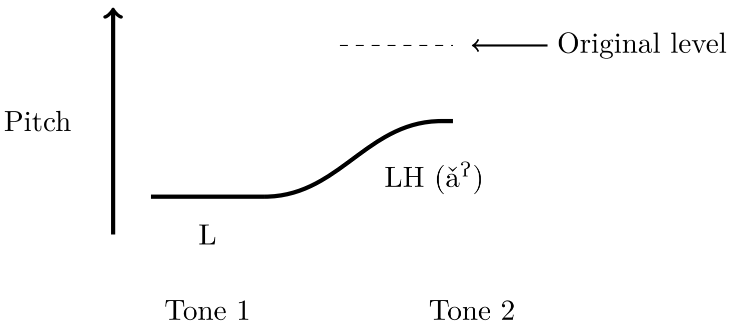
In the same sequence of Low and High tones, where the High is realized on a laryngealized low vowel, laryngealization and low vowel quality combine to lower the pitch of the High tone even more. This means that the pitch of the High tone is significantly lower than its target realization.



At this point, the pitch associated with the High tone is very far from its target. This might mean that the High is confusable with, for example, a Mid tone, which also has a pitch target in between that of a Low and a High tone. Because the High tone is too confusable with a Mid tone, and because it would take too much effort to realize the High with high enough pitch to reliably differentiate it from a Mid in this context, the grammar allows a neutralization of contrast, and low tone spread applies.[4]

---

[4] Interestingly, the neutralization is not to Mid or Low, as we might expect under this type of account. Instead, the neutralization is between H and LH. One point that might help in understanding why H



The illustrations above can be formalized in a Harmonic Grammar framework (Smolensky and Legendre, 2006; Hayes and Wilson, 2008; Pater, 2009; McCarthy and Pater, 2016), which is especially useful for modeling 'ganging-up effects' like those seen here. In this approach, every violation of a constraint is multiplied by that constraint's weight, and the violations are added together to produce a candidate's harmony score. The candidate with the highest harmony score wins. Below are constraints that would necessarily be active in a direct phonetics framework—constraints defining tones by their physical realization, and a constraint penalizing the neutralization of contrasts. As in Flemming (2001) a categorical constraint on contrast maintenance is violated whenever a contrast is neutralized.

- F0[H] = 260 HZ: Multiply constraint weight by Hz deviation from 260 Hz for a High tone.
- F0[L] = 200 HZ: Multiply constraint weight by Hz deviation from 200 Hz for a Low tone.

---

neutralizes to LH and not to M or L is that LH tones contain an H, which means that the underlying H can technically be understood to still be present in the output. Another consideration is that SMPM has no LM contours, but does have LH contours, so a rising tone might more readily signal the presence of an H in this environment than a level tone. In either case, though, H and LH do contrast on low, laryngealized vowels, so this process does still constitute a neutralization.

- MaintainContrast[h/lh]: Multiply constraint weight by -1 if contrast between H and LH is neutralized.

In a sequence of a Low and High tone, with the High tone linked to a non-laryngealized, high vowel, the High tone is realized at its target pitch. In this case, Candidate A has the highest harmony score, since neither the pitch target constraints nor the constraint requiring the maintenance of contrast incurs any violations. Candidate B, which neutralizes the contrast between H and LH, incurs a violation of the contrast maintenance constraint, thereby receiving a lower harmony score.[5]

(22)

L H | | V # i	MaintainContrast wt=100	f0[l] = 200 wt=10	f0[h] = 260 wt=4	**Harmony** **score**
☞ a. L (200) + H (260)				
b. L (200) + LH	-1			-100

In a sequence of a Low and High tone, with the High tone linked to a laryngealized high vowel, the pitch of the High tone is realized lower than its target value. This results in violations of the pitch target constraint for High tones, but these violations are not severe enough to favor Candidate B.

---

[5] Candidate B's LH tone is derived through tone spreading and not tonal epenthesis. However, because this still creates a LH tone and neutralize the H/LH distinction, it incurs a violation of MaintainContrast.

(23)

L H │ │ V # iˀ	MaintainContrast WT=100	F0[L] = 200 WT=10	F0[H] = 260 WT=4	**Harmony** **score**
☞ a. L (200) + H (240)			-20	-80
b. L (200) + LH	-1			-100

Finally, in a sequence of a Low and High tone, with the High tone linked to a laryngealized low vowel, the pitch of the High tone is realized even lower than its target value. The resulting violations of the pitch target constraint for High tones lead Candidate A to have a lower harmony score than Candidate B, which makes Candidate B the winner. That means that, in this case, neutralizing the contrast between H and LH is better than realizing the H, since the pitch of the H tone is too far from its target.

(24)

L H │ │ V # aˀ	MaintainContrast WT=100	F0[L] = 200 WT=10	F0[H] = 260 WT=4	**Harmony** **score**
a. L (200) + H (230)			-30	-120
☞ b. L (200) + LH	-1			-100

This framework is missing some of the ingredients commonly used in direct phonetics framework. It doesn't have, for example, *EFFORT constraints like those found in (Kirchner, 2000, 2004) or constraints defining the minimum allowable phonetic distance between contrasting categories (Flemming, 2001). However, it does illustrate how multiple coarticulatory pressures can combine to trigger a phonological alternation in a phonological model that directly incorporates gradiently-defined, coarticulatory pressures. I would like to argue that the approach sketched here, and any approach like it, is inadequate for the analysis of low tone spread in SMPM because its reliance on physical, coarticulatory pressures means that it is unable to capture an opaque interaction with a separate tone sandhi process in the language.

3.3.1.2 Opacity and its consequences for the direct phonetics analysis

There is a separate tone sandhi in process in SMPM, also initially described in Hedding (2019b), which derives word-final Low tones. This sandhi process, which I term **rise flattening**, flattens underlying Low-High Rising tones to a level Low tone (25).

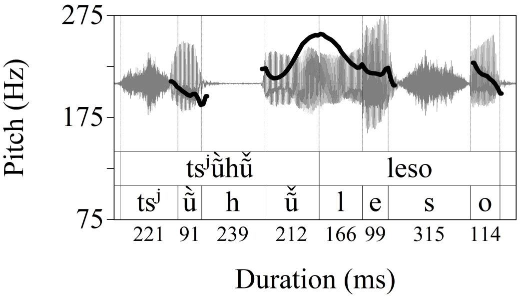
(25) **Rise Flattening:**

/LH # H/ → [L # H]

This can be seen in the following examples, where the underlying word-final Rise on *tsʲũ̀ʰũ̌* (‘turkey’) in (26) surfaces as a level Low tone in (27).

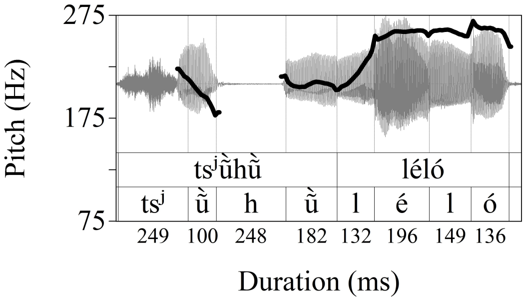
(26) **No rise flattening**

nì-^{n}tsì^{h}kù ts^{j}ũ̀^{h}ũ̌ le^{h}so
COMPL-chase turkey rabbit
‘The turkey chased the rabbit.’



(27) **Rise flattening**

nì-^{n}tsì^{h}kù ts^{j}ũ̀^{h}ũ̀ léló
COMPL-chase turkey skunk
‘The turkey chased the skunk.’



There are good reasons to believe that this process is phonological, which will be further outlined in Chapter 4. For now, it suffices to note that this process derives a Low tone whose pitch level and contour is essentially the same as an underlying Low tone. The important point here is that rise flattening may derive the conditioning environment for low tone spread to apply. Recall that low tone spread applies when a Low tone precedes a High-toned, laryngealized low vowel (28).

(28) **Low Tone Spread**

/v̀ # á^{ʔ}/ → [v̀ # ǎ^{ʔ}]

Because rise flattening derives a word-final Low tone, when it applies before an adjective that undergoes low tone spread, the process should apply. This can be seen in Figure 3.6.

However, Low tones derived from rise flattening do not trigger low tone spread. This can be seen in the following two examples, which show the word *kò^{h}ǒ* (‘snake’)

$$\boxed{\text{Input}} \qquad \boxed{\text{Rise Flattening}} \qquad \boxed{\text{Low Tone Spread}}$$

/v̌ # áˀ/ ⟶ [v̀ # áˀ] ⟶ [v̀ # ǎˀ]

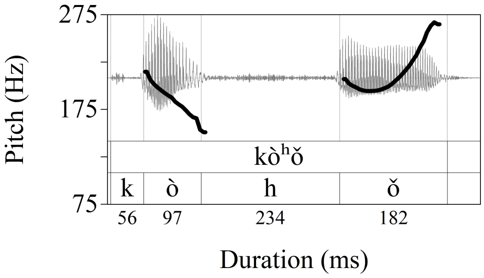
Figure 3.6: Expected feeding relationship between Rise Flattening and Low Tone Spread

in its base form (29) and when it has undergone rise flattening (30). Crucially, even though the final derived Low tone of 'snake' is followed by a high-toned, laryngealized low vowel, low tone spread does not apply.

(29) **No rise flattening**

kòʰǒ
snake

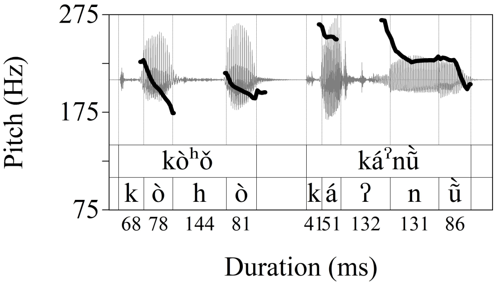
'A snake.'



(30) **Rise flattening, no low tone spread**

kòʰò	kàˀnȕ
snake	big

'A big snake.'



The fact that the derived Low tone in (30) does not trigger tone spread is, I argue, an insurmountable problem for a direct phonetics account of the process. The reason for this is that the direct phonetics account relies directly on the physical results of coarticulation to trigger the neutralization of contrast. That is, when coarticulatory pitch-lowering pressures of laryngealization, low vowel quality, and preceding Low tone combine, the High tone is either too articulatorily difficult to realize, is too confusable with another tone, or both. These articulatory and/or perceptual pressures overrule the pressure to maintain contrast between H and LH, and low tone spread applies. The problem that opacity poses for a direct phonetics account is that the *exact same* articulatory and perceptual pressures that presumably trigger low tone spread are present whether the conditioning environment for low tone spread involves an underlying Low tone or a derived Low tone. However, these pressures only trigger the alternation in one case but not the other.

Before completely refuting the direct phonetics approach on this basis, though, it is worth considering two potential ways that it might derive the opacity. One thing that would help a direct phonetics account get around the opacity problem is if derived Low tones like that in (30) are reliably higher than the pitch of the underlying Low tones that trigger low tone spread. If this is the case, then one might argue that the same physical coarticulatory pressures are *not* present in the context that triggers low tone spread and the context that does not. For example, if low tone spread is partially triggered by the articulatory difficulty of transitioning from a low to a high pitch, and the pitch of derived Low tones is not as low as that of underlying Low tones, then the transition from a derived Low to a High might be less articulatorily difficult because it requires less drastic of a change in pitch. If this were the case, then this decreased difficulty might mean that there is not enough cumulative coarticulatory pressure to drive neutralization of tonal categories and trigger low tone spread. However, this is not the case: there is essentially no difference between the pitch of underlying Low tones that trigger low tone spread and derived Low tones that do not.

This can be seen in the results of an informal production task carried out with Consultant 1, who produced carrier sentences containing N-Adj sequences, repeating each utternace 4 times. The N in the N-Adj sequence had either a final underlying Low tone or a final underlying Low-High rising tone, and the preceding tone and the voicing of the preceding consonant were held constant. The following adjective always began with a High tone, triggering preceding Low-High tones to undergo rise flattening and become derived Low tones. This setup allowed for the comparison of the pitch of underlying Low tones and derived Low tones, which is shown below in Figure 3.7. As can be seen, both begin and end at the same point, they follow a very similar trajectory, and the confidence intervals of the loess regression lines overlap for the entirety of the contour. This is highly suggestive of the claim that the pitch of underlying and derived Low tones is, for all intents and purposes, identical. What is more, even if the small (∼5 Hz) difference between them around step 4 in Figure 3.7 is consistent, this difference is highly unlikely to be large enough to trigger or block a neutralization process like low tone spread. It appears, then, that there are not distinct coarticulatory pitch-lowering pressures when the preceding Low is derived rather than underlying. As a result, the lack of application of low tone spread in the case of derived Lows cannot be said to result from a physical difference between derived and underlying Lows.

However, there are other possible ways that a direct phonetics account might be augmented to account for this opacity. One such approach is to make use of paradigm uniformity effects (Steriade, 2000), which, broadly speaking, refer to the tendency for members of a morphological paradigm to share the same value for some feature. Somewhat similarly to output-output faithfulness constraints (Benua, 1995), paradigm uniformity constraints can be used to enforce identity between one member of a morphological

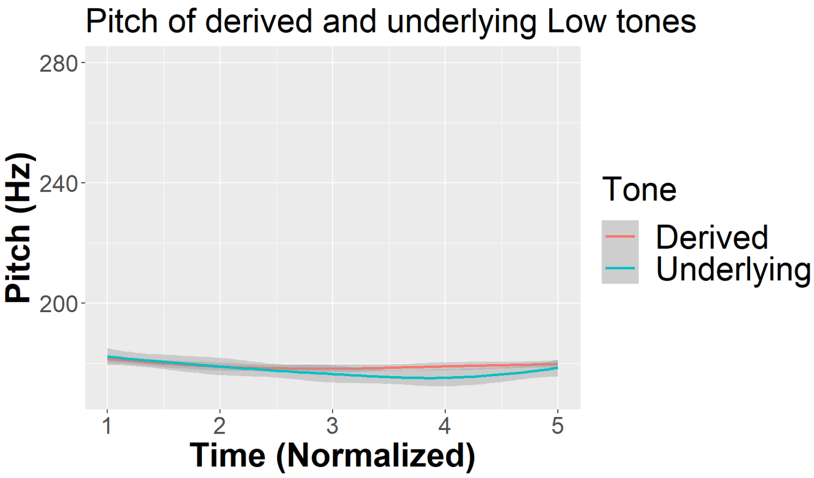


Figure 3.7: Pitch (Hz) of root-final derived and underlying L tones before an H-initial adjective (23 Underlying, 59 Derived).

paradigm and another member of that same paradigm. For example, consider the two words ‘condemn’ ([kʌnˈdɛm]) and ‘condemnation’ ([ˌkɑndəmˈneɪʃən]). Both share the same root, but the root in ‘condemnation’ ends in an [n] ([kɑndəmn + eɪʃən]) while the root in ‘condemn’ ends in an [m]. If the shared root is /kɑndəmn/, then this difference can be understood as being the result of word-final deletion of an [n] when it follows [m] (Borowsky, 1986).

(31) Word-final [n]-deletion

$/n/ \rightarrow \emptyset$ / m_#

This rule will apply to the root when it surfaces alone, but not when it is followed by a suffix, since the conditioning environment for deletion is not met:

(32) Application

/kʌndɛmn/ → [kʌnˈdɛm]

(33) Non-application

/kʌndɛmn + eɪʃən/ → [kɑndəmneɪʃən][6]

In this way, we can understand the presence and absence of the lack of root-final [n] as being the result of a general phonological process. This can be modeled using OT constraints, as well:

- *NN#: Assign one violation for sequence of two nasal consonants at the end of a word.
- MAX: Assign one violation for every segment in the input that does not have a correspondent in the output.

(34)

/kʌndɛmn/	*NN#	MAX
☞ a. [kʌndɛm]		*
b. [kʌndɛmn]	*!	

However, an issue arises when we consider the word ‘condemning’ ([kʌnˈdɛmɪŋ]). In this form, the underlying root-final [n] is deleted, even though it is followed by a suffix. This an apparent case of overapplication opacity: Here, the phonological rule of

---

[6] I do not analyze the variation in vowel quality, which is due to vowel reduction in unstressed syllables. Instead, I simply assume the underlying form /kʌndɛmn/.

word-final [n]-deletion has applied despite its conditioning environment not being met in /kʌnˈdɛmn + ɪŋ/. Under the current constraint ranking, we incorrectly predict that Candidate A will win, despite the fact that Candidate B is the real-word output:

(35)

/kʌndɛmn + ɪŋ/	*NN#	MAX
☹ a. [kʌndɛmnɪŋ]		
b. [kʌndɛmɪŋ]		∗!

One way to derive the overapplication of word-final [n]-deletion is to include a higher-ranked constraint enforcing similarity between a the form of a stem when it is modified by some affixes (i.e., ‘condemning’) and its base form, when it occurs as a free word (‘condemn’). In this case, the constraint might be something like the following:

- Paradigm Uniformity (PU): Assign one violation for every segment in a stem S that is not present in S when S occurs as a free word.

Ranking this constraint above MAX leads to the correct optimal candidate being chosen:

(36)

/kʌndɛmn + ɪŋ/	PU	*NN#	MAX
a. [kʌndɛmnɪŋ]	∗!		
☞ b. [kʌndɛmɪŋ]			∗

The final step is to restrict the PU constraint to certain morphological contexts, such as those involving inflectional affixes, in which case it will not apply to /kɑndəmn + eɪʃən/, which involves a derivational suffix. This restricted constraint will be called PU$^{\text{INFL}}$.

(37)

/kʌnˈdɛmn + eɪʃən/	PU$_{\text{INFL}}$	*NN#	MAX
☞ a. [kɑndəmneɪʃən]			
b. [kɑndəmeɪʃən]			*!

In this way, paradigm uniformity constraints can be utilized to derive opaque interactions without making recourse to abstract, non-surface representations. A potential approach to handling the opacity involved in low tone spread in SMPM, then, would be through the use of paradigm uniformity constraints. However, as I will show below, these types of constraints do not account for the opacity because, in each case, the same members of the paradigm are involved.

The reason that the paradigm uniformity constraint outlined above works for English is precisely because ‘condemnation’ and ‘condemning’ are two different morphological constructions and can be argued to be members of two different paradigms. In SMPM, though, the situation is different: The two surface forms [káˀnũ̏] and [kǎˀnũ̏] (‘big’) are the same exact lexical item, and as a result, they are in the same exact paradigmatic relationship with the stand-alone form [káˀnũ̏]. That is, the two surface forms of the adjective in (38) and (39) morphologically identical. The form in (38) is phonologically modified but morphologically identical to the form in (39)—their meaning is the exact same, and they are in the same syntactic relationship with the preceding noun.

(38) kîǹı káˀnũ̀ → kîǹı kǎˀnũ̀
pig big
‘A big pig.’

(39) kòʰǒ káˀnũ̀ → kòʰò káˀnũ̀
snake big
‘A big snake.’

The fact that the same exact lexical item in the same exact syntactic configuration both undergoes and does not undergo low tone spread means that paradigm uniformity cannot be used to derive the opacity see above. To illustrate this, consider the paradigm uniformity constraint outlined below:

- PARADIGM UNIFORMITY (PU): Assign one violation for every tone in a stem S that is not present in S when S occurs as a free word.

This is because a paradigm uniformity constraint like the one given above would have to be ranked below the markedness constraint driving low tone spread in the case of an underlying Low tone, as shown below. This is because the adjective [kǎˀnũ̀] in the optimal candidate has a tone not present in its base form [káˀnũ̀], namely the initial Rise.[7]

---

[7] I am using categorical constraints and a simplified OT analysis here for ease of exposition. A direct phonetics account would necessarily use gradiently-defined constraints and candidates, and a more thorough OT analysis of low tone spread and rise flattening can be found in §3.2.3 and §4, respectively.

(40)

/kînì # káˀnũ̀/	*áˀ	PU	*SPREAD
☞ a. [kînì # kăˀnũ̀]		*	
b. [kînì # káˀnũ̀]	*!		*

However, the same paradigm uniformity constraint would have to be ranked *above* that the markedness constraint driving low tone spread in the case of a derived Low. If the ranking of PU and *V́ˀ in (41) were flipped to correspond to the ranking in (40), then we would incorrectly predict Candidate B as the output. Instead, the ranking below is required to correctly select Candidate A, the attested surface form.

(41)

/kòʰŏ káˀnũ̀/	*LH # H	PU	*áˀ	*SPREAD
☞ a. [kòʰò káˀnũ̀]			*	
b. [kòʰò kăˀnũ̀]		*!		*
c. [kòʰŏ káˀnũ̀]	*!		*	

Because a paradigm uniformity constraint cannot be used to enforce similarity between an adjective and its stand-alone form in the presence of a derived Low but not in the presence of an underlying Low, a direct phonetics account cannot appeal to paradigm uniformity in accounting for the opaque interaction between rise flattening and low tone spread. This pitfall demonstrates that the opacity is fatal for a direct phonetics account: The adjectives are in the same paradigmatic relationship with their base form regardless of whether the preceding Low is derived or underlying, so even an augmented direct phonetics framework cannot derive the opacity.

3.3.1.3 Interim review

So far in this section, I have argued that, despite its highly-specific and multiply-triggered nature, low tone spread in SMPM does not lend itself to a direct phonetics analysis that defines phonological units and constraints at the level of fine-grained phonetic detail. The main reason for this is the opaque interaction between rise flattening and low tone spread: When derived Low tones precede adjectives that typically undergo low tone spread, the process does not apply. This is in spite of the fact that the same coarticulatory pressures are present in each case. Even if a direct phonetics account were augmented with the usual opacity-deriving mechanism of paradigm uniformity, such a framework is still unable to derive the opacity because the adjectives in question are the same morphological items, regardless of whether the preceding Low tone is derived or underlying.

It is clear, then, that a direct phonetics account of low tone spread falls short restricting its application only to cases involving underlying Low tones. As I will argue in the following section, a phonological framework whose units of representation and constraints are defined at a level more coarse-grained than that of fine-grained phonetic detail is able to straightforwardly derive the alternation, and to block its application when the conditioning environment includes a derived Low, and that it does so while broadly capturing the process's phonetic grounding.

3.3.2 Indirect phonetics

An indirect phonetics account defines phonological units at a more abstract level than that of fine-grained phonetic detail, such as phonological features or segments. This means that the internal structure of that phonological unit (i.e., the phasing of modal and creaky voice in laryngealized vowels) is not directly reflected in the phonological unit's representation, but rather filled in at a later stage in the derivation. As I will show, it is this characteristic that allows this type of approach to adequately account for low tone spread while nonetheless broadly capturing its phonetic grounding.

3.3.2.1 An indirect phonetics analysis

The phonological constraints necessary for deriving low tone spread are given below: The first penalizes High-toned laryngealized vowels, and the second penalizes High-toned low vowels.

- *v́ˀ: Multiply constraint weight by number of High-toned laryngealized vowels.
- *á: Multiply constraint weight by number of High-toned low vowels.

These two markedness constraints participate in a 'ganging-up effect' against three vanilla faithfulness constraints (Yip, 2002).

- *Spread: Assign one violation for every tone linked to two or more TBUs in the output.

- Ident[t]: Assign one violation for every tone whose value α is realized as β in the output.
- Max[t]: Assign one violation for every tone in the input that is not present in the output.

The faithfulness constraints are all more highly weighted than the markedness constraints, meaning that when only one of the markedness constraints is violated, the fully faithful candidate is the optimal candidate. This can be seen in the following tableau, which shows that low tone spread does not apply to a High-toned, laryngealized high vowel (42). Faith[t] subsumes Ident[t] and Max[t], since neither of these are violated by a winning candidate here. In (43), the winning candidate is the fully faithful candidate, Candidate A. This is because the violation incurred by having a laryngealized vowel not linked to a Low tone is lower than the violations incurred by Candidates B and C for spreading and changing the value of a tone, respectively.[8]

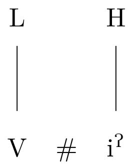
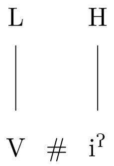
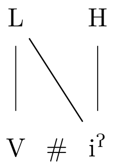
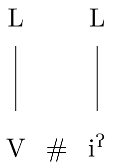
(42) **No low tone spread**

kînì ⁿtsíˀi
pig blue
'A blue pig.'

---

[8] This analysis and constraint formulation means that a candidate like Candidate B, which has an LH tone, does not violate *v́ˀ, even though the LH tone contains an H. This is because the constraint only penalizes High-toned laryngealized vowels, not laryngealized vowels with a Low-High tone. This point is unintuitive since the Low-High tone contains a High tone, but does not not trigger violations of *v́ˀ or *á. A more intuitive constraint would be something like vˀ[low], which would penalize any laryngealized vowel not linked to a Low tone. However, because the inductive grounding approach described later in the chapter requires constraints to be defined in negative terms (that is, to penalize specific configurations) rather than positive terms (that is, to penalize all configurations but a specific one), I use the constraint definitions above. Note, though, that this low tone spread presents a general problem for an analysis using negatively-defined markedness constraints: They must penalize a laryngealized vowel when it has only a High tone, but not when it has a Low-High tone.

(43)

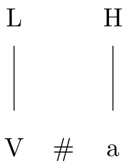
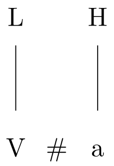
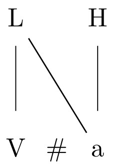
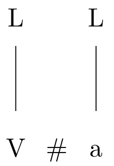
	FAITH[T] WT=10	*SPREAD WT=5	*v́^{ʔ} WT=3	*á wt=3	**Harmony** **score**
☞ a. 			-1		-3
b. 		-1			-5
c. 	-1				-10

Given that both of the markedness constraints have the same weight, this weighting also guarantees that low tone spread does not apply to high-toned low vowels that are not laryngealized, as in (44).

(44) **No low tone spread**

kînì k^{w}ã́ã̀
pig yellow
'A yellow pig.'

(45)

	FAITH[T] WT=10	*SPREAD WT=5	*V́ˀ WT=3	*á wt=3	**Harmony score**
☞ a. 				-1	-3
b. 		-1			-5
c. 	-1				-10

However, in a case where both of the markedness constraints are violated at the same time, the combination of their violations means that the fully faithful candidate has a lower harmony score than Candidate B, which violates *SPREAD. Candidate C, which violates FAITH[T] is still out, since its harmony score is the lowest.[9]

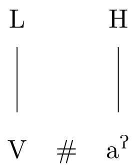
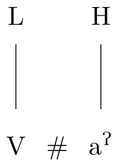
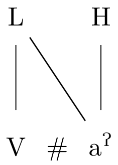
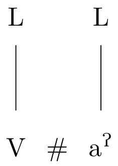
---

[9] Note that, under this constraint ranking, an input like *káˀ nũ̀* ('big') would undergo leftward low tone spread to become *kâˀ nũ̀*. This requires either the existence of separate *SPREAD constraints for each direction of spreading, or that a constraint penalizing falling tones on laryngealized vowels outweighs a constraint penalizing rises on laryngealized vowels.

(46) **Low tone spread**

ɾà sã̌ˀã́
3m Spanish
‘A Spanish man.’

(47)

	Faith[t] wt=10	*Spread wt=5	*v́ˀ wt=3	*á wt=3	**Harmony** **score**
a. 			-1	-1	-6
☞ b. 		-1			-5
c. 	-1				-10

Finally, the current constraint formulation and weighting predicts that preceding Mid tones could spread to ameliorate violations of *v́ˀ and *á. I assume that this is blocked by a high-ranking constraint prohibiting Mid-to-High contour tones.

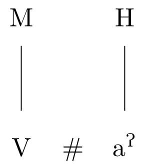
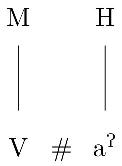
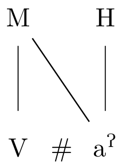
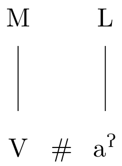
- *mh: Multiply constraint weight by the number of Mid-to-High contour tones in the output.

These contour tones are absent from SMPM's phonemic inventory, so a constraint on them is merited based on phonotactic restrictions. Alternatively, Mid tone might be phonologically unspecified in SMPM (it is not the target or trigger of any phonological alternations and is absent from the list of grammatical tones), in which case it might independently be blocked from spreading. Including this constraint allows the grammar to derive the non-application of tone spreading in (48) below.

(48) **No low tone spread**

le^{h}so ká^{ʔ}nũ̀
rabbit big
'A big rabbit.'

(49)

	FAITH[T] WT=10	*MH WT=5	*SPREAD WT=5	*V́^{ʔ} WT=3	*á wt=3	**Harmony** **score**
☞ a. 				-1	-1	-6
b. 		-1	-1			-10
c. 	-1					-10

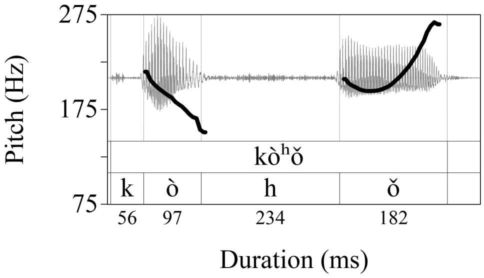
The process of low tone spread, then, can be straightforwardly analyzed in a framework whose units of representation and constraints are defined at a level more coarse-grained than that of fine-grained phonetic detail. As hinted at earlier, though, this type of analysis is advantageous because it is also able to account for the opaque interaction between low tone spread and rise flattening.

3.3.2.2 Deriving opacity

Recall that word-final Rising tones flatten to a Low tone when the next tone is High (50). However, Low tones derived via this sandhi process do not trigger low tone spread (51). This opacity proved to be a downfall of a direct phonetics analysis, since the physical coarticulatory pressures argued to be driving low tone spread are present in (51), but the process does not occur. Additionally, the usual augmentation via paradigm uniformity does not save this account, since the same lexical items both undergo low tone spread and do not undergo low tone spread. Because there is no way to tie the lexical item that does not undergo low tone spread in (51) to the base form, while not tying that same lexical item to the same base form in cases where it does undergo spread, it is not possible to derive the opacity via correspondence relations.

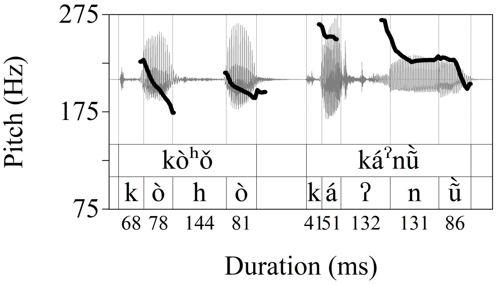
(50) **No rise flattening**

kò^{h}ǒ
snake
'A snake.'



(51) **Rise flattening, no low tone spread**

kò^{h}ò kà^{ʔ}nũ̀
snake big
'A big snake.'



While a direct phonetics account struggles to derive this opaque interaction, an indirect phonetics has recourse to derivation, which allows it to do so relatively straightforwardly. What is necessary is to derive low tone spread at an earlier phonological level than rise flattening. If low tone spread occurs before rise flattening, then at the point at which low tone spread occurs, rise flattening will not have created the opportunity for low tone spread to apply. This is very easily modeled via rule ordering, as shown below:

(52) Rule ordering

UR		**Low Tone Spread**		**Rise Flattening**		**Surface Form**
/kòʰǒ káˀnũ̀/	→	[kòʰǒ káˀnũ̀]	→	[kòʰò káˀnũ̀]	→	[kòʰò káˀnũ̀]

While type of interaction is easily modeled by rule ordering, it is more difficult in classic, parallel OT (McCarthy, 2007). In fact, in parallel OT, this opacity should be impossible to derive. The reason for this is that, if a constraint ranking derives low tone spread (/v̀ + áˀ/ → [v̀ + ǎˀ]) by preferring the output candidate [v̀ + ǎˀ] over the output candidate [v̀ + áˀ], then the dispreference for the output candidate [v̀ + áˀ] holds whether the initial Low is underlying or derived. As a result, there is no case in which the preference of output arguments changes. To illustrate this, I will walk through an analysis of rise flattening and show that, when the constraint ranking driving rise flattening is considered in parallel with the constraint ranking driving low tone spread, there is no way to block low tone spread from applying to derived Lows. The constraints involved in rise flattening are, once again, relative uncontroversial tonal markedness constraints (Yip, 2002).

- OCP[H]: Multiply constraint weight by the number of adjacent High tones.
- *CONTOUR: Multiply constraint weight by number of moras linked to two distinct tones.

These markedness constraints are, like the others, not strong enough to trigger changes on their own. For example, (53) shows that MAX[T] and IDENT[T] outweigh

OCP[H]—the output violates OCP[H] rather than modifying or deleting any of the underlying tones. Likewise, (54) shows that MAX[T] and IDENT[T] both outrank *CONTOUR, since the output violates *CONTOUR instead of modifying or deleting any of the underlying tones.

(53) **No tone change**

léló tsʲáʰku
skunk alive
‘A live skunk.’

(54) **No tone change**

kòʰǒ loˀo
snake small
‘A small snake.’

However, when both OCP[H] and *CONTOUR are violated at the same time, the outcome is different—the Low-High Rise becomes a simple Low tone. This can be seen below:

(55) /nì-ⁿtsìʰkù tsʲũ̀ʰũ̌ léló/ → nì-ⁿtsìʰkù tsʲũ̀ʰũ̀ léló
COMPL-chase turkey skunk
‘The turkey chased the skunk.’

This, then appears to be another ‘ganging-up effect,’ whereby violation of a higher-weighted constraint is preferable to violations of two lower-weighted markedness constraints. Note that I analyze the derivation of a Low tone from a Low-High Rise to be the result of deletion of the High member of the Low-High Rise.

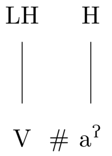
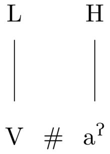
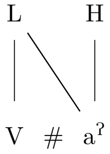
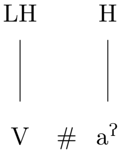
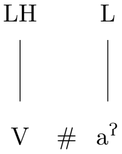
(56) Derivation of rise flattening

<table>
<tr><th>LH H
| |
V # V</th><th>IDENT[T]
WT=15</th><th>MAX[T]
WT=8</th><th>OCP[H]
WT=5</th><th>*CONTOUR
WT=5</th><th>Harmony
score</th></tr>
<tr><td>☞ a. L H
| |
V # V</td><td></td><td>-1</td><td></td><td></td><td>-8</td></tr>
<tr><td>b. LH H
| |
V # V</td><td></td><td></td><td>-1</td><td>-1</td><td>-10</td></tr>
<tr><td>c. H H
| |
V # V</td><td></td><td>-1</td><td>-1</td><td></td><td>-13</td></tr>
<tr><td>d. LH L
| |
V # V</td><td>-1</td><td></td><td></td><td>-1</td><td>-15</td></tr>
</table>

Now that the phonological derivation of rise flattening has been explored, it is possible to illustrate that a parallel account using the the constraint weightings to derive both low tone spread *and* rise flattening predict that the two will interact transparently. This is because a violation of MAX[T] is preferable to violating both OCP[H] and *CONTOUR, and a violation of *SPREAD is preferable to violating both *V̌^{ʔ} and *á. It follows, then, that violating both MAX[T] and *SPREAD is preferable to violating all four of the markedness constraints. This leads the analysis to incorrectly predict

Candidate B as the winner, though Candidate A is the attested output.[10]

(57)

	IDENT[T] WT=15	MAX[T] WT=8	*SPREAD WT=5	OCP[H] WT=5	*CONTOUR WT=5	*$\acute{V}^{ʔ}$ WT=3	*á wt=3	**Harmony score**
☹ a. 		-1				-1	-1	-14
☞ b. 		-1	-1					-13
c. 				-1	-1	-1	-1	-16
d. 	-1				-1			-20

The fact that parallel evaluation does not work for the interaction between rise flattening and low tone spread shows that what is needed is a model of OT that allows for variable constraint weighting depending on the level of derivation.[11] This can be

---

[10] Note that this pattern would also hold in classic OT and is unique to Harmonic Grammar. The reason is that, once a constraint ranking triggers triggers an alternation in a given context, it must always do so.

[11] Harmonic Serialism (McCarthy, 2000), which uses a constant constraint ranking but allows for only one change at a time between input and output, will not work here, either. Since the constraint ranking

accomplished in a framework like Stratal OT (Bermudez-Otero, 1999; Kiparsky, 2000), where distinct levels of derivation are associated with distinct subgrammars. These distinct levels of computation are often associated with the lexical and post-lexical divide in phonology, with lexical phonology referring broadly to within-word phonology and post-lexical phonology referring broadly to between-word phonology (Kiparsky, 1982). The fact that both low tone spread and rise flattening are between-words processes and therefore post-lexical is somewhat problematic, since it requires positing two distinct post-lexical grammars, though similar proposals have been made elsewhere (Kaisse, 1985; Sande et al., 2020).[12] However, there is some evidence that low tone spread applies in smaller (and thus, closer to the word-level) domains than rise flattening. Specifically, rise flattening applies when the conditioning environment is split between a subject and object DP (58), but low tone spread does not (59). Under the assumption that the prosodic boundary between subject and object DP is larger than the prosodic boundary between noun and adjective, this means that rise flattening can apply across larger boundaries than low tone spread.

(58) nì-^{n}tsì^{h}kù ts^{j}ũ̀^{h}ũ̀ léló
COMPL-chase turkey skunk
'The turkey chased the skunk.'

(59) nì-^{n}tsì^{h}kù kînì lá^{ʔ}n=ì
COMPL-chase snake old=1SG
'The pig chased my grandpa.'

So, rise flattening applies across larger prosodic domains than low tone spread.

---

is the same, when rise flattening applies, that constraint ranking should trigger low tone spread as the next intput-output change. Augmentation by incorporating faithfulness constraints to the underlying representation (as opposed to the input at that stage of the derivation), such as that proposed by Hauser et al. (2016), will not work here for the same reasons that paradigm uniformity does not work for the direct phonetics analysis—the same exact paradigmatic relationship is involved in both cases.

[12] Another possible analysis that would not require positing two post-lexical strata would be to analyze low tone spread as an instance of precompiled phrasal phonology (Hayes, 1990), in which case low tone spread would apply in the lexical stratum in a specific instantiation frame. Low tone spread is a good candidate for analysis in precompiled phrasal phonology because it appears to be sensitive to syntactic category (it only seems to apply in N-Adj pairs) and is derivationally earlier than other post-lexical rules.

This evidence is suggestive that the domain of application of low tone spread is smaller than that of rise flattening. If these two levels of derivation are associated with different strata, then the opaque interaction between low tone spread and rise flattening can be modeled. This can be done by more highly weighting MAX[T] in one stratum, and more highly weighting *SPREAD in the subsequent stratum.[13] Highly weighting MAX[T] in the first stratum results in rise flattening not applying. So, even though *SPREAD, *V́ˀ, and *á are weighted such that they would trigger low tone spread in its conditioning environment (see (49)), rise flattening is blocked and thus does not trigger low tone spread.

---

[13] There is independent evidence from the verbal domain that *SPREAD is lowly ranked in lexical phonology. Completive Low tone maps onto the leftmost syllable of a verb, whether that syllable is a prefix or the first syllable of the verb root. When Completive Low tone maps onto a prefix that is followed by a High-toned vowel, Completive Low tone spreads to the first syllable of the verb root, creating a Low-High Rise. However, this process applies to a wider range of High-toned vowels than low tone spread.

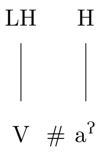
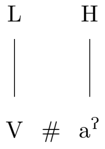
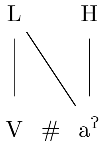
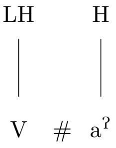
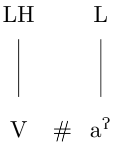
(i) sá-káʰsũ̀
CONT.CAUSE-toast
'Is toasting'

(ii) ⁿdátṹˀũ
CONT.converse
'Is conversing.'

(iii) sà-kǎʰsũ̀
COMPL.CAUSE-toast
'Toasted'

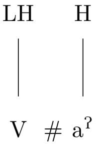
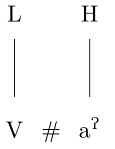
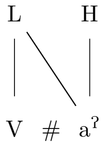
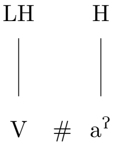
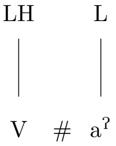
(iv) ⁿdàtũ̌ˀũ
COMPL.converse
'Conversed.'

(60) Stratum 1: Blocking rise flattening

	IDENT[T] WT=15	MAX[T] WT=15	*SPREAD WT=5	OCP[H] WT=5	*CONTOUR WT=5	*V́^{ʔ} WT=3	*á wt=3	**Harmony score**
a. 		-1				-1	-1	-21
b. 		-1	-1					-20
☞ c. 				-1	-1	-1	-1	-16
d. 	-1				-1			-20

Then, in stratum 2, the weight of MAX[T] is lowered, and the weight of *SPREAD is raised. This means that, though rise flattening applies, low tone spread is blocked. The result is that Candidate A, the attested surface form, is the candidate.

(61) Stratum 2: Blocking low tone spread

	IDENT[T] WT=15	MAX[T] WT=8	*SPREAD WT=15	OCP[H] WT=5	*CONTOUR WT=5	*V́ˀ WT=3	*á wt=3	**Harmony score**
☞ a. 		-1				-1	-1	-14
b. 		-1	-1					-23
c. 				-1	-1	-1	-1	-16
d. 	-1				-1			-20

So, by making recourse to derivation, an indirect phonetics analysis is able both to derive low tone spread in the cases in which it applies *and* to block its application when the preceding Low tone is derived by rise flattening. Though the analysis involves the complication of requiring multiple post-lexical strata, the indirect phonetics analysis is still superior to the direct phonetics analysis. The reason for this is that a direct phonetics cannot capture the opaque interaction between the two tone sandhi processes even if it is augmented with the usual mechanisms for deriving opacity in these types of frameworks. An indirect phonetics account, on the other hands, *can* be augmented to derive the opaque interaction. Though this augmentation makes the analysis more complicated, it is at least somewhat motivated by the fact that rise flattening applies across larger domains than low tone spread. In the end, a framework that has been augmented to account for a special case is preferable to a theory that cannot be augmented to account for the same special case.

3.3.2.3 Interim review

In this section, I have argued that low tone spread is problematic for direct phonetics frameworks, despite its highly-specific, multiply-triggered nature. The reason for this is that, in these frameworks, it is the physical, coarticulatory pressures that trigger alternations. A logical extension of this fact is that whenever the physical coarticulatory pressures that trigger low tone spread are present, the process should apply. However, this is not always the case: When the Low tone trigger of low tone spread is derived rather than underlying, low tone spread does not apply. I showed that workarounds in a direct phonetics account, such as appealing to differences in pitch level between derived and underlying Lows, or using paradigm uniformity to account for opacity, do not work. Instead, an indirect phonetics framework whose units of representation and constraints are defined at a level more coarse-grained than that of phonetic detail is able to easily derive the process of low tone spread and is also able to account for its opaque interaction with rise flattening if implemented in a derivational framework like Stratal OT.

However, indirect phonetics frameworks like the one outlined above, which derive multiply-triggered phonological alternations via the interaction of phonological constraints, have been criticized for not transparently reflecting the coarticulation-driven nature of these types of alternations (Lionnet, 2016). In the following section, I show that the constraints involved in the analysis of low tone spread can be tied to learners' phonetic experience speaking and listening to their language through a process of phonological constraint induction outlined in Hayes (1999). As a result, though the phonological constraints do not directly model coarticulation, they do reflect it at a broader level, since their introduction into the grammar during the process of language learning was driven by learners' experience with coarticulation.

3.4 Phonetic grounding

As outlined in §3.1.3, the phonetic grounding of this alternation seems relatively straightforward: Both laryngealization and low vowel quality lower F0, so it makes sense that a laryngealized low vowel would be more likely to be a target of low tone spread. However, the process cannot be driven directly by the coarticulatory pressures that seem to underlie it, since there are cases in which these pressures are present but low tone spread does not occur. This indirect relationship between coarticulatory pressures and phonological processes that reflect them is consistent with a process of phonological constraint induction like those outlined in Hayes (1999), Smith (2004), and Flack (2007), though the proposal in this section is given in terms of the proposal in Hayes

(1999). In this type of model, language learners posit phonological constraints based on their articulatory and perceptual experience, with a pressure toward simple constraints, and then incorporate those constraints in their phonological grammar if they do a good job of ruling out articulatorily and/or perceptually difficult configurations and allowing easy configurations

This approach has two crucial ingredients: The first is a phonetic 'map,' and the second is a process of constraint evaluation. In this approach, language learners use their experience with the processes of articulation and perception in their language to create a phonetic map that details the difficulty of producing or perceiving particular configurations of sounds. They also have access to the set of phonological primitives in their language, and they use these features and combinations of them to determine the space of logically-possible phonological constraints. For example, given the phonological features [+/- voice] and [+/- nasal], the logically-possible phonological constraints are *[+ voice, -nasal], *[+ voice, + nasal], *[-voice, -nasal], *[-voice, +nasal]. Once they have determined the possible constraint space, they use their phonetic map to evaluate potential phonological constraints based on how good those constraints are at (1) ruling out articulatorily or perceptually difficult configurations, and (2) allowing articulatorily or perceptually easy configurations. Constraints are then compared to each other, with direct comparisons being between constraints of similar complexity. Those that are most effective within a complexity-based group of constraints are introduced into the phonological grammar, where they are ranked by the same process by which learners eventually arrive at their language's constraint ranking.

So, there are three crucial steps in an inductive grounding approach: The first is the generation of a phonetic map, the second is the generation of phonological constraints, and the third is the evaluation of the effectiveness of those constraints relative to each other. In this section, I will describe how this process might be applied to the generation of the markedness constraints used in the indirect phonetics analysis of low tone spread, starting with the constraint *v́^{ʔ} (assign one violation for every High-toned laryngealized vowel). As I will argue, this approach provides a link between phonological and their phonetic bases without referring directly to the physical structures involved.

The first step is the generation of a phonetic map that details the difficulty or ease with which a given configuration of phonological units is produced or perceived. The formulation of this type of map is highly complex and outside the scope of this dissertation, but in order to have some numerical measure of phonetic difficulty, I will make a relatively arbitrary and speculative connection between acoustic measures and articulatory difficulty. Specifically, I will assume that producing a modal-then-creaky sequence is more difficult than maintaining modal voice, and that large pitch falls are more articulatorily difficult than smaller pitch falls. It seems reasonable to assume that phasing two phonation types involves more articulatory effort than maintaining modal voice. Also, the assumption that larger pitch falls require more articulatory effort than smaller pitch falls might have a basis in prior research, though the connection is perhaps tenuous: The speech of pitch change varies linearly with the size of pitch change (Xu and Sun, 2002). If faster changes in pitch require more articulatory effort than slower changes in pitch, than the size of a pitch fall might be at least somewhat related to a measure of articulatory difficulty. In any case, these assumptions allow us to map pitch contours to measures of articulatory difficulty in a one-to-one manner, such that pitch changes are more difficult the larger they are. This is important because, as mentioned earlier, laryngealization causes pitch lowering. However, the degree of pitch lowering is not identical in all cases. Instead, pitch falls by different amounts depending on the tone and phonation type of the vowel it is realized on.

The data used to determine the degree of pitch fall for tones in laryngealized vowels was gathered by asking Consultant 1 to produce target words in the carrier phrase in (62), with the target word either modal or laryngealized.

(62) kã̂ˀ=ĩ̀ ___ βitsĩ̀
POT.say=1SG ___ now
‘I will saw ___ now.’

There were 89 productions of laryngealized words, and 66 productions of modal words. For laryngealized vowels, the pitch of the first vowel was measured into the onset of laryngealization, and the pitch of modal vowels was measured until the end of the vowel. Examples of the spliced portion of laryngealized vowel (left) and modal vowel (right) are given in Figure 3.8.

A Praat script extracted the difference between the maximum and minimum pitch. This value was used as the degree of pitch fall. The resulting values are illustrated in Figure 3.9.

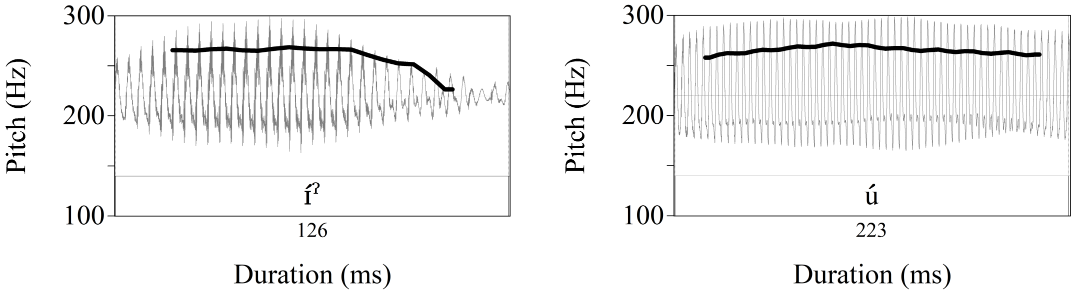


Figure 3.8: Representative examples of analyzed portions of laryngealized (left) and modal (right) vowels.

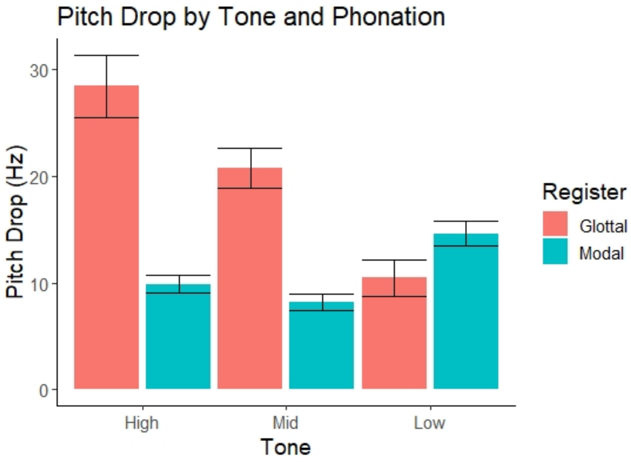


Figure 3.9: Pitch drop for High, Mid, and Low tones by phonation type (34 High Glottal, 28 Mid Glottal, 27 Low Glottal, 31 High Modal, 19 Mid Modal, 16 Low Modal)

As can be seen above, the pitch of laryngealized vowels with a High or Mid tone consistently falls more than that of modal vowels with a High or Mid tone. Curiously, the pitch of Low tones on modal vowels falls more steeply than it does in laryngealized vowels, and also more steeply than the pitch of High or Mid tones on modal vowels. This might be indicative that Low tones in modal vowels are consistently produced as slightly falling. That being said, the main point of interest for our present purposes is the pitch of laryngealized vowels. As the table above shows, degree of pitch fall for laryngealized vowels decreases monotonically with tone level: The pitch fall for High-toned laryngealized vowels is greatest, followed by the pitch fall for Mid-toned laryngealized vowels, and then the smallest fall is seen in Low-toned laryngealized vowels.[14] Using the highly simplified heuristic of converting the degree of pitch excursion directly into some measure of articulatory difficulty, this means that the production of a High tone on a laryngealized vowel involves more articulatory difficulty than the production of a Mid or a Low tone on a laryngealized vowel. Additionally, the production of a High tone on a laryngealized vowel involves more articulatory difficulty than the production of a High tone on a modal vowel, with the reasoning being that phasing modal and non-modal voice is requires more articulatory effort than maintaining modal voice.

The next step is constraint generation: This is done by freely combining the phonological primitives in a learner's language, and then grouping them into constraint 'neighborhoods' based on their featural makeup. For any given constraint, its constraint neighborhood is made up of all those constraints that can be derived from it by changing one primitive phonological element, either by addition or deletion of one of these elements, or by changing a feature value. Constraints are also organized by whether or not they are more complex than the target constraint (if the target constraint's structural description is properly included in a neighbor constraint, that neighbor constraint is more complex). The constraint neighborhoods for the phonological constraints used in the direct phonetics analysis are given in Table 3.6.

---

[14] It might be the case that the smaller fall in Low-toned laryngealized vowels is the result of a floor effect. Impressionistically, Low-toned laryngealized vowels start with a lower pitch than other Low tones.

Note that, though there are featural analyses of tonal systems (see, e.g., Yip, 2002:Ch. 3), I consider the tonal units High, Mid, and Low to be the phonological primitives of SMPM's tone system. This means that neighbor constraints for tonal constraints on tonal specification may be determined by deleting a tone (i.e., *v́ˀ and *vˀ), adding a tone (i.e., *v́ˀ and *v̌ˀ),[15] or changing from one level tone to another (i.e., *v́ˀ and *v̀ˀ).

Constraint	Neighbor constraints	
	of equal or lesser complexity	**of greater complexity**
*v́ˀ	*vˀ, *v̀ˀ, *v́	*v̂ˀ, *v̌ˀ
*v́[+LOW]	*v́[-LOW], *v[+LOW], *v̀[+LOW]	*v̂[+LOW], *v̌[+LOW]

Table 3.6: Constraint neighborhoods for indirect phonetics constraints

Once a constraint neighborhood has been determined, the phonetic effectiveness of the target constraint is compared with the phonetic effectiveness of all of its neighbor constraints of equal or lesser complexity. As noted in Table 3.7, High tones on laryngealized vowels have the largest pitch drop out of any level tone. Under the assumption that large pitch changes are more difficult than small pitch changes, this means that *v́ˀ rules out configurations that are more difficult than those ruled out by *vˀ and *v̀ˀ.

---

[15] I analyze contour tones as being made up of two level tones (Leben, 1973; Goldsmith, 1976). There are various typological and SMPM-internal reasons for this, including that H- and HL- marked vowels both trigger flattening of a preceding LH Rising tone. This joint behavior makes sense when one considers both as starting with a H tone, since we can appeal to first member of the tone in each case being H, but it requires more stipulation if one considers High and Falling tones to be completely separate tones.

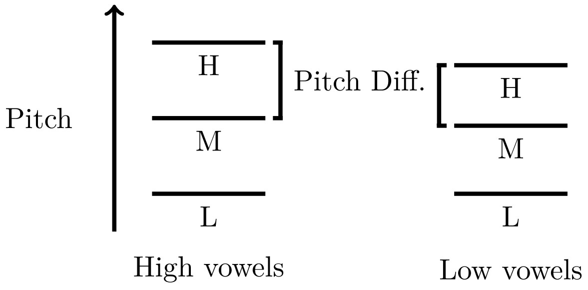
Because rising tones and falling tones contain two level tones each, I assume that *v̌ˀ, *v̂ˀ are *more* complex than *v́ˀ because they involve the addition of another tone. Additionally, I do not include *v́ʰ in this inventory, since the change from vˀ to vʰ likely involves both a change in the [+/- constricted glottis] feature as well as the [+/- spread glottis] feature.

High	~28 Hz
Mid	~21 Hz
Low	~10 Hz

Table 3.7: Average pitch falls on laryngealized vowels by tone (34 High, 28 Mid, 27 Low)

Additionally, since laryngealized vowels involve phasing of phonation types, *v́^{ʔ} rules out configurations more difficult than those ruled out by *v́. So, it appears that *v́^{ʔ} is phonetically grounded, since it rules out configurations that are more difficult than those ruled out by its neighbor constraints of equal or lesser complexity.

The other constraint used in the phonological analysis, *v́[+LOW], is likely also phonetically grounded, but potentially on separate grounds. In SMPM, High-toned low vowels have lower pitch than High-toned high vowels (Figure 3.3). Intrinsic F0 differences are also greatest at the high end of the pitch range, and weak or non-existent at the low end (Whalen and Levitt, 1995). So, if High-toned low vowels have a lowered pitch but Mid-toned have a pitch that is lowered to a lesser extent, then it is likely that the distance between High and Mid tones on low vowels is smaller than the distance between High and Mid tones on non-low vowels, as illustrated below:



If this is the case, then then under the naive assumption that the acoustic distances here are representative of perceptual distance, it is possible that the pitch of High-toned low vowels is somewhat confusable with the pitch of Mid-toned low vowels by virtue of not being as far away along the dimension of pitch. If this is the case, then *v́[+LOW] might be said to rule out a configuration that is perceptually difficult, since it is more confusable with another category. Under this view, then, *v́[+LOW] would be penalize structures that are more perceptually confusable with other categories, and the other constraints on low vowels (*v[+LOW], which penalizes Mid-toned low vowels, and *v̀[+LOW], which penalizes Low-toned low vowels) penalize less marked configurations.[16] Finally, *v́[-LOW] does not penalize marked structures, since F0 on non-low vowels is higher. Using this line of reasoning, then, the constraint *v́[+LOW] is also phonetically grounded, since it is more effective in ruling out difficult configurations than its neighbor constraints of equal or lesser complexity.[17]

So, the phonological constraints motivating the application of low tone spread in SMPM can likely be posited by a learner using a process of phonological constraint induction. The learner, having experience with the articulatory and perceptual difficulty of certain configurations, evaluated potential phonological constraints and promotes those that are more effective than their neighbors of equal or lesser complexity into the phonological grammar. As I have shown here, this process could plausibly lead to the

---

[16] This point assumes that, since intrinsic F0 effects on Mid tones should be smaller, Mid-toned low vowels are more easily-recognizable as Mid-toned than High-toned low vowels are recognizable as High-toned, meaning *v́[+LOW] rules out a more perceptually difficult configuration than *v[+LOW].

[17] Another possibility might be that it is more articulatorily difficult to produce a High tone on low vowels than it is on high vowels. This might be possible under the 'tongue-pull' hypothesis of intrinsic F0, which roughly states that in higher vowels, the tongue pulls on the hyoid bone, which causes an increase in stiffness of the vocal folds and thus an increase in F0 (Sapir, 1989). Under this approach, higher pitch is easier to produce on higher vowels, and by consequence it is more difficult to produce the same high pitch on low vowels.

two key markedness constraints used in the indirect phonetics analysis, namely *V́ˀ and *v́[+LOW]. The other constraints in the analysis, such as the faithfulness constraints or the phonotactic constraint *MH, are either learned through other constraint inducation processes or are innate, and their interaction with the induced markedness constraints gives rise to the pattern of low tone spread.

This process of constraint induction is able to provide the link between the coarse-grained phonological constraints and the fine-grained, coarticulatory pressures underlying them. That is, the constraints *v́ˀ and *v́[+LOW] are induced on the basis of learners' experience with coarticulatory pressures, so they broadly reflect the coarticulatory basis of the process of low tone spread. However, they do not directly encode these pressures in terms of fine-grained coarticulatory detail. Instead, they encode these pressures indirectly through constraints on configurations of abstract, phonological categories. Because of this coarse-grained level of granularity, induced constraints can interact with the rest of the grammar to produce results that are inconsistent with the results predicted by a phonological framework that makes no distinction between phonological and phonetic units of representation and constraints.

3.5 Review

In this section, I have shown that low tone spread is not amenable to an analysis in a direct phonetics framework that defines phonological units of representation and constraints in terms of fine-grained phonetic detail. The primary reason for this is that direct phonetics frameworks, by definition, rely directly on the physical structures involved in certain sound configurations to trigger alternations. However, whether or not the physical configuration that triggers low tone spread is underlying or derived determines whether or not the process applies. Even an augmented direct phonetics account is unable to account for this opacity, since it relies directly on the physical pressures involved and cannot make reference to their derivational history.

On the other hand, an indirect phonetics account is able to derive the process of low tone spread, and it can be augmented with strata to account for the opaque interaction between rise flattening and low tone spread. Additionally, the constraints involved in the analysis can plausibly be induced by SMPM learners on the basis of their articulatory and perceptual experience with different configurations of tone and phonation type. The framework of constraint induction allows for an analysis of low tone spread that (1) satisfactorily derives the process and its opaque interaction with rise flattening, and (2) reflects its basis in coarticulatory pressures, albeit indirectly.

3.6 Consequences

The fact that the highly-specific, multiply-triggered process of low tone spread is successfully modeled only by an indirect phonetics account is important because it highlights the type of phonological framework that is necessary for analysis of sound patterns in SMPM and other languages. Specifically, I would like to argue that any model of the sound patterns of SMPM must make use of at least two distinct levels of representation. One level corresponds broadly to what is usually thought of as phonology, and its units of representation are relatively coarse-grained (e.g., segments, features). Another corresponds to what is often labeled as phonetics, and its units of representation are much more fine-grained (e.g., Hz, dB). As noted earlier, this is by no means an claim original to this dissertation, since it has been made in a variety of works by a number of researchers (Keating, 1996; Hayes, 1999; Zsiga, 2000; Smith, 2005; Kingston, 2007; Cohn, 2007; Bermúdez-Otero, 2011, a.o.). However, it is unique in that its empirical basis is an alternation of the sort that is commonly used to argue *for* direct phonetics frameworks, namely a multiply-triggered process that appears to be the result of cumulative coarticulatory pressures. Even in a case where the coarticulatory pressures appear to be directly triggering the alternation, the alternation can be shown to be affected by abstract factors like derivational ordering.

If the claim that analysis of sound patterns requires multiple levels of description and analysis holds outside of SMPM as well—and I will argue later than it does—then it provides evidence against phonological models that posit no separation between phonology and phonetics, such as the direct phonetics model outlined earlier. Though they capture the phonetic grounding of phonological alternations, the fact that they do so by making direct reference to the physical phonetic structures involved means that they face significant difficulty in accounting for cases in which (1) the alternation has an opaque relationship with another process in the language, or (2) the alternation is phonetically 'unnatural,' meaning that it is either apparently unmotivated by any synchronic phonetic pressures (i.e., 'crazy rules,' Bach and Harms, 1972), or does not ameliorate the phonetic pressures apparently underlying it. Ruling out models of this type allows us to make a cut in the models of the phonology-phonetics interface that are able to account for patterns like the one described in this chapter. Specifically, frameworks to the right side of the dashed line in Figure 3.10, such as a unified model of phonology and phonetics, are unable to account for low tone spread in SMPM.

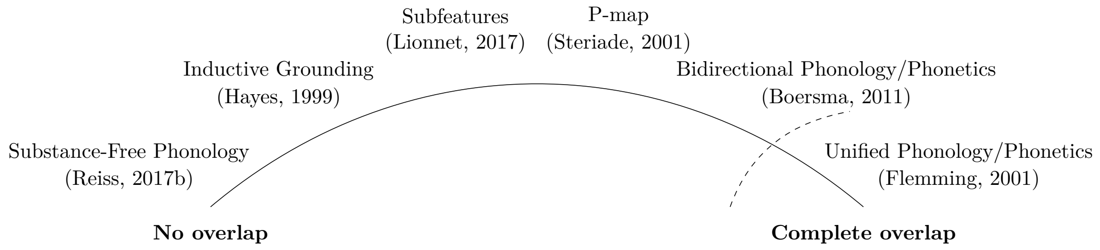


Figure 3.10: Continuum of frameworks of the phonology-phonetics interface

The existence of differing levels of granularity in the units of representation of a language's sounds necessitates two distinct levels of representation and computation, one which corresponds to what is often thought of phonology, and another which corresponds to what is often thought of as phonetics. These levels of description and analysis are distinct, but nonetheless connected. One way in which they interact with each other is likely though learning, where a process like inductive grounding is used to form pieces of the computational machinery of phonological grammar. The following chapter concerns another way in which the two systems are connected. Specifically, I will claim that, contrary to the predictions of a strictly feed-forward model of phonology and phonetics, the phonetic system is more than simply interpretational. That is, instead of simply converting phonological units of representation in phonetic units, it also has some influence over which phonological candidate makes it to the surface.

Chapter 4

Motivating a connection between phonology and phonetics

4.1 Introduction

In many modular theories of phonology, the relationship between the phonetics and phonology is essentially uni-directional: Phonology operates over phonological units, and the output of phonological computation is fed into the phonetic system, which converts phonological units into physical events that exist in time and space. This conversion process is sometimes portrayed as universal (Chomsky and Halle, 1968) and sometimes as language-specific (Zsiga, 2000), but it is often the case that the phonetic system is thought of as interpretational. That is, the job of phonetics is to convert phonological representations into phonetic ones, and vice versa, but its job stops there. In this view, there are no cases in which the process of converting phonological representations into phonetic ones fundamentally alters those phonological representations themselves. This is illustrated in Figure 4.1 below.

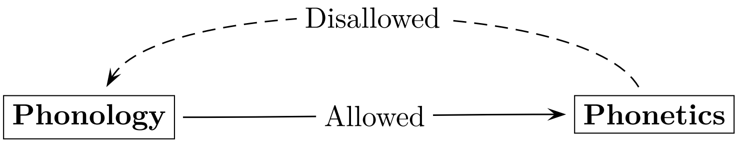


Figure 4.1: Uni-directional phonology-phonetics interface

In this chapter, I describe and analyze a pattern of reduction of laryngealized roots in SMPM that challenges the assumptions of uni-directional, modular accounts of the phonology-phonetics interface.

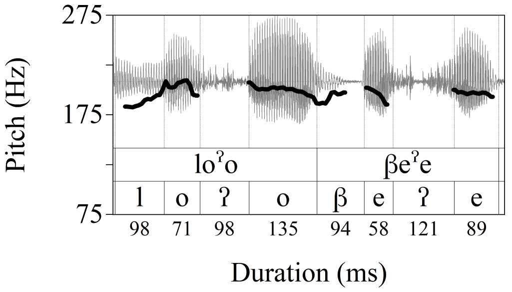
In SMPM, the acoustic correlates of laryngealization, some of which were already described in Chapters 2-3, can vary greatly. In some tokens, which I term *unreduced*, laryngealized vowels contain aperiodicity, glottal closure, and significant pitch and amplitude drops. In other tokens, which I term *reduced*, many of these characteristics appear to be greatly weakened or altogether absent. Additionally, there are many ‘in-between’ forms, where some acoustic correlates of laryngealization appear to be weakened or missing, but others are not. An example of the two ends of this continuum is given below, where the laryngealized words [lo^{ʔ}o] (‘small’) and [βe^{ʔ}e] (‘house’) surface in an unreduced form in (1) and in a reduced form in (2).[1]

---

[1] The transcription of reduced forms of laryngealized words with a single vowel and no [ʔ] is not a claim about their phonological representation, but simply meant to convey the distinction between unreduced and highly reduced laryngealized roots in transcription.

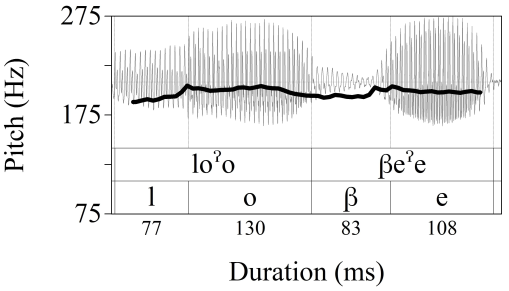
(1) **Unreduced**

ʃĩni rà loˀo βeˀe koni
see.COMPL 3M small house yest.
'The boy saw a house yesterday.'



(2) **Reduced**

ʃĩni rà lo βe koni
see.COMPL 3M small house yest.
'The boy saw a house yesterday.'



I generally refer to the weakening or apparent loss of some of the acoustic correlates of laryngealization as *laryngeal reduction*. Laryngeal reduction in SMPM bears the hallmarks of a phonetic process under the views presented above. First, it is a process that is driven primarily by speech rate. In abstractionist theories of phonology, speech rate is a phonetic factor that is not relevant to a language's phonology (McCarthy, 1986: 249-250; Keating, 1996: 263; Myers, 2000: 265-266). That is, speech rate involves how quickly articulators move from one target to another, and it should not have an impact on the abstract mental representations that phonology works with. In other words, speech rate is a physical factor, while phonology is abstract. Second, laryngeal reduction is gradient, since there are many 'intermediate stages' of reduction. Given that gradience is often associated with phonetic sound patterns and categoricity with phonological patterns, this point provides another argument in favor of analyzing laryngeal reduction as a phonetic process. Finally, the process has no clear, phonologically-defined conditioning environment—it appears to be able to apply to any laryngealized root in any phonological environment. While unconditioned processes can be modeled phonologically (e.g., redundancy rules; Stanley, 1967), the fact that laryngeal reduction involves an alternation means that a phonological account must be able to define exactly when it should occur. Without reference to some phonological configuration, this is much more difficult.

Despite the phonetic characteristics of laryngeal reduction, though, it is sometimes correlated with a change in phonological representation. The evidence for this correlation comes from the fact that highly reduced laryngealized roots often undergo a process of phonological tone sandhi that never applies to unreduced laryngealized roots. I will argue that this correlation is best understood as being the result of a phonological process of mora deletion that is conditioned by the phonetic factor of speech rate, and as a result that the phonological computation may be influenced by phonetic factors. This result is used to further narrow down the hypothesis space of frameworks of the phonology-phonetics interface, excluding models that allow for no interaction between phonology and phonetics.

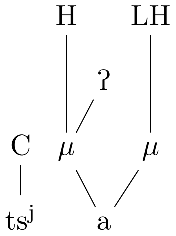
4.2 Background

4.2.1 Phonological representation

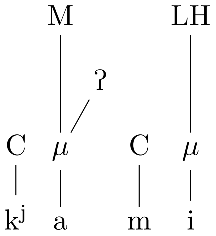
Given that the empirical focus of this chapter is on laryngealized roots, it is worth reiterating here the proposed phonological structure of laryngealized roots outlined in Chapter 2. I analyze these roots as underlyingly bi-moraic (c.f. Penner, 2019 for

Ixtayutla Mixtec), and I assume laryngealization is linked to the first of the two moras because, in roots of the shape CV^{ʔ}CV, laryngealization precedes the medial consonant. Because the tone of each mora is not predictable but rather lexically determined, I analyze laryngealized roots has having tones underlyingly linked to each mora. A phonological representation that captures all of these characteristics is shown for the roots *ts*^{j}*á*^{ʔ}*ǎ* (‘Tecomaxtlahuaca’) and *k*^{j}*a*^{ʔ}*mǐ* (‘gourd/pumpkin’) below.

(3)



(4)



This representation captures all of the main points reviewed here: [ʔ] is a suprasegmental feature associated with the first mora of the root, and each mora is associated with its own tone. This representation serves as the basis for the analysis of laryngeal reduction and its interactions with tone sandhi discussed later.

4.2.2 Laryngeal reduction across Mixtec

SMPM is by no means the only Mixtec language in which laryngeal reduction occurs, but Mixtec languages do appear to vary in whether laryngealized roots undergo severe reduction. In discussions of the phonetic realization of contrastive laryngealization in Coscatlán Mixtec (Zendejas, 2014:72-74) and San Pedro Tulixtlahuaca Mixtec (Becerra Roldán, 2019:112-116), there is no note of highly reduced realizations of laryngealized words—they are either produced with glottal closure, creaky voice, or both. However, several works on other Mixtec languages discuss a process very similar to SMPM's laryngeal reduction. Laryngealized roots in these languages may lose their laryngealization (also called 'glottalization') in certain contexts. These analogous processes are usually analyzed as involving the phonological deletion of the laryngeal feature and/or a vowel, and they have received a fair amount of attention in the Mixtec literature. For example, Pike and Small (1974:122-124) write that, in Coatzospan Mixtec, glottalized morphemes often "lose their glottal stop" in normal speech when they are not the rightmost member of a word phrase, but that "in slower, slightly emphasized speech, the same sequence of morphemes may ... have two or more glottalized or lengthened syllables." Gerfen (2013) updates this claim, showing that underlying laryngealization only surfaces in positions of phrasal prominence. For example, (5) shows a sequence of two underlyingly laryngealized roots, but only the second surfaces with laryngealization when the two roots are combined to form a larger constituent. This is because only the rightmost root here receives phrasal stress.

(5) /tʲɨʔᶤβi/ /βaʔa/ → [tʲɨβi βaʔa]
'to push' 'well'
'To push well' (Gerfen, 2013:62)

Laryngeal reduction has also been described in Chalcatongo Mixtec by Macaulay (1996), who analyzes reduced laryngealized roots as having undergone two separate phonological processes—glottal deletion and vowel deletion—in connected speech.

(6) /bàʔà/ → [bàà] → [bà]
'Good' (Macaulay, 1996:42)

Finally, Penner (2019) notes that when two laryngealized roots are combined to form a noun-noun compound in Ixtayutla Mixtec, the first often loses its laryngealization.

(7) /juˀù/ + /kúˀú^{L}/ = [jù-kùˀú]
'mouth' + 'bush'
'Bathroom.' (Penner, 2019:254)

The previous examples show that analogues of laryngeal reduction in other Mixtec languages have been analyzed as involving phonological change. However, phonologically-identical laryngealized roots also vary greatly in their realization. For example, in a production study conducted by Gerfen and Baker (2005) with speakers of Coatzospan Mixtec, participants were presented with a wordlist containing one word at a time, and asked to pronounce each of them. They saw the same list multiple times, and as a result pronounced the same word multiple times. Below are two separate productions of the same word by the same consultant in the same task.

Given that these are both productions of the same word in isolation, they both presumably have the same phonological representation. However, the form on the left correlates well with 'unreduced' laryngealized roots in SMPM, given that it contains creaky voicing, a dip and rise in amplitude, and is relatively long (509 ms). The form on the right, however, correlates well with highly reduced laryngealized roots in SMPM, given that it contains periodic voicing throughout, a dip in amplitude but no subsequent

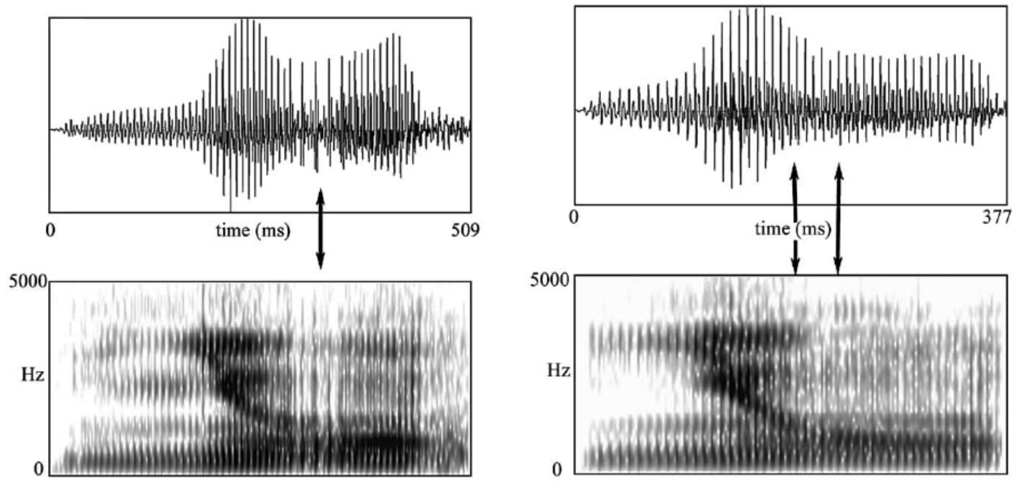


Figure 4.2: Productions of the word [ɲṵː] (‘ground’) (Gerfen and Baker, 2005:314-315)

rise, and is shorter (377 ms).

It is clear, then, that a process analogous to laryngeal reduction occurs in other Mixtec varieties, and reduced laryngealized roots are often analyzed as having undergone a phonological process that deletes laryngealization or a vowel. However, productions of laryngealized roots that ostensibly have the same phonological representation can vary greatly, and this variation seems to track relatively well with laryngeal reduction in SMPM. This conglomeration of facts raises the question of whether laryngeal reduction in SMPM should be analyzed as phonological or phonetic in nature. That is, are the weakening and apparent loss of some of the acoustic correlates of laryngealization in SMPM the result of a phonological process that deletes a vowel and/or laryngealization, as proposed for Chalcatongo Mixtec in Macaulay (1996), or is it a phonetic process whereby a single phonological representation may be realized with a range of acoustic correlates? In order to answer this question, the following sections investigate the acoustics of unreduced and reduced laryngealized roots, as well as the conditioning environment(s) and driving factors behind laryngeal reduction, concluding that the process does not appear to have a clear phonological conditioning environment and is driven primarily by speech rate.

4.2.3 Methods

Before beginning the investigation of the acoustic correlates of laryngealization and laryngeal reduction, I would like to outline the general methods used to gather some of the data analyzed throughout. Specifically, the data used in the aggregated amplitude, pitch, and H1-H2 plots was collected at a point in the research process where Consultant 1 had developed a meta-linguistic awareness of laryngeal reduction and was able to differentially produce unreduced and highly reduced versions of laryngealized roots. To investigate these acoustic measures in unreduced and highly reduced laryngealized roots, Consultant 1 was presented with a carrier phrase containing a target word, and was asked to produce the phrase five times in its unreduced form. After that, they were asked to produce the same sentence five times with reduced forms of the words. For non-laryngealized roots (CVV and CVCV), Consultant 1 was asked to produce each target word in the carrier sentence five times. The fifth and final of each set of productions was excluded from analysis to avoid any effect of intonation. Pitch, amplitude, and H1-H2 data were illustrated in aggregated plots by extracting pitch and amplitude values from target words produced in carrier sentences using Praat scripts, and H1-H2 values were extracted using VoiceSauce (Shue, 2010). These were then illustrated in R (R Core Team, 2013), often using ggplot (Wickham, 2016). In the following sections, the specifics of other tasks not described here are given.

4.3 The characteristics of laryngeal reduction

In answering the question of whether or not laryngeal reduction in SMPM should be understood as a phonological process that deletes a vowel and/or laryngealization, it is first useful to examine the acoustics of unreduced and highly reduced laryngealized roots. This point is relevant for two reasons: First, under a feed-forward model of the phonology-phonetics interface, one might assume that phonological alternations are somewhat straightforwardly reflected in the acoustic signal. According to this view, if a phonological representation has a straightforward phonetic mapping, then a change in phonological representation should have a straightforward consequence in the phonetics. Second, in many approaches to phonology, whether a process is categorical or gradient is indicative of its status as phonological or phonetic, respectively (i.e., Keating, 1996). In this regard, examining the acoustics of laryngealized roots might yield some evidence as to its phonological or phonetic nature.

I begin this section by walking through some of the acoustic correlates of unreduced and highly reduced laryngealized roots, showing that highly reduced roots vary from their unreduced counterparts in terms of amplitude, pitch, and duration, but that reduced and unreduced variants pattern together in their H1-H2 values and maintenance of pitch targets. I use these points to show that, though some of the acoustic correlates of laryngealization are greatly weakened or absent in highly reduced laryngealized roots, at least some remain relatively robust, suggesting against a phonological analysis of laryngeal deletion. After outlining these characteristics, I discuss the gradience of the process, showing that laryngealized roots surface not only as unreduced or highly reduced, but in many intermediate forms. Under the assumption that phonological processes are categorical while phonetic processes are gradient, this provides another piece of support against a phonological analysis of laryngeal reduction.

4.3.1 Pitch

In general, the surface pitch of words that undergo laryngeal reduction is consistent with their underlying tonal specification. That is, in most cases the pitch targets associated with a word's tonal melody appear to be maintained in highly reduced forms. This can be seen in the aggregated pitch plots below for the unreduced and highly reduced forms of laryngealized roots with an L-H melody and a H-L melody. The grey shaded portion of the plots indicates the section of the unreduced forms in which laryngealization interrupts pitch (determined by the distribution of NA pitch values across all unreduced examples), meaning that that portion of the pitch plot for the unreduced forms is often discontinuous and less reliable.

Finally, even tri-tonal melodies like H-LH appear to have their pitch targets maintained when appearing in a reduced form. I only know of one laryngealized root with this melody, which is the place name *tsʲáˀ ǎ* ('Tecomaxtlahuaca'), so there are very few tokens. However, the similarity in pitch plots can nonetheless be seen below.

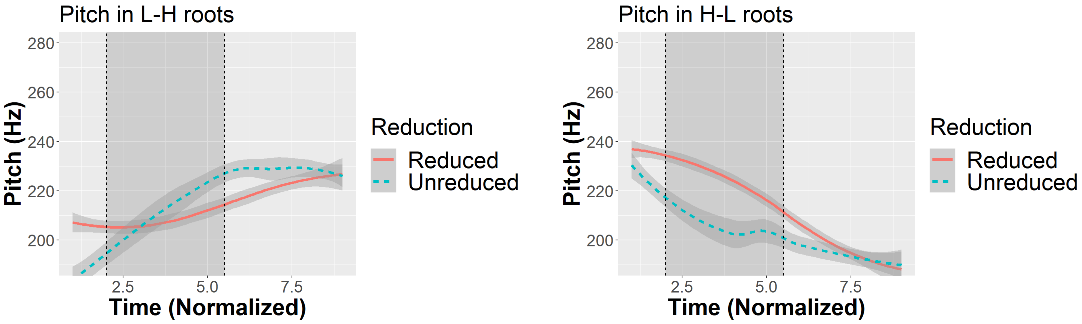


Figure 4.3: Aggregated pitch plots of unreduced and highly reduced productions of laryngealized roots with an L-H melody (left, 31 unreduced, 30 reduced) and a H-L melody (right, 24 unreduced, 17 reduced).

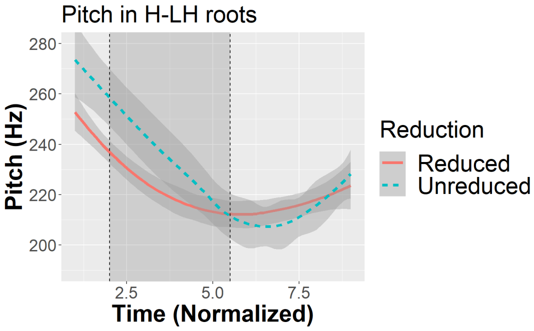


Figure 4.4: Aggregated pitch plot of unreduced and highly reduced productions of [tsʲáˀǎ] ('Tecomaxtlahuaca,' 8 unreduced, 8 reduced).

The fact that pitch targets are transparently maintained in reduced laryngealized words can be taken to suggest that laryngeal reduction does not involve the deletion of tones. In other words, it appears that the phonological tones associated with unreduced forms survive even in highly reduced laryngealized words.

4.3.2 Amplitude

One acoustic correlate of laryngealization is amplitude; amplitude tends to dip at the onset of laryngealization and to rise at its offset, creating a falling-then-rising amplitude contour. This is shown below, where the arrows indicate the direction of the amplitude contour.

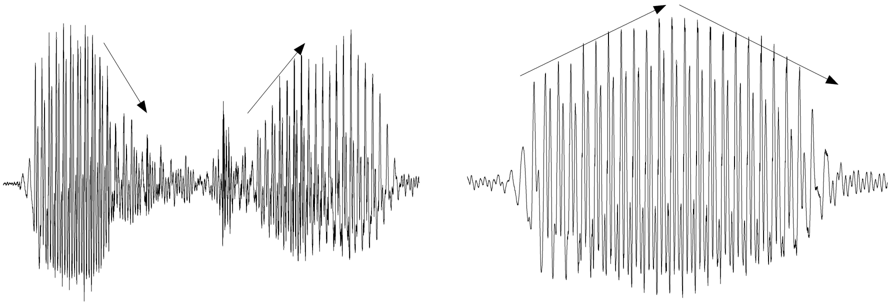


Figure 4.5: Unreduced (left, ∼280ms) and reduced (right, ∼140ms) forms of the laryngealized word [βe^{ʔ}e] spoken in the same position by Consultant 1. Arrows indicate direction of amplitude contour.

In order to test the robustness of the effect of laryngeal reduction on amplitude contour, I normalized amplitude across all productions of unreduced, reduced, and CVV roots gathered using the informal production task described in the Methods section. Amplitude was normalized using a Praat script which extracted the maximum amplitude in dB for each token, and then extracted the mean amplitude across 9 equidistant time windows for the vocalic portion of each token. The mean amplitude for each time window was divided by the maximum amplitude, yielding a normalized amplitude measure of the proportion of the token's maximum amplitude value for each time window. As seen below, the amplitude of unreduced laryngealized roots dips somewhat drastically in the first third of the vocalic portion of the root, followed by a rise in the second third. By contrast, both modal, CVV roots and highly reduced laryngealized roots have a relatively flat amplitude contour.

So, unreduced and highly reduced laryngealized roots have reliably distinct

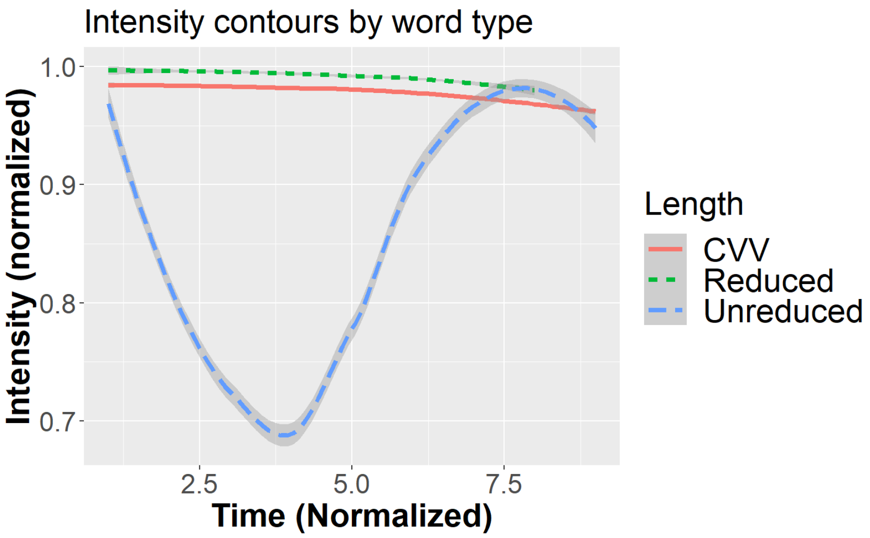


Figure 4.6: Normalized intensity contours for unreduced and reduced laryngealized roots as well as modal, CVV roots (131 CVV, 183 Unreduced, 184 Reduced).

amplitude contours. While amplitude dips and rises in unreduced forms, it stays relatively steady in highly reduced forms.

4.3.3 Duration

Another typical characteristic of highly reduced roots is that they are relatively short. This is to be expected because, as shown in §4.3, laryngeal reduction happens most often in fast speech, and vowels in fast speech tend to be shorter. However, one relevant question is whether the durational reduction of laryngealized roots in fast speech is equivalent to or more drastic than the durational reduction of non-laryngealized roots in fast speech. To investigate this, another informal production task was carried out to measure the duration of laryngealized and non-laryngealized roots in both slow and fast speech. For this task, both Consultants 1 and 2 produced target words in a carrier sentence, and they were prompted to utter each sentence first slowly and then quickly. Laryngealized roots (CV^{ʔ}V), mono-syllabic long-vowel roots (CVV), and bi-syllabic roots with two mono-moraic short vowels (CVCV) were used as target words, and the duration of each vowel was measured from the beginning of its steady-state. For CV^{ʔ}V words, vowel duration covered the entire vocalic portion of the word, including any creak or glottal closure.[2] For CVCV words, the duration of vowels in the first syllable (σ_1) were measured separately from the duration of vowels in the second syllable (σ_2). Considering the small number of consultants and items per condition, I only report basic descriptive statistics here—the mean and median duration, standard deviation, and total number of tokens for each vowel type and prompted speech rate are given below.

---

[2] Because I analyze laryngealization in SMPM as non-modal phonation, I include portions of creak and glottal closure in vowel duration because they are acoustic correlates of laryngealization, which is itself a part of the vowel.

(8) Duration (ms) by vowel type in prompted slow and fast speech for Consultant 1

Vowel type	Slow				Fast				Mean difference across rates
	Mean	Median	SD	N	Mean	Median	SD	N	
CV^{ʔ}V	260	258	52	15	84	83	17	15	176 (68%)
CVV	221	215	48	14	116	126	27	15	105 (48%)
CVCV, $\sigma 1$	110	110	33	15	72	76	22	15	38 (35%)
CVCV, $\sigma 2$	83	81	31	15	66	58	17	15	17 (20%)

(9) Duration (ms) by vowel type in prompted slow and fast speech for Consultant 2

Vowel type	Slow				Fast				Mean difference across rates
	Mean	Median	SD	N	Mean	Median	SD	N	
CV^{ʔ}V	230	223	51	12	103	99	21	12	127 (55%)
CVV	210	200	59	11	102	98	15	12	108 (51%)
CVCV, $\sigma 1$	110	111	29	12	63	65	19	12	47 (43%)
CVCV, $\sigma 2$	79	80	14	12	60	63	11	12	19 (25%)

There are some clear trends in the data. First, there is an obvious difference in duration by prompted speech rate for all vowel types, with the fast productions having lower durations than the slow productions. The mean duration of CVV vowels drops by 105 ms (48%) for Consultant 1 and by 108 (51%) for Consultant 2. The mean duration of mono-moraic vowels in the first syllable of a bi-syllabic word drops by 38 ms (35%) for Consultant 1 and by 47 ms (43%) for Consultant 2. The mean duration for the second syllable drops by 17 ms (20%) for Consultant 1 and 19 (25%) for Consultant 2. The mean duration of CV^{ʔ}V vowels drops the most for Consultant 1, with a difference of 176 ms (68%), and by less for Consultant 2, with a difference of 127 ms (55%). For Consultant 1, the difference in duration for CV^{ʔ}V vowels is particularly striking, since they are much longer in slow speech than CVV vowels but much shorter in fast speech. For Consultant 2, however, the duration loss in CV^{ʔ}V and CVV vowels is about the same. In other words, for one of the consultants surveyed here, the durational reduction of CV^{ʔ}V words in fast speech appears to be more drastic than the durational reduction of CVV words. This tendency is consistent with the phonological analysis argued for later in §5.5, which claims that highly reduced CV^{ʔ}V roots have phonologically short vowels at least some of the time, while CVV roots retain their long vowels in fast speech. One unresolved point, though, is that the apparent difference in durational reduction between CV^{ʔ}V and CVV roots is seen for Consultant 1's productions but not for Consultant 2's, even though the phonological analysis presented later applies to both Consultant 1 and Consultant 2's speech. This point is curious, but I leave it to future study to determine whether the durational variation between the consultants' productions is indicative of, for example, a differential application of the phonological reduction described later.

Ultimately, though, this type of comparison requires significantly more data. Either way, it is apparent that the highly reduced CV^{ʔ}V roots have a shorter duration than their unreduced counterparts, and for at least one consultant, this difference may be larger than the expected effect of fast speech.

4.3.4 The maintenance of laryngealization

In other Mixtec languages, processes similar to SMPM's laryngeal reduction have been analyzed as involving deletion of a laryngeal feature (i.e., Macaulay, 1996), while purportedly phonologically-identical laryngealized roots may nonetheless appear in unreduced and reduced forms in Coatzospan Mixtec (Gerfen and Baker, 2005). Additionally, wide variation in the realization of laryngealization is also shown for Yalálag Zapotec by in Avelino (2010). In fact, Gerfen and Baker (2005) showed in a lexical decision task that Coatzospan Mixtec listeners can distinguish laryngealized words from their non-laryngealized counterparts based solely off of a very small dip in f0 or amplitude alone, even with periodic vocal fold vibration throughout the vowel. It is clear, then, that the absence of some acoustic correlates of laryngealization from the signal does not automatically mean that laryngealization is absent or unrecoverable from the acoustic signal.

In fact, an acoustic analysis of highly reduced forms of laryngealized words in SMPM suggests that some acoustic correlates of laryngealization are maintained even through reduction. The relevant measure here is H1-H2, which is the difference in amplitude between the first harmonic and the second harmonic. H1-H2 is generally lower in creaky voice than in modal voice (Kreiman et al., 2010; Keating et al., 2015). I examined the H1-H2 values of highly unreduced and highly reduced tokens of laryngealized words, alongside the H1-H2 of modal vowels in bi-syllabic CVCV words.[3] Fewer measurements were taken for highly reduced laryngealized roots because of their shorter duration and the need for a sufficiently long time window to accurately measure H1-H2. To plot all values together, some modification of X-axis values was required. As a result, the longer horizontal distance between points 3 and 4 for CVCV and Unreduced roots is not indicative of longer time between points; instead, it is there so that the mid-point of the reduced roots can be plotted in its respective place.

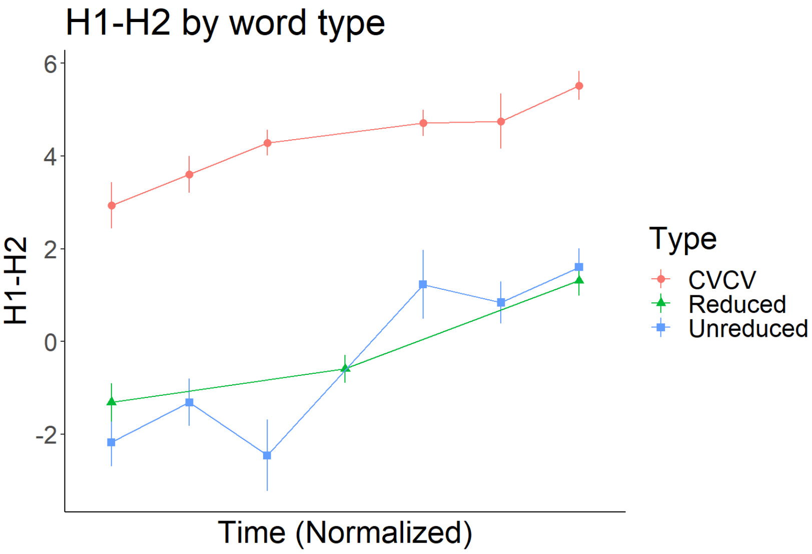


Figure 4.7: Time-normalized H1-H2 (dB, not formant-corrected) means with error bars for the vocalic portions of mid-vowel, modal CVCV roots; mid-vowel unreduced laryngealized roots; and mid-vowel, reduced laryngealized roots. Data produced by Consultant 1 (35 Unreduced, 32 Reduced, 32 CVCV).

---

[3] Only words with mid vowels were examined because formant correction was inconsistent due to the mistracking of formants for many data points, presumably caused by the consultant's high f0 (Garellek, 2019:18). Low vowels were not used because of the potential effect of nasal poles on harmonic amplitude (Simpson, 2012), and high vowels were not used because of the potential effect of the first formant on the amplitude of the harmonics.

In general, the mean H1-H2 for both unreduced and reduced laryngealized vowels is ∼4-5dB lower than the corresponding portions of modal vowels. Though this difference may seem relatively small, it is well above the reported H1-H2 just-noticeable difference (JND) of 2.6 dB for Gujarati listeners (Kreiman et al., 2010) and the JND of 2.72 dB for Mandarin listeners (Kreiman and Gerratt, 2010). This fact, combined with the findings of Gerfen and Baker (2005) that Coatzospan Mixtec listeners are robustly sensitive to even minute changes as signals of the presence of contrastive laryngealization, suggests that vowels in reduced laryngealized words are robustly different from modal vowels in a way that is likely perceptible to listeners. However, I have not established that SMPM listeners use H1-H2 in the perception of laryngealization, and the perception of phonation type by Gujarati and Mandarin listeners cannot be used to conclusively reason about the perception of phonation type by SMPM listeners. Still, given the currently-available evidence, it seems more likely than not that the differences in H1-H2 are significant enough that even highly reduced laryngealized words can still be identified as laryngealized. That is, the acoustic evidence does not suggest that laryngeal reduction involves the deletion of a laryngeal feature from the phonological representation, but rather perhaps a weakening of some of the acoustic correlates of laryngealization under the articulatory and perceptual pressures of fast speech.

4.3.5 Gradience

Finally, it is worth exploring the phonetic gradience of laryngeal reduction. So far in the discussion of the acoustic effects of the process, the tasks have focused on words that met the criteria of 'clearly unreduced' or 'clearly reduced.' However, it is the case that there are many tokens of laryngealized words that do not fit so neatly into these two categories. That is, while there are highly reduced forms and highly unreduced forms, there are many that fall somewhere in between. This fact is especially apparent in the acoustic consequences of laryngealization, which manifest in SMPM as a cline from glottal closure to apparent modal voice.[4] For example, the waveforms in 4.8 all represent the vocalic portion of the word [lo^{ʔ}o] ('small'), uttered by the same consultant in the same syntactic position.

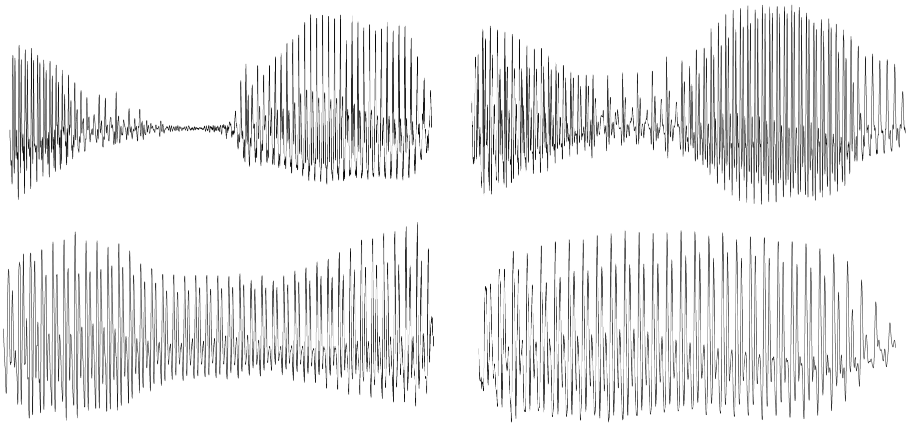


Figure 4.8: Four productions of the vocalic portion of the word *lo*^{ʔ}*o* ('small') by Consultant 1 in the same syntactic position.

The form on the top left lasts ∼300ms and shows complete glottal closure corresponding to laryngealization, and the form on the top right lasts ∼285ms and shows creaky voice, evidenced by the widely spaced glottal pulses. The form on the bottom left lasts ∼180ms and shows regular vocal fold vibration throughout, but with

---

[4] Interestingly, this squares with Gordon and Ladefoged's (2001) characterization of phonation types as existing along a continuum from open to closed vocal folds.

a dip in intensity in the middle of the vowel. Finally, the form on the bottom right lasts ~140ms and shows regular vocal fold vibration and no apparent dip in intensity due to laryngealization. This gradient realization of laryngealization is common throughout SMPM and highlights the phonetic characteristics of laryngeal reduction.

One might wonder whether acoustic measures that correlate with laryngeal reduction motivate the existence of two discrete categories (unreduced vs. reduced), or if they motivate a view of laryngeal reduction as ranging along a single continuum of the degree of reduction. In other words, one might wonder whether unreduced and reduced words represent two ends of a continuum with many intermediate forms (i.e., all realizations in Figure 7 are more or less equally likely), or whether words tend surface as unreduced or reduced only, with very few intermediate forms (i.e., only the top-left and bottom-right forms in Figure 7 are common, while the top-right and bottom-left forms are uncommon). In order to test this question, I conducted a production task aimed at measuring acoustic correlates of laryngealization across a large number of tokens produced at various rates of speech.

The relevant acoustic measures analyzed were the degree of amplitude dip and the duration of the vocalic portion of the word.[5] The reasons to use these measures are as follows: First, as mentioned before, unreduced laryngealized roots have a large amplitude dip and rise, while highly reduced roots do not (Figure 5). Additionally, the previous section showed that reduced forms have a shorter duration than unreduced

---

[5] H1-H2 was not analyzed because the data were not controlled for vowel height or nasalization. As mentioned in §4.4, issues with formant tracking made formant-corrected H1-H2 unreliable, and nasalization affects the amplitude of lower harmonics.

forms. If laryngealized words tend to be produced as either highly unreduced or highly reduced, then the distributions of these measures should be bi-modal. That is, using the example of amplitude dips, there should be two ‘kinds’ of productions: Those with a large amplitude drop, and those with little to no amplitude drop. However, if laryngealized words range along a single continuum of reduction, then there should not be two main ‘kinds’ of productions; instead, words should vary along a continuum in terms of the degree of their amplitude dip. That is, the distributions for these measures should be uni-modal or relatively flat. Below, I present the details and results of this investigation.

In this task, both Consultant 1 and Consultant 2 were asked to produce carrier sentences containing target words five times, with the first repetition being produced very slowly and each subsequent repetition being produced more quickly than the last, with the result that the fifth and final was produced very quickly. If laryngealized words tend to surface as unreduced or reduced, with few ‘in-between’ forms, then the consultants should tend to switch at some point in the productions between the unreduced forms and the reduced forms. However, if laryngealized words are realized along a cline of the strength of laryngealization, then there should be many intermediate forms between highly unreduced and highly reduced forms in this task. Both Consultant 1 and Consultant 2 took part in this task. Consultant 1 produced 145 tokens and Consultant 2 produced 164 tokens. There were 18 different target words of the shape CVˀV, which varied in consonant onset and vowel quality.

For each token, amplitude dip was measured by finding the difference between the maximum and minimum amplitude for the vocalic portion of a laryngealized word, with a larger difference correlating to a larger dip,[6] and duration was measured beginning at the steady state of the vowel. The resulting data sets contained the amplitude dip and duration for each token by each consultant. In order to test whether the acoustics merited the postulation of multiple categories, I applied Hartigans' dip test for uni-modality (Hartigan and Hartigan, 1985) to the data sets comprised of the range of amplitude dip and duration measures, with the dip test applied separately for each measure for each consultant. Hartigans' dip test provides a 'dip' value, which is a measure of the deviation of an empirical distribution (in this case, the set of recorded values for an acoustic measure across all tokens for one consultant) from a projected uni-modal distribution whose mode and data points are designed to fit the empirical data. The 'dip' is the greatest distance between the empirical distribution and the projected uni-modal distribution at any point. In general, a higher dip value correlates to a less uni-modal distribution. The reason for this is that, when there is more than one mode in the empirical distribution, the mode of the projected uni-modal distribution will be set in between the multiple empirical modes, resulting in a greater maximum distance between the two distribution functions. When the empirical distribution only has one mode, the maximum distance between the two functions will be smaller, since the mode of the projected distribution is closer to the mode of the empirical distribution. This difference is illustrated for a bi-modal and uni-modal empirical distribution in Figure 8 below.

---

[6] All tokens have *some* difference between maximum and minimum amplitude, but the differences for fully reduced words are much smaller than for unreduced words.

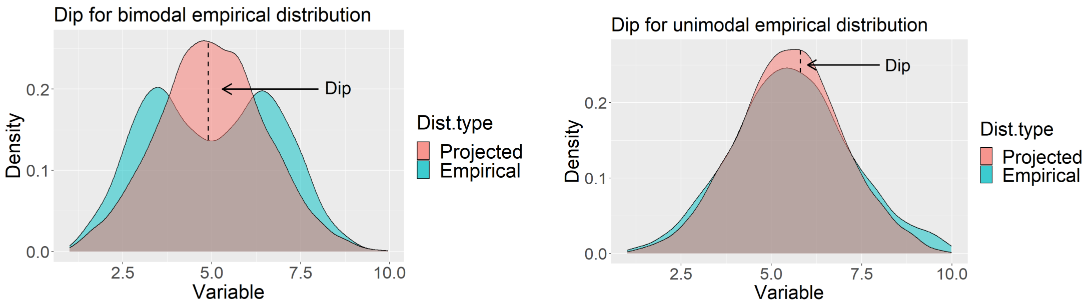


Figure 4.9: Illustration of dip for example bi-modal and uni-modal empirical distributions.

The dip test provides two outputs. The first is the dip described above. Its maximum value is 0.25, and the greater the value, the less uni-modal the distribution. The second output is the p-value, which is the probability that the data would be observed under the null hypothesis (in this case, that the empirical distribution is uni-modal). In other words, a p-value under 0.05 means that the likelihood that the empirical distribution would have resulted from a uni-modal distribution is less than 5%, and this result is considered significant. The plots of the actual distributions are given below in Figure 4.10.

For each measure for each consultant, I computed Hartigan's dip test statistic for measuring uni-modality in R using the 'diptest' package (Maechler, 2021). For Consultant 1, the null hypothesis of uni-modality could for the degree of amplitude dip could not be rejected ($D = .027$, $p = .727$), and the same result obtained for the distribution of durations ($D = .035$, $p = .283$), which is somewhat surprising given the plot of the empirical distribution. For Consultant 2, the degree of amplitude dip was determined not to be uni-modal ($D = 0.05$, $p < .01$), but the null hypothesis of

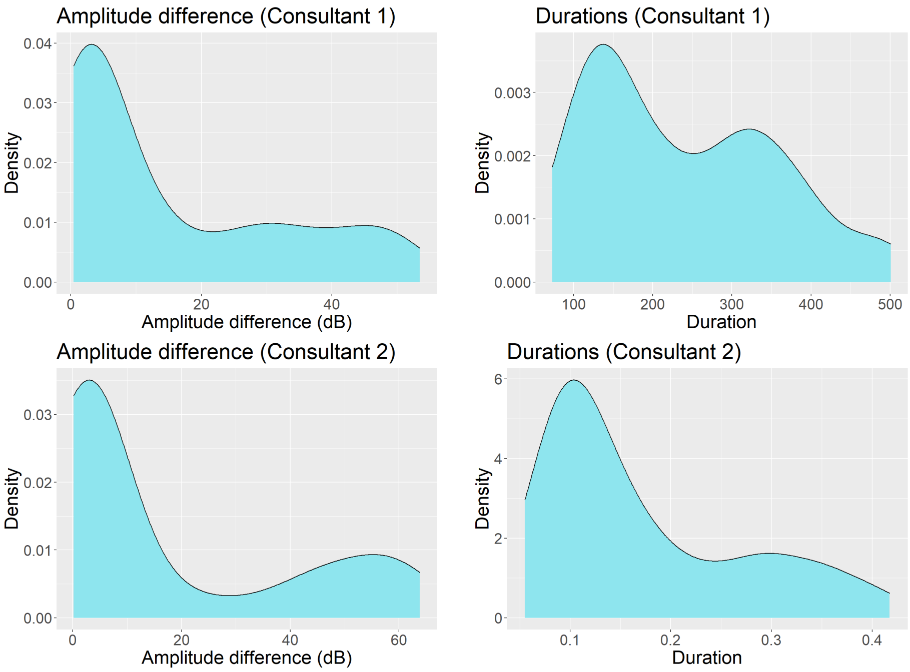


Figure 4.10: Raw distributions of amplitude dip and duration across all productions for both consultants.

uni-modality for the distribution of durations could not be rejected (D = .019, p = .974). In other words, three out of the four distributions of these acoustic measures of laryngeal reduction could not be determined to be non-uni-modal, except for the degree of amplitude dip for Consultant 2. By this measure for Consultant 2, laryngeal reduction is somewhat categorical in that words can be sorted into two categories, even though there is some overlap between them. However, by duration for both consultants, and by the degree of amplitude drop for Consultant 1, laryngeal reduction appears to exist along an acoustic continuum because laryngealized roots cannot be sorted into two separate distributions.

It is important to note, though, that one should be cautious in assuming that the presence or absence of acoustic categories directly corresponds to the presence or absence of phonological categories. There are two reasons for this. The first is that I have not shown (nor do I have the necessary evidence to show) that these acoustic correlates are used by SMPM listeners as auditory cues to the presence or absence of laryngealization, just that they correlate with the apparent strength of laryngealization in the acoustic signal. The second is that speech perception is multi-dimensional (see e.g., Lisker, 1986 for cues to English stop voicing contrasts), with perceptual dimensions being non-linear to acoustic dimensions and highly interdependent (e.g., Holt and Lotto, 2010). In the light of possible non-linear transforms between acoustic and perceptual space, it is technically possible for a uni-modal distribution along an acoustic dimension to be mapped to a bi-modal distribution along a corresponding perceptual dimension, and, in principle, vice versa. Because of this, I take caution in reasoning from these acoustic distributions to phonological distributions, and a perceptual study is necessary in order to confidently determine the effect of these acoustic dimensions on listeners' categorization of the signal. That being said, if the acoustic distributions are taken as suggestive evidence, they provide support for a view of laryngeal reduction in SMPM as highly gradient. That is, when measured in terms of amplitude dip and duration, words do not tend to fall neatly into acoustic categories of 'unreduced' and 'reduced.' Instead, they tend to fall somewhere along a continuum of reduction.

4.3.6 Review

In this section, I have outlined various acoustic correlates of laryngeal reduction in SMPM. In general, laryngealized roots tend to have an interrupted pitch contour in their unreduced form and an uninterrupted pitch contour in their reduced form. That being said, it appears that all phonological tones are retained through laryngeal reduction, since their pitch targets are all present even in highly reduced roots. Another difference between unreduced and reduced laryngealized roots is found in their amplitude contour; unreduced roots have a steep dip and subsequence rise in amplitude, while highly reduced roots do not. Highly reduced roots also have a much shorter duration than their unreduced counterparts, and this reduction in duration appears, at least for Consultant 1, to potentially be greater than the expected reduction due to fast speech.

Even though unreduced and highly reduced laryngealized roots have distinct acoustic correlates, they do share some: H1-H2, a cue to non-modal phonation often used by listeners, is roughly equivalent between unreduced and highly reduced laryngealized roots, to the exclusion of modal vowels. This point suggests that a phonological analysis of laryngeal reduction in SMPM as the phonological deletion of laryngealization is less desirable, since at least some acoustic correlates of the contrast are robustly available in the acoustic signal. Additionally, the high degree of gradience of laryngeal reduction make a phonological analysis that involves the categorical deletion or modification of a phonological feature more difficult, though by degree of amplitude dip, Consultant 2's productions are non-unimodal.

However, some alternations that are thought of as phonological are nonetheless gradient, either in their rate of application or in their extent of application. For example, the optionality of *t*/*d*-flapping in English can be explained by making reference to grammar-external factors like production planning (Wagner, 2012; Kilbourn-Ceron et al., 2016; Kilbourn-Ceron, 2017), and some phonologically-complete alternations are nonetheless gradiently realized, as in the case of incomplete neutralization in Chinese Tone 3 sandhi (Du and Durvasula, 2020). In this light, the optionality and gradience of laryngeal reduction and the maintenance of some acoustic correlates of laryngealization in highly reduced forms do not necessarily preclude a phonological analysis. However, if it is also found that the driving factors behind laryngeal reduction are non-phonological, this would provide yet further evidence against a phonological analysis of the process. Because of this, I now turn to the conditioning environment and driving factors of laryngeal reduction.

4.4 The phonological environment and driving factors of laryngeal reduction

In this section, I examine the distribution of the application of laryngeal reduction, showing that it may apply in essentially any phonological context, though there are certain prosodic positions in which laryngealized roots are less likely to be highly reduced. Additionally, I provide evidence that one of the main driving factors behind laryngeal reduction is speech rate, which is usually considered to be an extra-grammatical factor not taken into account in phonological computation (McCarthy, 1986: 249-250; Keating, 1996: 263; Myers, 2000: 265-266; c.f. Kaisse, 1985). These points—the lack of a clear phonological conditioning environment and the drastic effect of speech rate on laryngeal reduction—combine to provide further evidence against a phonological analysis of the alternation. However, as I will argue in §5, the phonological behavior of unreduced and highly reduced laryngealized words does motivate a distinction in phonological representation between unreduced and at least some highly reduced laryngealized roots.

4.4.1 Conditioning environment

As mentioned earlier, an optional process may nonetheless be phonological. However, the optionality of phonological processes is often linked to grammar-external factors. For example, one hypothesis is that production planning plays a role in whether a phonological process applies or not (Wagner, 2012; Kilbourn-Ceron et al., 2016; Kilbourn-Ceron, 2017). Under this view, utterances are planned in chunks, and these chunks do not always align perfectly with the conditioning environment of phonological alternations. Take, for example, *t/d*-flapping in English, the process by which /t/ and /d/ can be realized as the flap [ɾ] between vowels (e.g., 'bet' vs 'betting,' 'bed' vs 'bedding'). In terms of a phonological rule, it can be formulated as follows:

(10) **t/d-flapping**

/t,d/ → [ɾ] / V_(#)V

This process is optional across word boundaries, since the word ‘add’ in the phrase ‘add or subtract’ can be pronounced as [æd] or [æɾ]. But this optionality can be understood in terms of whether the alternation’s phonological conditioning environment is entirely contained in the same production planning window. If the entire conditioning environment is contained in the same planning window, then the process applies; if only a portion of the conditioning environment is contained in the planning window, then the process does not apply. This application or lack of application can be visualized below, where one box represents a single production planning window. In (11), the entire conditioning environment is contained in the same planning window, so flapping applies. In (12), the conditioning environment is split across multiple planning windows. Because the conditioning environment is not met within a single planning window, flapping does not apply.

(11) **Application**

$\boxed{\text{...Vd\#V...}} \rightarrow$ [...Vɾ#V...]

(12) **Non-Application**

$\boxed{\text{...Vd}}\ \boxed{\text{\#V...}} \rightarrow$ [...Vd#V...]

In this light, it is worth examining whether there is a phonologically-defined conditioning environment for laryngeal reduction because, if it is the case that laryngeal reduction happens consistently in a given phonological environment, then an appeal to something like production planning might be made to account at least for the gradience in its rate of application, if not for the gradience in its extent of application. This section examines the various phonological environments in which laryngealized roots undergo reduction. As a preview, it appears that there is not a clear conditioning environment for laryngeal reduction, since it can happen anywhere in an utterance, though there are some prosodic tendencies.

The first point to establish is that laryngeal reduction applies to roots of all syntactic categories, so long as they are of the shape (CV)CVˀV. The following examples show verb roots (13)-(14); noun roots (14)-(15), (18); a preposition (17); and an adjective and adverb (18) in an unreduced and reduced form. Additionally, laryngeal reduction appears to apply to roots in many different prosodic positions. The following examples show laryngeal reduction applying to utterance-initial roots (14), (18); utterance-medial roots (13)-(18); to stand-alone DPs (15)-(16); to roots that are subparts of a DP (14), (18); to roots that are the heads of complex DPs (*ñã̀ˀ=ĩ̂* in (19)); to roots that are separated from utterance-final position only by a weak pronoun (13), (17); and even sometimes to morphologically-complex laryngealized words (*ñã̀ˀ=ĩ̂* in (19)). These examples show that laryngeal reduction happens in all sorts of environments.

(13) nũ̀ʰnĩ̌ tʃíˀì/tʃî ɾà
corn plant.CONT 3M
‘He is planting corn.’

(14) sá-kʷáˀa/sá-kʷâ nà tũ̀ˀũ/tũ̀ ⁿdáˀβì
CAUS-go(?).CONT 3N.PL word poor
‘They study Mixtec.’

(15) ʃîʃi tsìmã̂ˀã̀/tsìmã̂ pã́ã̀
eat.COMPL raccoon bread
‘The raccoon ate the bread.’

(16) ʃĩn=ì tsʲòˀo/tsʲò koni
see.COMPL=1SG root yesterday
‘I saw a root yesterday.’

(17) sáá kàʰtʃi ñá ʃĩ́ˀĩ̀/ʃĩ̂ nà
SO SAY.COMPL 3F with 3N.PL
‘That’s what she said to them.’

(18) loˀo/lo kʷéˀe/kʷê jùˀù/jù péðɾo
small very mouth Pedro
‘Pedro’s mouth is very small.’

(19) tàʃĩ amígò ñã̀ˀ=ĩ̂/ñã͡ĩ tsʲàˀá/tsʲǎ ⁿdàˀ=î
give.COMPL friend POSS=1SG salsa hand=1SG
'My friend gave me salsa.'

Though reduction is relatively free in its distribution, there are two positions in which laryngealized roots are much less likely to reduce, and these are when they are utterance-final and when they are under narrow focus, which triggers fronting of the focused argument (Ostrove, 2018; Hedding, 2019a). For example, the following sentences show that a reduced laryngealized root is dispreferred utterance-finally and in a focus-fronted position.

(20) tàˀβi ʧúʰtu kòˀǒ/#kǒ
break.COMPL cat plate
'The cat broke the plate.'

(21) **Question**

nã̀ã́ nàkàβa nũ̀ʰũ̌ ñũˀũ̌?
what fall.COMPL face ground
'What fell to the ground?'

(22) **Answer**

βeˀe/#βe nàkàβa nũ̀ʰũ̌ ñũˀũ̌
house fall.COMPL face earth
'The *house* fell to the ground.'

However, the dispreference for laryngeal reduction utterance-finally lessens if the word in question has been previously mentioned. For example, in the following discourse, laryngeal reduction of an utterance-final root is possible:

(23) **Question**

tàˀβi ʧúʰtu kòˀǒ?
break.COMPL cat plate
'Did the cat break the plate?'

(24) **Answer**

ã̀hã̌, tàˀβi ʧúʰtu kǒ
yes, break.COMPL cat plate
'Yes, the cat broke the plate.'

In fact, somewhat surprisingly, laryngealized roots that are used as fragment answers may appear in a reduced form, again if the root has been previously mentioned. Note also that in this case, the laryngealized root is under information focus.

(25) **Context**

kʷìˀi	ɾa	ʃì̃ˀǐ	βa	nìʃì̃ʰjǒ	nũ̀ʰũ̌	maɾía
fruit	and	mushroom	EMPH	exist.COMPL	face	María

‘María had fruit and a mushroom.’

(26) **Question**

ⁿtsʲàá	já	ʃàʃi	ñá?
which	3SG.N	eat.COMPL	3SG.F

‘Which did she eat?’

(27) **Answer**

✓ kʷì
fruit

‘Fruit.’

So, it appears that laryngealized roots may reduce in nearly every prosodic configuration, except for utterance-finally and under narrow focus. However, having previously mentioned the root in the discourse makes reduction possible even in these positions.

The lack of reduction in these environments might be used as evidence for an analysis of the process as prosodically-conditioned, under a line of reasoning somewhat like the following: Laryngeal reduction could be analyzed as applying whenever a laryngealized root is in a particular prosodic configuration (e.g., it is a foot that is non-final in a prosodic word), and it doesn't apply when the root is not in that environment. Under this view, the fact that reduction does not apply to utterance-final or focus-fronted roots is evidence for this view—utterance-final and bare, focus-fronted roots can never be non-final in a prosodic word, so reduction is blocked in these environments. Finally, under this account, a word that is previously mentioned in the discourse context would have a different prosodic structure than one that has not been previously mentioned, and it is this difference which allows exceptional reduction of previously-mentioned roots. Similar information-structural differences in prosodic organization have been found in, for example, Yanbian Korean (Jun and Jiang, 2019). So, under this type of view, laryngeal reduction might be a prosodically-conditioned phonological process that applies in a consistent environment, and never occurs outside of that environment.

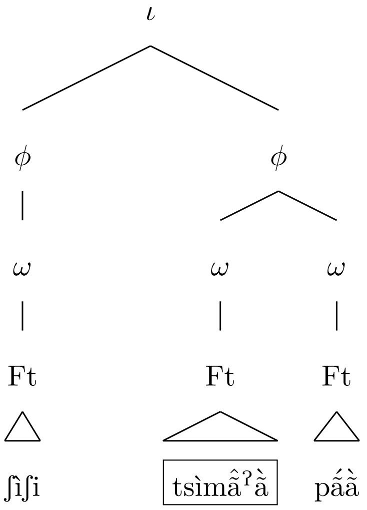
There are several reasons why it is unlikely that laryngeal reduction is conditioned and/or blocked by phonologically-defined prosodic structure. The first is that an analysis of laryngeal reduction as applying within a specific, phonologically-defined prosodic configuration predicts that it should apply in all and only the cases in which that specific prosodic structure is present. For instance, we might adopt the proposal above that laryngealized roots that are non-final in a prosodic word undergo reduction (c.f. Pike and Small, 1974; Gerfen, 2013). However, as we saw in (13)-(19), laryngeal reduction may apply to roots in nearly any syntactic configuration. It is difficult to imagine a syntax-prosody mapping process that allows so much variance that a subject DP like the one in (15), repeated below, would be the rightmost foot in a prosodic word in one utterance but not in another. The prosodic schemata required under the potential proposal here are illustrated in (29)-(30), using units from Prosodic Hierarchy Theory (Selkirk, 1984; Nespor and Vogel, 1986) for illustration .

(28)	ʃìʃi	tsìmã̂ˀã̀/tsìmã̂	pã́ã̀
	eat.COMPL	raccoon	bread

‘The raccoon ate the bread.’

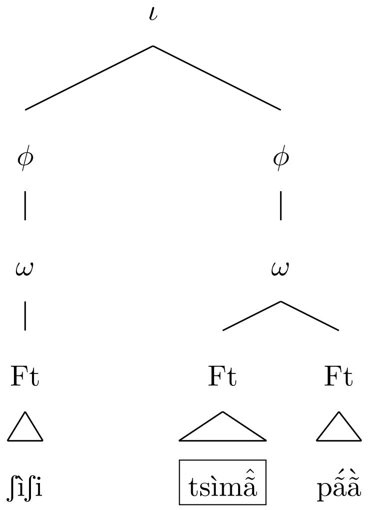
(29) **Non-application of reduction**

(ω-final Ft)



(30) **Application of reduction**

(non-ω-final Ft)



This type of analysis, then, requires a large amount of variation in how a given syntactic constituent is mapped to a prosodic constituent, but it is precisely this type of variation that would be required in order to analyze laryngeal reduction as being a phonological process triggered by a specific prosodic configuration. This high degree of variation is one reason that I do not pursue this type of analysis.

The second reason I do not pursue this view is that there are independent phonetic factors that likely account for the dispreference for laryngeal reduction utterance-finally and under narrow focus. For example, in Yoloxóchitl Mixtec, words lengthen utterance-finally (DiCanio et al., 2020) and under narrow focus (DiCanio et al., 2018).

Though I do not have the data to show that this is the case for SMPM, it does seem reasonable to suppose that words might lengthen in these contexts, given that these patterns are seen in another Mixtec language and are relatively robust cross-linguistically (Cambier-Langeveld and Turk, 1999; Chen, 2006; Fletcher, 2010). This is important because duration is an important factor in reduction. For example, shorter duration correlates with greater undershoot of target formant values in vowels, and longer duration with less undershoot (i.e., Lindblom, 1963; Moon and Lindblom, 1994). In this light, it is not suprising that laryngeal reduction is dispreferred in exactly those positions where duration is lengthened (i.e., utterance-finally and under narrow focus). If laryngeal reduction is analogous to vowel target undershoot, then we would expect it to apply less readily under lengthening, which is the pattern we see. So, durational lengthening in these contexts might help to explain the dispreference for laryngeal reduction without requiring the positing of a phonological conditioning environment, which is independently undesirable because it requires a one-to-many syntax-to-prosody mapping.

4.4.2 Interim review

It appears, then, that laryngeal reduction in SMPM does not have a clear, phonologically-defined conditioning environment. Though there are some tendencies against reduction utterance-finally and under narrow focus, these tendencies are just as readily explained as the result of durational lengthening as they are as the result of specific prosodic structures phonologically blocking the application of laryngeal reduction. These points together make a phonological analysis of reduction significantly more difficult.

The final place to investigate in order to determine whether laryngeal reduction should be viewed as phonological or phonetic is in the factors that drive it: If these driving factors also lie outside of the phonological grammar proper, then this might be another nail in the coffin of a phonological analysis. We have already seen hints that duration is correlated with reduction, and also that previous mentions appear to make reduction more likely. In the following section, I will show that speech rate appears to be the main driving factor behind laryngeal reduction, and that there is likely also an effect of previous mentions. This point is important because speech rate is often considered an extra-grammatical factor not taken into account in the phonological grammar proper.

4.4.3 The effect of speech rate

The clear effect of speech rate on laryngeal reduction, and a potential effect of previous mentions, can be seen in the results of an informal production task carried out with Consultant 1. In this task, Consultant 1 was asked to produce utterances twice, once at a slow rate of speech and once at a fast rate of speech. An example of this process is given below:

(31) **Linguist:**

I am going to ask you to translate a sentence from Spanish into Mixtec, and to say the sentence twice, the first time slowly and the second time quickly. How do you say ‘The raccoon ate the bread’?

(32) **Consultant:**

a. **Slow repetition**

ʃìʃi tsìmã́ˀã̀ pã́ã̀
eat.COMPL raccoon bread
‘The raccoon ate the bread.’

b. **Fast repetition**

ʃìʃi tsìmã̂ pã́ã̀
eat.COMPL raccoon bread
‘The raccoon ate the bread.’

This task tested the effects of speech rate and previous mentions on laryngeal reduction. In (32-a), the word for *raccoon* has not been previously mentioned and is given at a slow rate of speech. In (32-b), however, *raccoon* has been previously mentioned and is produced at a fast speech rate. In a subsequent elicitation session, the same sentence would be presented again, with the consultant asked to produce the sentence first quickly and then slowly. In this case, the first repetition would not have been previously mentioned but would be uttered at a fast rate of speech. The second repetition would have been previously mentioned but uttered at a slow rate of speech. As a result of this set-up, the effect of previous mentions and speech rate could be somewhat reliably disentangled.

30 sentences total were used, 25 of which contained laryngealized roots and 5 fillers which did not. Each sentence was produced in both orders of speech rate (slow-then-fast and fast-then-slow, with order of speed varied), and the slow-then-fast and fast-then-slow productions of the same sentence were almost never prompted within the same elicitation session. Some sentences contained multiple laryngealized words, and only those laryngealized roots that were non-final and non-focused were analyzed, since utterance-final and narrow-focused laryngealized roots are less likely to reduce. The

result of this setup is 120 total sentence productions (30 sentences x 2 productions x 2 rates) with 98 analyzable productions of laryngealized roots.

For the purposes of this task, laryngealized roots were labeled as unreduced if their amplitude showed a clear dip followed by a rise, or if there was any visible creaky voice and/or glottal closure followed by modal voice in the waveform and spectrogram, as shown in the waveform to the left. Words were labeled as reduced if they had no obvious dip and subsequent rise in amplitude and no visible creak followed by modal voicing, as shown in the waveform to the right. For this task, these criteria tended not to conflict.

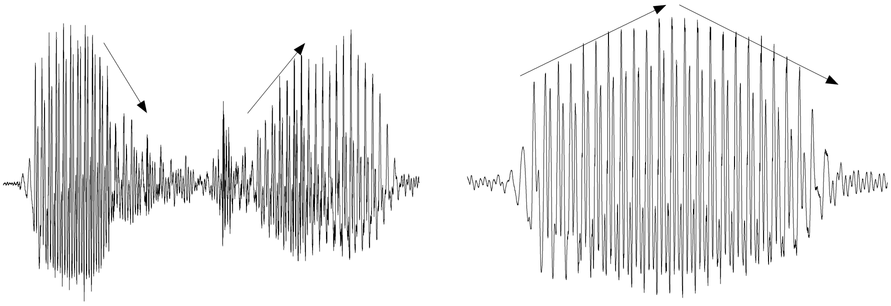


Figure 4.11: Unreduced (left, ∼280ms) and reduced (right, ∼140ms) forms of the laryngealized word [βe^{ʔ}e] spoken in the same position by Consultant 1. Arrows indicate direction of amplitude contour.

The results of this task were telling: 26/49 laryngealized roots met the criteria for being classified as reduced in fast speech, while only 1/49 met the criteria in slow speech. It appears that whether or not a word was previously mentioned may have also had an effect on reduction rates in the fast repetition, but not in the slow production.

(33) Reduction rates by production speed and previous mention.

	Fast production	Slow production	
Previously mentioned	16/25	0/24	16/49
Not previously mentioned	10/24	1/25	11/49
	26/49	1/49	**Total**

Though the task outlined above lacks the rigorous control expected of a production study, it can be taken as suggestive that speech rate has a large effect on the probability of laryngeal reduction. A chi-square test comes out as significant (χ-square: 35.11, df $=$ 3, p $<$ .001), suggesting that speech rate, whether a word has been previously mentioned, and whether it appears in a reduced form are not independent. Looking at the raw values suggests that speech rate is the main driving factor behind reduction in this task, since the difference in reduction rates between the fast and slow productions is much higher than the difference between previously-mentioned and not-previously-mentioned productions.

The task described above was able to separate the potential influence of previous mentions from an influence of speech rate, showing that speech rate appears to be the main driving factor in laryngeal reduction. Another informal production task also shows the influence of speech rate on laryngeal reduction, but does not control for previous mentions. However, it was conducted with both consultants instead of just one,[7] and it contains significantly more analyzable tokens, so it is worth reporting the results here. The task in question is the speech rate manipulation described in §4.5. In this task, both consultants produced target words in carrier sentences five times, with the first repetition being produced very slowly and each subsequent repetition being produced more quickly than the last, with the fifth and final being produced very quickly. These results provide a window into the effect of speech rate on laryngeal reduction at a more fine-grained level than that described above, since it involved changing speech rate gradually across five productions, instead of a single fast-slow binary. As stated in §4.5, Consultant 1 produced 145 tokens (29 tokens at each speed), and Consultant 2 produced 164 (32 tokens at each speed, with one set of productions thrown out because it only contained four repetitions). There were 18 distinct target words of the shape CV^{ʔ}V, which varied in consonant onset and vowel quality. There were no filler items. The reduction rates are given for each consultant below:

---

[7] Consultant 1 participated in both tasks. However, the two tasks were separated from each other by about 1.5 years, so it is unlikely that taking part in the first task had a very large effect on Consultant 1's productions in the second one.

(34) Reduction rates by production speed for both consultants.

	Production Speed (slow → fast)					
	1	**2**	**3**	**4**	**5**	
Consultant 1	0/29	3/29	12/29	20/29	25/29	60/145
Consultant 2	0/32	5/32	10/32	21/32	31/32	67/160
	0/61	8/61	22/61	41/61	56/61	**Total**

The rate of reduction clearly increases for both Consultants as speech rate increases, suggesting again that speech rate is a principal driving factor in laryngeal reduction. This has been shown across two tasks, then—one in which speech rate was not confounded with previous mentions, and one with more tokens and participants but in which speech rate was confounded with previous mentions (as speech rate increased, so did the number of times the target word had been previously mentioned). It is clear, then, that speech rate is at least one of the main driving factors in determining whether or not a laryngealized root will surface in a reduced form.

Another potential tendency illustrated here and bolstered by the discussion in §4.1 is for laryngealized roots to be more likely to reduce when they have been previously mentioned than when they have not. This tendency is not surprising—it is common for words to be reduced if they have been previously mentioned in discourse (Bard et al., 2000; Warner, 2011).

4.4.4 Review

In this section, I have argued that laryngeal reduction in SMPM does not have a clear phonologically-defined conditioning environment, and that positing one requires permitting an unusually high degree of variance in syntax-prosody mapping. Specifically, laryngeal reduction appears to apply to laryngealized roots regardless of prosodic context, and the dispreference for utterance-final and focus-fronted reduction can likely be explained by making reference to increased duration in these contexts. I have also shown that one of the most important driving factors behind laryngeal reduction is speech rate—laryngealized roots are much more likely to reduce in fast speech than in slow speech. This final point is important because sensitivity to speech rate is often used to diagnose a sound pattern as non-phonological (McCarthy, 1986:249-250; Keating, 1996:263; Myers, 2000:265-266; c.f. Kaisse, 1985). These points, taken alongside the process's gradient nature and robust maintenance of H1-H2 differences even in highly reduced roots, conspire to point toward a non-phonological analysis of the phenomenon. That is, the acoustics, conditioning environment, and driving factors behind laryngeal reduction suggest that there is no change in phonological representation between unreduced and highly reduced laryngealized roots. However, as I will show in the following section, at least some highly reduced laryngealized roots do have a distinct phonological structure from their unreduced counterparts. The evidence for this conclusion comes from the interaction of laryngeal reduction with an independent phonological process of tone sandhi.

4.5 Tone sandhi and mora deletion

In this section, I describe a phonological process of tone sandhi in SMPM that reliably distinguishes between rising tones linked to a single mora and rising tonal melodies that span two moras. The interaction of this sandhi rule with laryngeal reduction provides evidence that at least some highly reduced laryngealized roots are mono-moraic instead of bi-moraic. This fact means that laryngeal reduction is sometimes correlated with a change to the abstract phonological representation. In order to make this point, I describe the relevant tone sandhi process as well as its interaction with laryngealized words with an L-H melody.

4.5.1 Tone sandhi

Unlike many other Mixtec languages where there is widespread evidence of floating tones and tone sandhi, relatively few instances of phonological tone sandhi have been described for SMPM. However, one process described in Hedding (2019b) is relatively robust. When a word-final LH contour tone is followed by a word-initial H tone, the word-final LH tone flattens to L. This process, which I term rise flattening, is schematized in (35) and illustrated in the following examples. In (36), the word-final rise of *tsʲũ̀hũ̌* (‘turkey’) surfaces faithfully. In (37), it surfaces as a flat low tone.

(35) **Rise flattening**

/LH # H/ → [L # H]

(36) **Non-application**

ni-ⁿtsìʰkù tsʲũ̀hũ̌ leso
COMPL-chase turkey rabbit
'The turkey chased the rabbit.'

(37) **Application**

ni-ⁿtsìʰkù tsʲũ̀hũ̀ léló
COMPL-chase turkey skunk
'The turkey chased the skunk.'

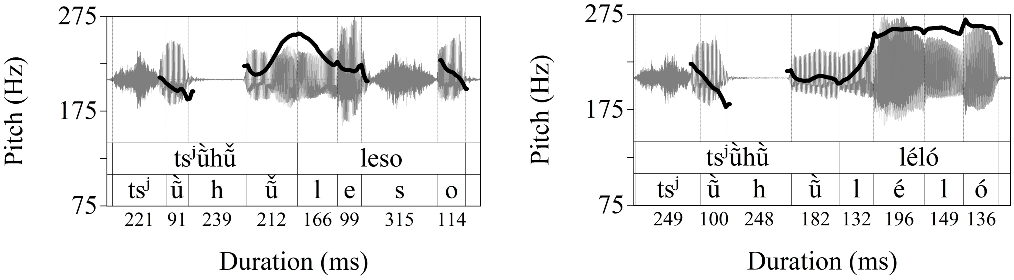


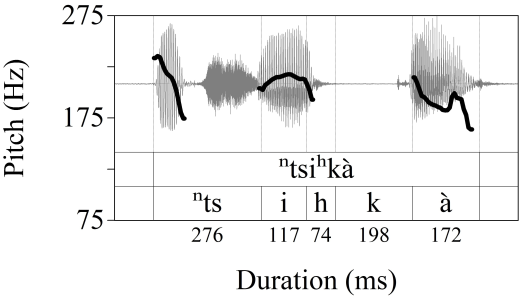
Figure 4.12: Waveforms with pitch tracks showing non-application (left) and application (right) of tone sandhi.

It is important to note that this tone sandhi process is phonological, not phonetic, and involves a change in abstract representation. First, it is categorical, occurring across-the-board when its conditioning environment is met. This is also true across speech rates—it applies in fast and slow speech, and even when there is a pause between words. The second reason to believe that the process is phonological is that it is neutralizing, or very near-neutralizing, as argued in Chapter 3. This can be seen in the following examples. (38) and (39) show a near-minimal pair, with (38) having a final L tone and (39) having a final LH tone. (40) shows that the underlying final LH of 'banana' surfaces with a pitch contour very similar to that of the underlying final L in 'chest' in (38).

(38) **Underlying Final L**

ⁿtsiʰkà

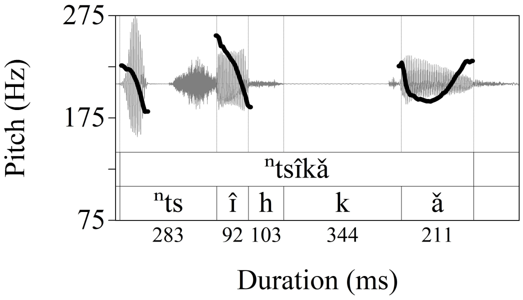
‘Chest’



(39) **Final LH**

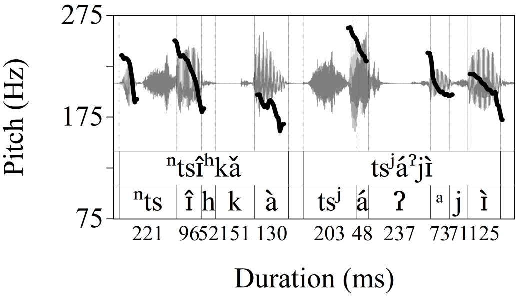
ⁿtsîʰkǎ

‘Banana’



(40) **Derived Final L**

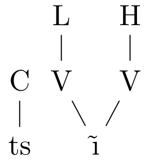
ⁿtsîʰkà tsʲáˀjì
banana rotten
‘A rotten banana.’



Another piece of evidence for the phonological status of tone sandhi in SMPM comes from the fact that not all *surface* rises from L to H undergo this process: When the L and H are linked to separate moras, and thus do not form a contour unit, the tone sandhi process described above does *not* take place. This can be seen clearly in words with a bi-moraic, mono-syllable template (CVV) with an L-H melody. For example, the word in (41) has a L-H melody, with the L linked to one mora and the H linked to the other (41). The tonal melody of this word is realized as rising pitch (42). As (43) shows, this rise does not flatten before an H tone:

(41) *tsĩ̀ĩ́*

‘Rat.’



(42) **Non-application**

tsĩ̀ĩ̀ tsĩ̀ĩ́ kolo
COMPL.bite rat turkey
‘The rat bit the turkey.’

(43) **Non-application**

tsĩ̀ĩ tsĩ̀ĩ́ táʰtè
COMPL.bite rat man
‘The rat bit the man.’

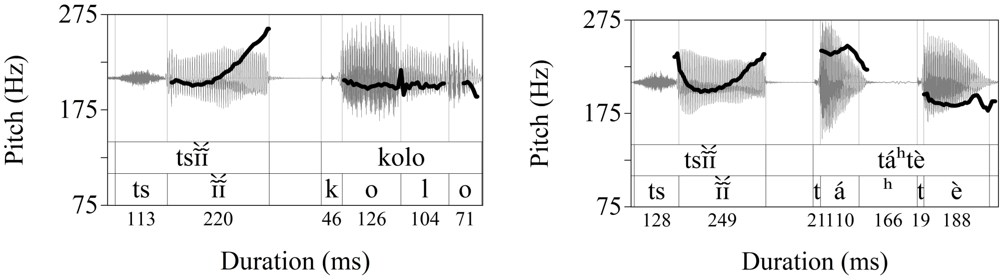
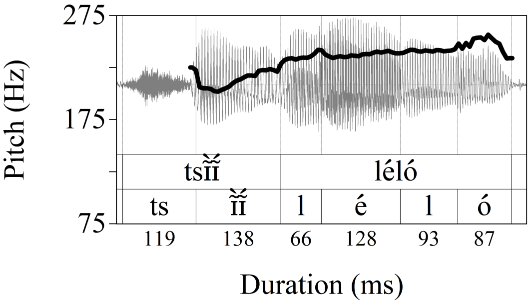


Figure 4.13: Waveforms with pitch tracks showing non-application of tone sandhi to an L-H sequence linked to two moras.

Importantly, rise flattening does not apply to bi-moraic, L-H melodies even in fast speech. This can be seen in the following example, where the rise on the word for ‘rat’ has about the same duration as the vowel hosting the derived L on the word for ‘banana’ in (40). In other words, the lack of application of rise flattening to L-H melodies on bi-moraic, mono-syllabic words is not due to the increased duration of these vowels, but rather due to their distinct phonological structure.

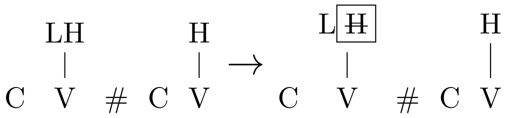
(44) **Non-application**

ʃǐnì tsĩ̀ĩ́ léló
see.COMPL rat skunk
‘The rat saw the skunk.’

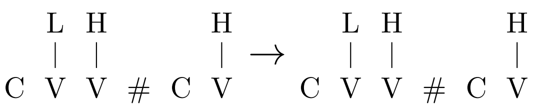


There is a clear difference, then, between LH contour tones linked to a single mora and L-H sequences linked to separate moras, illustrated by the two schematizations below. While LH contours linked to a single mora undergo H deletion, L-H sequences linked to two moras do not.

(45) **Tone sandhi**



(46) **No tone sandhi**

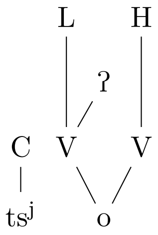


4.5.2 Interim review and prediction

So far, I have argued that rise flattening is a phonological process, and that it applies to LH contour tones that are linked to a single mora, but not to L-H melodies where the L and H tones are linked to separate moras. This fact, considered alongside the phonetic nature of laryngeal reduction, leads to the following prediction: Because laryngealized roots with a L-H melody have the same tonal alignment as CVV roots with a L-H melody (47)-(48), rise flattening should not apply to laryngealized roots with a L-H melody. Additionally, because laryngal reduction does not appear to be a phonological process, then whether or not a laryngealized root undergoes reduction should not change whether or not it undergoes tone sandhi.

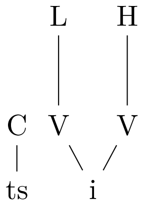
(47) tsʲòˀó

‘Flea.’



(48) tsĩ̀ĩ́

‘Rat.’



However, as I will show in the following section, the facts are not so simple. While the unreduced forms of laryngealized words do not undergo tone sandhi, the reduced forms *do* often undergo sandhi. This fact suggests that the phonological representation in (47), is not the only one associated with laryngealized words. Instead, words that are highly reduced can have a separate representation. The consequence of this asymmetry is that laryngeal reduction is at least sometimes correlated with phonological change, despite its purportedly phonetic nature.

4.5.3 Tone sandhi and laryngeal reduction

Laryngealized roots with an L-H melody alternate between an unreduced form, in which the L and H are separated by laryngealization, and a reduced form, which has a continuous rising contour. Figure 12 shows representative examples of this alternation for both consultants in the sentences (49) and (50), where unreduced and reduced forms of the L-H word *tsʲòˀó* are in an environment that does not trigger tone sandhi. In (49), the L and H tones surface on either side of laryngealization. In (50), the L and H tones have formed a rising contour.

(49) **Non-sandhi environment**

tsȉȉ [tsʲòˀó] kolo
COMPL.bite flea turkey
'The flea bit the turkey.'

(50) **Non-sandhi environment**

tsȉȉ [tsʲǒ] kolo
COMPL.bite flea turkey
'The flea bit the turkey.'

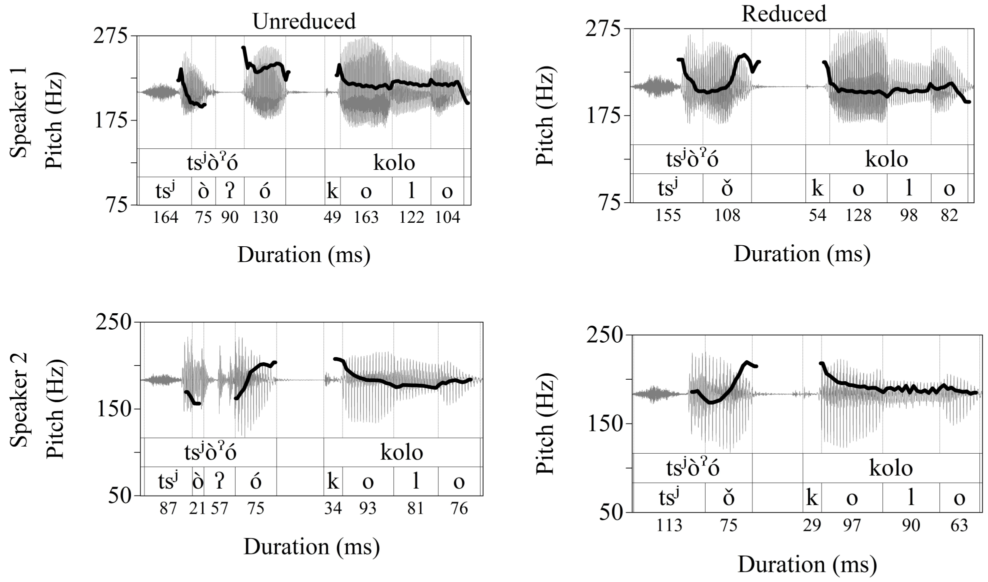


Figure 4.14: Representative examples of unreduced and reduced laryngealized words with an L-H melody in a non-sandhi-triggering environment for both consultants.

When the same laryngealized root with an L-H melody is placed before an H-initial word, creating the environment for tone sandhi, there is a distinction between the unreduced and reduced forms. The unreduced form surfaces faithfully, with an L and H tone separated by laryngealization. However, on the highly reduced form of the same root, the expected L-H rise surfaces as a flat L tone, showing that tone sandhi *has* taken place. Figure 13 shows representative examples of this alternation applying to the laryngealized root *tsʲòˀó* in sentences (51) and (52) for both consultants. This pattern is in contrast to (43), which shows that bi-moraic L-H melodies do not undergo sandhi.

(51) **Non-application of sandhi**

tsȉȉ tsʲòˀó táʰte
COMPL.bite flea man
'The flea bit the man.'

(52) **Application of sandhi**

tsȉȉ tsʲò táʰte
COMPL.bite flea man
'The flea bit the man.'

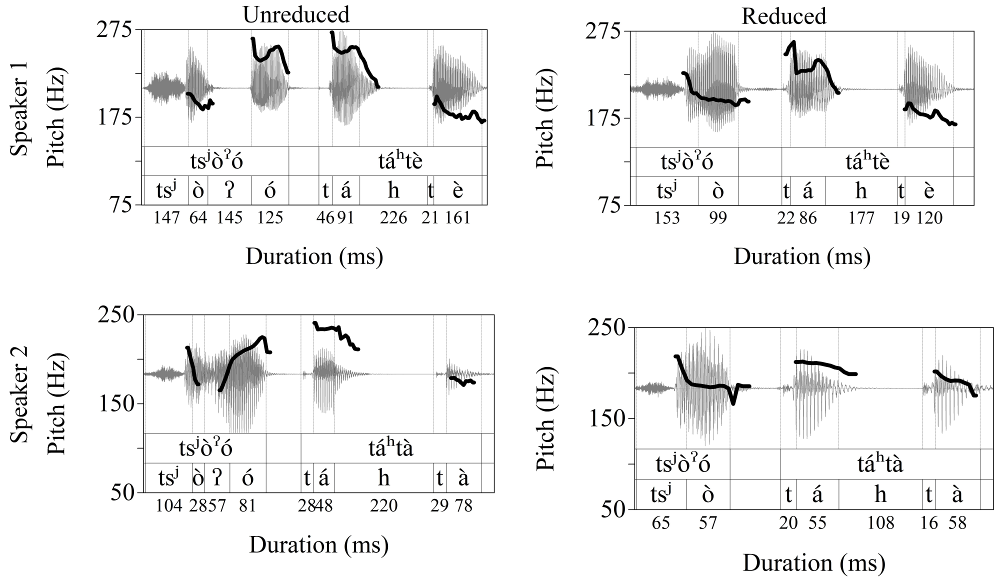


Figure 4.15: Representative examples of unreduced and reduced laryngealized words with an L-H melody in a sandhi environment for both consultants.

There are cases in which the application of rise flattening to highly reduced laryngealized roots does not occur, but the pattern shown above is nonetheless relatively consistent. This can be seen in the pitch plot in Figure 4.16 below, which shows aggregated pitch contours for unreduced and highly reduced laryngealized roots with an L-H melody in the conditioning environment of tone sandhi. The pitch contour of highly reduced roots tracks relatively well with the pitch contour of underlying L tones in the same context. Finally, it is worth noting that I have not seen any cases of rise flattening applying to an intermediate case of laryngeal reduction, such as one that has no creak but still has an amplitude dip and rise.

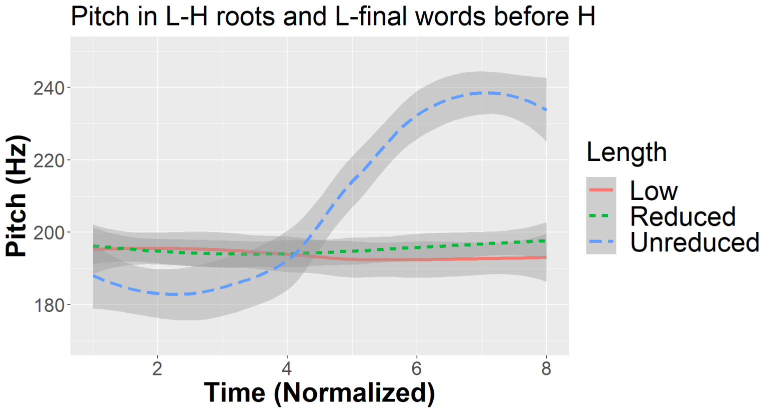
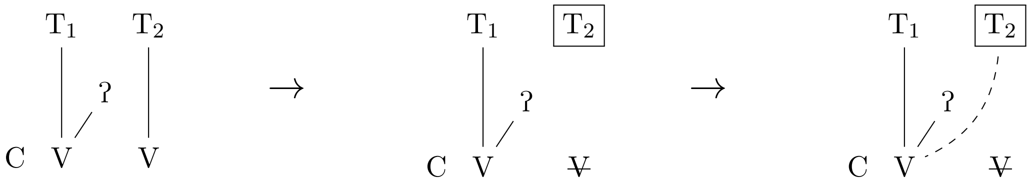


Figure 4.16: Pitch (Hz) before an H tone for highly reduced and unreduced productions of laryngealized roots with an L-H melody, as well as vowels with an underlying L tone (14 Unreduced, 13 Reduced, 14 Low).

We have seen, then, that the phonological process of tone sandhi applies to LH contour tones linked to a single mora (37), but not to a L-H melody linked to two moras (43)-(44). We have also seen that tone sandhi does not ever apply to the unreduced form of laryngealized roots that have a L-H melody. However, it is the case that sandhi applies to many highly reduced forms of roots with L-H melodies.

The fact that highly reduced and unreduced forms of laryngealized roots with an L-H melody behave differently with respect to tone sandhi shows that their phonological representation is at least sometimes categorically distinct: While unreduced roots have a melody consisting of a sequence of L and H linked to separate moras, the reduced forms of the words have apparently been re-analyzed as containing an LH contour tone linked to a single mora. I take this fact as evidence that laryngeal reduction, though a phonetic process, often correlates with the phonological deletion of a mora and the re-association of tone to the remaining mora. Given that laryngealization is maintained on the reduced forms of laryngealized words (§3.4), the deleted mora is the second, which is not linked to laryngealization.

(53)



This result is highly consequential because it provides evidence that laryngeal reduction at least sometimes correlates with a phonological alternation: Highly reduced laryngealized roots often have a different abstract, categorical representation. This result is unexpected given the phonetic characteristics of the alternation outlined in §3-4.

4.5.4 Interim Review

In this section, I have demonstrated that laryngeal reduction is sometimes associated with a change in phonological representation, namely the deletion of a mora. The evidence for this claim comes from a phonological tone sandhi process that applies to LH contour tones linked to a single mora but not to L-H melodies spanning two moras. This process does not apply to unreduced laryngealized roots with an L-H melody, consistent with their bi-moraic nature, but does apply to many highly reduced forms of the same type of root, suggesting that they are phonologically mono-moraic. This fact is surprising when one considers the phonetic characteristics of the process laid out in §3-4, and it suggests that a phonetic process (laryngeal reduction) is correlated with a phonological process (mora deletion). In the following section, I consider the consequences of this slate of properties for an analysis of laryngeal reduction and tone sandhi in SMPM, arguing that laryngeal reduction and mora deletion are two separate processes, the former phonetic and the latter phonological, that are connected to each other in SMPM's sound system. Given the tight relationship between the two and the phonetic conditioning of laryngeal reduction, I argue that the phonological process of mora deletion is conditioned by the phonetic factors, such as speech rate, and thus constitutes an instance of a phonological process that is conditioned by purportedly phonetic factors.

4.6 Consequences

§3 showed that laryngeal reduction is a highly gradient process that does not appear to result in wholesale deletion of tone or a laryngeal feature, since H1-H2 remains reliably distinct between highly reduced laryngealized roots and modal vowels. §4 showed that laryngeal reduction cannot be clearly shown to occur in a specific, phonologically-defined environment, even if that environment is defined in terms of prosodic structure, and that it appears that speech rate is the main driving factor behind the process. These points suggest that laryngeal reduction is a phonetic process that does not reflect a change a in phonological representation. Despite this, §5 showed that some highly reduced laryngealized roots are phonologically distinct from unreduced laryngealized roots, suggesting that laryngeal reduction is at least sometimes correlated with a change in phonological representation. In other words, an apparently phonetic process of laryngeal reduction is correlated with an apparently phonological process of mora deletion.

I would like to argue that mora deletion and laryngeal reduction are two separate processes, but that the two influence each other. That is laryngeal reduction and mora deletion are not the *same* process—not all highly reduced laryngealized roots have undergone mora deletion, as evidenced by the fact that reduced laryngealized roots with an L-H melody do not always undergo rise flattening—but they are also connected to each other—if a laryngealized root is highly reduced, then it is much more likely to undergo mora deletion. Likewise, if a laryngealized root has undergone mora deletion, then it will almost certainly surface in what appears to be a highly reduced form. There are two primary ways that this connection might be thought of: The first is to say that the two processes are perhaps related to each other diachronically, with laryngeal reduction being the phonetic precursor of mora deletion, but that they are independent of each other synchronically. The second possibility is that laryngeal reduction and mora deletion are distinct processes, but that their conditioning factors are the same. Specifically, the purportedly phonetic factors that cause drive laryngeal reduction also drive mora deletion. I will briefly walk through each possibility here, arguing against the first and for the second.

4.6.1 Rule scattering

One possible analysis of the SMPM facts is to say that laryngeal reduction and mora deletion are independent processes—one phonological and one phonetic—that are related diachronically to each other, but do not synchronically interact. For example, the relationship between the two might be an instance of rule scattering (Bermúdez-Otero, 2015). Rule scattering describes a situation in which related sound patterns exist independently at different levels of a language's grammar (e.g., both lexically and post-lexically, or, in the case of SMPM, both post-lexically and phonetically), with the two processes being diachronically related to each other but nonetheless synchronically distinct.

An example of rule scattering is English palatalization, which exists both as a phonological rule (i.e., *press/pressure*, where the final /s/ in *press* becomes an [ʃ] in *pressure*) and as a phonetic process of coarticulation (i.e., *press your point*, where the final /s/ in *press* is coarticulated with the following [j] and is produced as something close to an [ʃ]). Here, a phonological process of palatalization (/s/ → [ʃ] / _+j) coexists with a similar phonetic process of coarticulation, whereby an [s] becomes more palatal when coarticulated with a [j] across a word boundary (see Zsiga, 1995; 2000 for discussion). These two rules are likely diachronically related to each other, with the phonetic, coarticulatory process presumably having been phonologized and incorporated into the grammar of English at some point. That being said, they are synchronically distinct in several ways. For example, phonological palatalization is categorical and unaffected by speech rate, but coarticulatory palatalization is gradient and highly dependent on speech rate.

In a similar way, the phonological process of mora deletion in SMPM might be thought of as coexisting with a phonetic process of reduction, with the two nonetheless being independent in the same way that phonological and phonetic palatalization are in English. Here, mora deletion would be a phonological alternation that occurs in a predictable set of circumstances. When it applies, the resulting laryngealized root surfaces in what appears to be a highly reduced form, simply because that is how mono-moraic laryngealized outputs are implemented by the phonetic system. Independently, bi-moraic laryngealized roots may nonetheless surface with a variety of degrees of reduction due to the phonetic process of laryngeal reduction. This is schematized in Figure 4.17 below, which shows that phonologically-reduced and phonetically-reduced laryngealized roots could overlap significantly at the reduced end of a continuum of reduction.

However, an analysis of laryngeal reduction and mora deletion in SMPM as scattered rules which are related diachronically but independent synchronically is not motivated empirically. The reason for this is that the relationship between the two processes in SMPM is different from the relationship between the phonological and phonetic palatalization processes in English. For example, the phonological rule of palatalization in English is clearly distinct from the phonetic process of palatal coarticulation—though coarticulation is gradient and affected by speech rate, the phonological rule has neither of these characteristics. Instead, it applies whenever its phonological conditioning en-

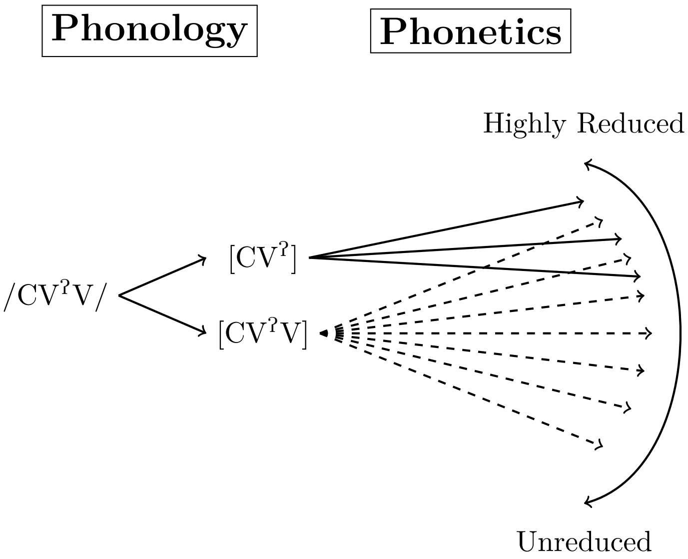


Figure 4.17: Illustration of the phonetic implementation of mono- and bi-moraic laryngealized roots.

vironment is met, even in slow speech. Unlike in English, though, there is no evidence in SMPM that the phonological process of mora deletion is independent of speech rate. Instead, the main factor driving it appears to be speech rate, just as it is with laryngeal reduction.[8]

This raises an issue for a purely phonological analysis of mora deletion—if one of the principal conditioning factors of the alternation is non-phonological, then how can it be derived in the phonological grammar? Given that there is apparently no phonological distinction between environments in which laryngeal reduction and mora deletion may or may not occur, there is no clear way for mora deletion to be modeled purely in the phonological grammar. Instead, it appears that laryngeal reduction—

---

[8] Laryngealized function roots, such as prepositions, sometimes appear in a highly reduced form even in normal/slow speech. However, I know of no prepositions which have a L-H melody, which makes it difficult to test whether mora deletion has applied in these cases.

either the process itself, or the factors that trigger it—are at least partially behind the application of mora deletion. Because of the influence of speech rate and the lack of phonologically-definable conditioning factors, an analysis in which mora deletion is not synchronically tied to laryngeal reduction is stipulative and not empirically justified.[9]

4.6.2 Consequences for the interface

If laryngeal reduction and mora deletion are not synchronically independent of each other, then there must be some dependency between them. It appears that the linking piece between the two is speech rate: Laryngeal reduction applies much more readily in fast speech, and it is to reduced laryngealized roots in fast speech that mora deletion is most likely to apply. As noted above, there do not appear to be phonological configurations conditioning or blocking the application of either process, meaning that the triggering factors appear to be phonetic. This result is important, since it suggests that a phonological process can be driven synchronically by the purportedly phonetic factor of speech rate. In the case that one wishes to keep phonological and phonetics separate—as I argued in Chapter 3—and in the case that one wishes to keep considerations of speech rate out of the phonological grammar—as I will argue for in Chapter 5—then a phonology-phonetics framework that is able to account for laryngeal

---

[9] Note that, even if this analysis were augmented by making reference to a production-planning account like that of Wagner (2012) or Kilbourn-Ceron et al. (2016) in order to derive variability in the application of mora deletion, the lack of phonological conditioning factors still keeps the analysis from adequately modeling the process. The reason for this is that the essence of a production-planning account is that gradience in the application of external sandhi processes can be boiled down to whether the entire phonological conditioning environment for the sandhi process is present in the production planning window or not. In SMPM, where there does not appear to be a phonologically-defined conditioning environment, one cannot appeal to planning windows to derive the gradience because there is no phonological conditioning environment for them to contain.

reduction and mora deletion in SMPM must necessarily allow for some sort of synchronic interaction between phonology and phonetics. Motivating this view and elucidating the nature of this interaction is the topic of Chapter 5, but it is nonetheless possible to come to an interim conclusion here. Specifically, since phonology and phonetics appear to be interacting synchronically in SMPM, frameworks which disallow any such type of interaction are ruled out. When combined with the conclusion from Chapter 3 that phonology and phonetics must be distinct, this conclusion narrows down the possible space of frameworks of the phonology-phonetics interface that are able to account for SMPM's sound system. Specifically, only those frameworks that lie in between the two dotted lines on the continuum have the necessary characteristics to model low tone spread, laryngeal reduction, and mora deletion in SMPM. These frameworks all posit a distinction between phonology and phonetics, but have some mechanism built in by which the two may interact. The combination of these features is crucial: Without a separation between phonology and phonetics, a framework is unable to model the opaque interaction between low tone spread and rise flattening. However, without some sort of interaction between phonology and phonetics, a framework is unable to model the relationship between laryngeal reduction and mora deletion in SMPM.

4.7 Conclusion

In this chapter, I have described and analyzed a process of laryngeal reduction in SMPM that is similar to other phonological reduction processes described in

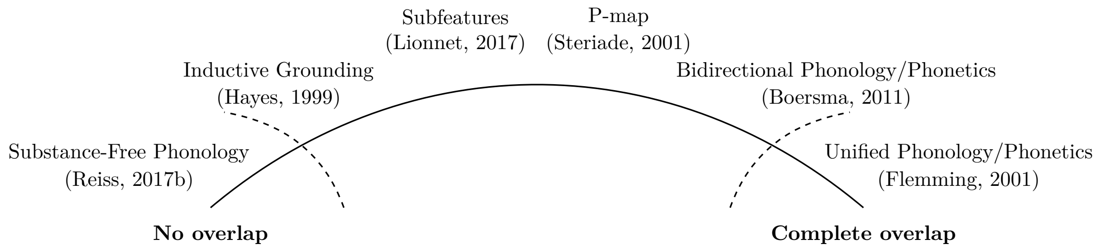


Figure 4.18: Continuum of frameworks of the phonology-phonetics interface

other Mixtec languages (Pike and Small, 1974; Gerfen, 2013; Macaulay, 1996; Penner, 2019). Given that this type of reduction has been analyzed phonologically in other varieties (Macaulay, 1996; Gerfen, 2013), but that phonologically-identical laryngealized roots may nonetheless have vastly different acoustic characteristics (Gerfen and Baker, 2005), I examined the acoustic correlates, conditioning environment, and driving factors behind laryngeal reduction. Analysis of the acoustics of unreduced and highly reduced laryngealized roots in §3 showed no apparent phonological deletion of tones or laryngealization, since pitch targets and H1-H2 values are robustly maintained even in highly reduced laryngealized roots. Additionally, it was shown that laryngeal reduction is highly gradient, and laryngealized roots cannot be neatly binned into discrete acoustic categories corresponding to 'unreduced' and 'reduced.' Then, §4 showed that there is no clear, phonologically-defined conditioning environment for laryngeal reduction, since roots of this type can reduce in essentially any phonological environment. Even tendencies against reduction of utterance-final and focus-fronted laryngealized roots are likely explainable in terms of durational lengthening in these contexts. Additionally, the main factor driving laryngeal reduction appears to be speech rate, which is often considered to be extra-phonological. This constellation of facts points to an analysis of laryngeal reduction as a phonetic process that does not reflect a change in phonological representation.

However, through an investigation of a phonological tone sandhi process in the language, §5 showed that laryngeal reduction is often correlated with a change in phonological representation, such that unreduced roots are bi-moraic, but highly reduced roots are often phonologically mono-moraic. This alternation was analyzed as a phonological process of mora deletion, and it was argued that laryngeal reduction and phonological mora deletion are two distinct processes that are nonetheless conditioned primarily by the same factors. These factors are largely extra-grammatical, given the lack of phonological conditioning environment for laryngeal reduction alongside the speech-rate-driven and gradient nature of the process. The picture that emerges is one in which the phonological alternation of mora deletion is influenced primarily by factors that lie outside of the phonological grammar proper. Given that these facts require an interaction between phonology and the purportedly phonetic factor of speech rate, I have argued that frameworks which posit complete separation between phonology and phonetics are not adequately able to model this pattern. When considered alongside the conclusion of Chapter 3 that phonology and phonetics are distinct systems, the conclusion is that the frameworks that are able to model low tone spread, laryngeal reduction, and mora deletion in SMPM are those that distinguish between phonological and phonetic levels of representation, but still allow the two to interact synchronically. It is the nature of this interaction to which I turn in the following chapter.

Chapter 5

Framework construction and comparison

5.1 Introduction

Throughout the previous two chapters, I have argued that there are at least two necessary features of any successful model of the interface between phonology and phonetics: The two systems must constitute distinct levels of representation and analysis, as evidenced by the inability of a direct phonetics framework to adequately model low tone spread, but they must also be allowed to influence each other, as necessitated by inductive grounding and the influence of speech rate on mora deletion. If these requirement are imposed on frameworks modeling the phonology-phonetics interface, then the hypothesis space is slightly narrowed. However, the models that are incompatible with these requirements only constitute the very ends of the continuum, and those that are consistent with these intuitions implement them in different ways. For example, the subfeatural account of Lionnet (2016, 2017) divides phonology and phonetics into separate systems, but allows the phonological grammar to refer to units that constitute an abstraction over expected, phonetic coarticulation. This type of interaction between phonetics and phonology is very different than, for example, models of phonetically-grounded constraint induction (Hayes, 1999; Smith, 2004; Flack, 2007), which allow phonetics to influence phonology by determining the content of at least some phonological constraints. Yet other proposals, such as Steriade's P-Map (2001, 2008), maintain that language users' phonetic knowledge establishes fixed rankings of phonological constraints. So, though many frameworks of the interface embody the characteristics I have argued for thus far, they do so in different ways. With the aim of elucidating the similarities and differences between the remaining viable models of the interface, this chapter motivates and outlines one way of implementing the characteristics I have argued are necessary for any framework, and then compares it to several other approaches, showing that the characteristics can be fulfilled in a variety of ways, but that each distinct way of applying them makes different typological predictions and has different amounts of empirical coverage.

5.2 Speech rate in phonology

I argued in Chapter 4 that speech rate has an effect on the application of a phonological process of mora deletion in SMPM. Importantly, other cases of speech rate influencing phonology have been reported in the literature. This section reviews these processes, showing that the cases of speech rate influencing phonology can be divided into two types of interactions. In most instances, increased speech rate apparently widens the domain of application of an extant phonological process. However, there are other cases that cannot be analyzed in this way, since they do not involve an alternation applying across larger domains than usual. In each case, though, the alternation triggered by an increase in speech rate is one that is already used elsewhere in the language's phonology. Crucially, it appears to be the case that speech-rate-driven alternations are also attested at normal rates of speech, with the difference between rates being that the alternation applies either across different prosodic boundaries or in different environments. An example of this characteristic can be seen in SMPM, in which mora deletion is a fast-speech process that has rate-independent analogues. The first analogue is found with function words like prepositions, which often surface in a highly reduced form even in slow speech. Crucially, non-laryngealized roots also undergo this process, like the breathy root [nũ̀ʰũ̌] ('face/to') in (1-b).[1]

(1) /ʃã̀ˀã̀/ → [ʃã̀]
foot
'For.'

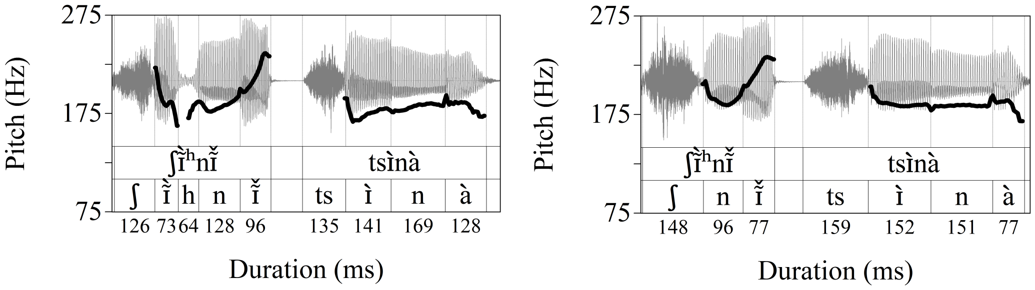
(2) /nũ̀ʰũ̌/ → [nũ̌]
face
'To.'

A second analogue is found in complex DPs. If a fricative-initial noun is non-final in a complex DP, its initial vowel can be elided, as seen in (3), where the word [ʃĩ̀ʰnĩ̌] ('head') surfaces both with and without its initial vowel.

---

[1] Unfortunately, I know of no prepositions with an L-H melody, so it is difficult to tell whether phonological mora deletion or severe phonetic reduction applies in these cases. However, given that the highly reduced forms can even occur in slow speech (though they do not always), it is likely that they reflect a categorically distinct representation.

(3) ká^{ʔ}nũ̀ ʃĩ̀^{h}nĩ̌/ʃnĩ̌ tsìnà
big head dog
‘The dog’s head is big.’



This elision process can apply to fricative-initial nouns regardless of the voicing of the medial consonant, and also regardless of the height of the elided vowel. For example, (3) showed the process applying to a noun with a medial voiced consonant, and (4) shows it applying to a noun with a medial voiceless consonant. Additionally, (3) and (4) show elision applying to high vowels, but (5) and (6) show the process applying to a mid and low vowel, respectively.

(4) ʃì^{h}tà/ʃtà ká^{ʔ}nũ̀
tortilla big
‘A big tortilla.’

(5) sò^{h}kò/skò tʃútu
shoulder cat
‘The cat’s shoulder.’

(6) sà^{h}tà/stà tʃú^{h}tu
back cat
‘The cat’s back.’

Both the reduction of function words and vowel elision can apply (though optionally) even in slow speech, and the latter process is insensitive to factors like consonant voicing and vowel height that would feasibly influence coarticulatory processes like gestural overlap (Mo, 2007). These characteristics point to their being phonological alternations. This point is important because it means that the rate-dependent process of mora deletion in SMPM described in the previous chapter has rate-independent analogues. Specifically, it appears that rate-dependent mora deletion is the application of an existing family of vowel-deletion processes in a new environment, namely to laryngealized, non-function words in fast speech. As I will show in the rest of this section, this characteristic also holds of rate-dependent alternations outside of Mixtec—they all have a rate-independent analogue elsewhere in the language's phonology. To illustrate this, I will first walk through cases in which speech rate widens the domain of application of a phonological process, and then I will discuss a case in which it appears to trigger an existing phonological alternation in a new environment.

Before beginning, though, it is important to note that many speech-rate-driven alternations have been argued not to result from phonological change, but rather from the influence of speech rate on articulation. For example, the loss of word-initial, pretonic schwa in fast speech in English (e.g., 'potato' vs. 'ptato') has in some cases been analyzed as a phonological process of schwa deletion (Zwicky, 1972; Kaisse, 1985), but in other cases as not involving phonological deletion but rather gestural overlap (Browman and Goldstein, 1992; Davidson, 2006). As a result, I limit my investigation of fast-speech phenomena to those that are demonstrably phonological, with their phonological nature hinging on their affecting demonstrably-phonological alternations or interacting with other demonstrably-phonological alternations. The resulting set of languages and alternations is small (6 languages, 7 including SMPM), but some trends are clear.

5.2.1 Domain widening

The first two instances of interaction between speech rate and phonology come from Mandarin and Nantong Chinese. In Mandarin Chinese, as described by Cheng (1966), a Tone 3 (falling-then-rising, indicated below with ‘v̌’) becomes a Tone 2 (rising, indicated below with ‘v́’) when followed by another Tone 3. In a string of five words that are all specified for Tone 3, the tonal output varies by speech rate. In slow speech, Tone 3 sandhi only applies between words separated by a weak prosodic boundary. In faster speech, Tone 3 sandhi applies across stronger prosodic boundaries, and in rapid speech, Tone 3 sandhi applies across all prosodic boundaries. This is shown in (7)-(9), where the words that have undergone Tone 3 sandhi in the output are bolded. The parentheses correspond to syntactic constituents and, by extension, prosodic constituents.

(7) **Slow Speech:**

/((lǎo lǐ) (maǐ (měi jǐu)))/ → [((**láo** lǐ) (maǐ **(méi** jǐu)))]

‘Old Li buys good wine.’

(8) **Faster Speech:**

/((lǎo lǐ) (maǐ (měi jǐu)))/ → [((**láo** lǐ) **(maí (méi** jǐu)))]

‘Old Li buys good wine.’

(9) **Rapid Speech:**

/((lǎo lǐ) (maǐ (měi jǐu)))/ → [**((láo lí) (maí (méi** jǐu)))]

‘Old Li buys good wine.’ (Cheng, 1966:150-151)

A similar change is described for Nantong Chinese in Ao (1993). In Nantong Chinese, any syllable that is not word-final and not the leftmost member of a foot loses its tone, and its surface tone is determined by rightward spreading of the leftmost tone. This can be seen in (10), where the low tone on the second syllable (indicated by ‘v̀’) is non-final and not the head of a foot. It is overwritten by the H portion of the underlying mid-to-high contour of the preceding vowel (indicated as a unit by ‘v̌’ and as a sequence by ‘v̄.v́’):

(10) /jěŋ.kə̀ŋ.xβ̩̌/ → [(jēŋ.kə́ŋ).(xβ̩̌)]

‘Man-made lake.’ (Ao, 1993:114-115)

The application of this tone deletion and spreading process varies within the word depending on speech rate. Because of the lack of long, mono-morphemic words in the native lexicon, the author uses a loan word for ‘Bolivia’ to illustrate this point. Using the term ‘foot’ a bit loosely here, each syllable makes its own foot in slow speech (11), while in normal speech the first two syllables combine to form a foot (12). Finally, in fast speech, the first three syllables form a foot to the exclusion of the final syllable, which is privileged (13). Since the ‘foot’ is the domain of the application of this tone deletion and spreading process, the initial M tone spreads in increasingly larger domains as speech rate increases, suggesting that as speech rate increases, the amount of material in a single prosodic domain increases.[2]

---

[2] Ao (1993) differentiates between LM and MH contours in this example; I do not do so here for convenience’s sake.

(11)	**Slow Speech:**	(12)	**Normal Speech:**	(13)	**Fast Speech:**
	(pū)(lî)(vě)(â)		(pūlī)(vě)(â)		(pūlīvē)(â)
	‘Bolivia.’		‘Bolivia.’		‘Bolivia.’ (Ao, 1993:136)

The above example shows that the domain of application of tone deletion and spread increases with speech rate. In slow speech, no tone is deleted. In normal speech, the tone of the second syllable, but not the third, is deleted. In fast speech, the tones of the second and third syllables are deleted, and the initial M tone spreads to them.

Another similar case involves the devoicing of high vowels in Japanese. This process applies word-internally, and between a voiceless consonant and a pause (Hasegawa, 1979) and has been argued to be a phonological process (Tsuchida, 1997; c.f. Jun and Beckman, 1993). It can be seen in (14), where the first and third syllables of the word for ‘season’ have devoiced vowels, with the first being devoiced between two voiceless consonants, and the second being devoiced between a voiceless consonant and a pause.

(14) ki̥setu̥

‘Season.’ (Hasegawa, 1979; as reported in Kaisse, 1985:25)

This process is usually word-internal, but in fast speech, the domain of high vowel devoicing expands to include the initial consonant of the following word. This means that if a vowel is in between two voiceless consonants across a word boundary in fast speech, high vowel devoicing applies. This can be seen in that the final vowel of the word *iku* is devoiced in (15).

(15) tokyo e iku̥ hito
Tokyo to go person
‘The person who goes to Tokyo’ (Hasegawa, 1979; as reported in Kaisse, 1985:25)

Another instance in which speech rate apparently widens the domain of application of a phonological alternation is found in Modern Hebrew, as reported by Bolozky (1977). In Modern Hebrew, the vowel [e] is often deleted when roots are preceded by a clitic in normal speech:

(16) [a + yeladím] → [ayladím]
‘The children.’

In fast speech, a similar process occurs in multiple environments, not just between clitics and roots.[3]

(17) [ʔéyfo # $\boxed{\text{amaxbéret}}$ # ʃelì] → [ʔéyfo $\boxed{\text{amaxbért}}$ ʃelì]
‘Where is my notebook?’ (Bolozky, 1977:231)

This [e]-deletion is highly likely to be phonological, given that it interacts transparently with stress placement. When [e]-deletion would create a stress clash, then it is either blocked (18), or it causes stress retraction to the previous syllable (19).

---

[3] Bolozky (1977) writes that this process occurs to avoid stress lapses, but Bolozky and Schwarzwald (1990) provide several examples in which the process does not appear to avoid apply in order to avoid stress lapses.

(18) [$\boxed{\text{atá}}$ # $\boxed{\text{mevín}}$ # otí] → *[$\boxed{\text{atá}}$ $\boxed{\text{mvín}}$ otì]

‘Do you understand me?’

(19) [$\boxed{\text{atá}}$ # $\boxed{\text{mevín}}$ # otí] → [$\boxed{\text{àta}}$ $\boxed{\text{mvín}}$ otì]

‘Do you understand me?’ (Bolozky, 1977:231)

The interaction between [e]-deletion and stress placement in (18)-(19) suggests that, at the level at which stress is assigned, the vowel [e] is categorically absent from the phonological representation.

A final instance of rate-conditioned domain widening is found in the application of vowel coalescence in Mandar, an Austronesian language of Indonesia, as described in Brodkin (2022). In Mandar, when a word ends in two adjacent vowels, the two vowels occupy different syllables when the word occurs at the right edge of a maximal ϕ-phrase. Additionally, the first of the two vowels bears stress, as seen in (20). When the same word is non-final in a maximal ϕ, the two word-final vowels coalesce into one. The phonological status of this coalescence can be seen in that it causes stress to shift back one syllable, and the newly stressed vowel undergoes lengthening (21).

(20) [$\boxed{\text{ma.ˈlaː.iː}}$ $]_{\phi}$ =mi mwaˈne.na iˈniː.naː
return =NOW.AGR husband NAME
‘Nina’s husband came home.’

(21) [$\boxed{\text{ˈmaː.le}}$=mi iˈniː.naː $]_{\phi}$
return=NOW.AGR NAME
‘Nina came home.’ (Brodkin, 2022:18)

At normal rates of speech, vowel coalescence is restricted to non-ϕ-final environments, meaning that the verb ‘return’ in (20) cannot undergo coalescence and subsequent stress shift. However, at faster rates of speech, coalescence is possible even in this position, as seen in (22). In this case, a faster speech rate widens the domain of application of vowel coalescence.

(22) ˈmaː.le=mi mwaˈne.na iˈniː.naː
return=NOW.AGR husband Nina
‘Nina’s husband came home.’ (Brodkin, p.c., 2022)

The examples outlined here from Mandarin Chinese, Nantong Chinese, Japanese, Modern Hebrew, and Mandar provide evidence that speech rate can influence phonology by apparently expanding the domain within which a phonological process may apply. That is, as speech rate increases, so do the boundaries across which a phonological alternation’s conditioning environment may be defined. For example, in Mandarin Chinese, Tone 3 sandhi applies across larger prosodic boundaries in fast speech, and in Japanese, high vowel devoicing applies across word boundaries, not just word-internally. Another rate-influenced alternation in Italian is less straightforwardly analyzable in terms of domain widening.

5.2.2 Non-domain widening

In Italian, as described in Nespor (1987), identical vowels across a word boundary often undergo a process of ‘vowel degemination’ in fast speech in Italian. This process, shown in (23), deletes the second of two identical vowels across a word boundary, and it may cross relatively large prosodic boundaries.

(23) /mólto offensívo/ → [móltoffensívo]

'Very offensive.' (Nespor, 1987:71)

The reason to believe that it is the second vowel that deletes is that vowel degemination may not apply if the second vowel is stressed:

(24) Dicono che **mangiava álghe** e nient'altro (*mangiavalghe 'ate seaweed')

'They say that he ate seaweed and nothing else.' (Nespor, 1987:73)

Vowel deletion has the potential to create stress clashes, which are highly marked in Italian. In these cases, deletion of one of the vowels at the word edge would result in two adjacent primary stresses:

(25) /pianterá arbústi/ → *[pianterárbústi]

'He will plant bushes.' (Nespor, 1987:74)

Stress clashes are disallowed in Italian, and there is a separate phonological process that repairs them: When two primary stresses are adjacent, the first primary stress retracts to the closest stressed syllable to the left, even if that syllable would have normally received secondary stress. Importantly, in cases like (25) where fast-speech vowel deletion creates a stress clash, the separate phonological process of stress retraction occurs.

(26) /pianterá arbústi/ → [pianterárbústi] → [piánterarbústi]
‘He will plant bushes.’ (Nespor, 1987:74)

In other words, the fast-speech process of vowel deletion feeds the regular phonological process of stress retraction, meaning that in Italian, the fast-speech process vowel degemination is demonstrably phonological because it may feed a separate phonological process.

Notably, the interaction of speech rate and phonology in Italian is not straightforwardly analyzable as the widening of the prosodic domain of application of a categorical phonological process. This is because there are some word-internal cases in which vowel degemination is blocked, even in fast speech:

(27) Ha tante idée/*idé ma non conclude mai niente
‘He has many ideas but never accomplishes anything’ (Nespor, 1987:71)

It is not the case, then, that vowel degemination across word boundaries is simply the application of a word-internal phonological process across word boundaries. Instead, it appears to be a phonological process whose application is influenced by speech rate. This is not to say that it is an entirely ‘new’ phonological process, though, since there are many other rules that cause the deletion of one of two adjacent vowels in Italian (Nespor, 1987:70), such as vowel deletion in determiners before vowel-initial words ( ).

For example, the vowel

(28) la elica → l'elica

'The propeller.' (Nespor, 1990, as reported in Garrapa et al, 2021:4)

So, it appears that the fast-speech process of vowel degemination in Italian is not as amenable to an analysis in terms of a widened domain of application of a phonological process. Instead, it appears to be a phonological process whose application is more or less directly conditioned by speech rate.

5.2.3 Interim Review

In this section, I have outlined six cases in which speech rate has been shown to have an influence on the application of phonological processes. In Mandarin Chinese, Nantong Chinese, Japanese, Modern Hebrew, and Mandar, faster speech correlates with a wider domain of application of an extant phonological process: In each case, an alternation that occurs within a certain phonological domain applies across larger domains as speech rate increases. Another case of speech-rate-influenced phonology was shown for Italian, but this case is different in that it cannot simply be analyzed as the widening of the domain of an exceptionless phonological process.

At this point, it is possible to make a point that I believe is highly important: In all of these cases, the phonological alternation whose application is affected by speech rate is already an existing phonological process in the language in question, or is at the very least an analogue to a family of existing processes. For example, Italian vowel degemination, which applies vowel deletion in a new phonological context, can be thought of as an analogue to existing deletion rules. In this case, an unfaithful phonological mapping in one environment (rate-independent vowel deletion) is generalized to a new environment (rate-dependent vowel degemination) in a way similar to the ‘free rides’ of McCarthy (2005). In other words, it does not appear to be the case that speech rate may create entirely new phonological alternations; instead, it appears to ‘recycle’ existing ones. The observation that all of the processes conditioned by speech rate are independent phonological processes with a life of their own may not be surprising, but it is not logically necessary. One might imagine, for example, that fast speech could introduce a process of neutralization that is categorical but absent in slow speech. However, the (admittedly small) set of processes surveyed here does not find such application at fast speech rates of a phonological process absent at other rates, and I do not know of any process with these characteristics. This point is important because the apparent restriction of speech rate to affecting only independent phonological changes suggests that speech rate *interacts* with a language’s phonology, but that it does not *change* a language’s phonology; it only makes more likely the application of some extant phonological rule.[4]

I will argue in the following section that this characteristic of speech-rate-

---

[4] This is not to say that the means of interaction between speech rate and phonology in all of these cases is identical, but rather that it does not appear to be the case that speech rate directly changes phonology. In fact, the domain-widening cases lend themselves rather well to an analysis in terms of production planning, such as that outlined earlier for t/d-flapping in English (Wagner, 2012; Kilbourn-Ceron et al., 2016; Kilbourn-Ceron, 2017). The basic idea is that, as speech rate increases, the amount of material in a planning window increases, leading to application of the process in question across larger boundaries. However, as discussed earlier, the SMPM case is not amenable to a production planning account, and the other speech-rate-driven alternation in Italian is likely not explainable in these terms, either. It is on the basis of these cases that the forthcoming model is developed.

influenced phonological alternations motivates a view in which speech rate is an extra-grammatical factor not taken into account in the phonology proper. However, given its influence on phonology, it necessitates a model of the phonology-phonetics interface in which phonetic factors, among which is speech rate, may nonetheless influence which phonological candidate makes it to the surface. I will outline one such model, illustrate it through an analysis of mora deletion in SMPM, and then show that it can be applied to other cases in which phonetic factors are argued to trigger phonological changes.

5.3 Modeling the interaction

The apparent influence of speech rate on the application of phonological alternations necessitates a formal phonological framework that is able to account for it. Under an approach that differentiates phonology from phonetics, which I argued for in §2, there are several ways in which speech rate might be allowed to influence phonology: The first is to say that speech rate is directly taken into account in the phonological grammar, and the second is to relegate speech rate to the phonetic system but to allow a certain amount of interaction between phonology and phonetics. I will argue against the first approach and in favor of the second on the basis of the apparent indirect nature of speech rate's interactions with phonology.

5.3.1 A direct influence of speech rate

One potential way to model the effect of speech rate on phonological grammar is to allow speech rate to directly affect phonological computation. This could take the form of allowing for multiple phonological grammars according to speech rate, or by allowing speech rate to affect phonological constraint ranking/weight, as proposed in Coetzee (2016). In an account like this, the ranking or weight of phonological constraints is influenced by a numerical scaling factor tied to speech rate. This factor scales faithfulness constraints down as speech rate increases, capturing the intuition that faster speech rates result in more faithfulness violations due to processes like elision and assimilation (many of which are arguably non-phonological). I would like to argue that any model in which speech rate influences phonological constraint rankings necessarily predicts the existence of phonological processes at one rate of speech that are categorically absent from another.

In order to illustrate this point, it is useful to consider approaches to modeling the effects of speech style (e.g., formal vs. casual speech) on phonological processes, some of which allow speech style to directly affect phonological constraint ranking. There are several implementations of this approach, including a categorical model like that of Van Oostendorp (1997) in which faithfulness constraints are more highly ranked in formal speech than in casual speech, leading to categorically distinct grammars for different speech registers. In a gradient approach like that of Boersma and Hayes (2001:Appendix C) or Coetzee and Pater (2011:426-427), a 'style' factor gradiently affects the weight of phonological constraints in Stochastic OT or Noisy HG, leading to a change in the rate of application of some phonological processes in proportion with the level of formality of the speech situation. What is important here is that both of these models must be (and are) able to account for instances of categorical phonological distinctions driven by speech style. Such an alternation can be found in register-sensitive consonantal alternations in Samoan. In Samoan, the consonants /t/, /n/, and /r/ contrast with /k/, /ŋ/, and /l/, respectively, in a register associated with Western activities (Mosel and Hovdhaugen, 1992; Duranti, 1981:360). These contrasts are neutralized in outside of this register.

Both Van Oostendorp's and Boersma and Hayes' models must be able to handle this sort of alternation, and it appears that they can—for Van Oostendorp, a 'formal speech grammar' would rank highly the faithfulness constraints motivating the maintenance of consonant place contrasts. For Boersma and Hayes, the 'style' factor would have to have a large enough effect on the same phonological constraints' weight that it would trigger the categorical or near-categorical maintenance of place contrasts in formal speech. In other words, both of these approaches must allow for speech style to cause the categorical or near-categorical application of a phonological process in one speech register but not another.

It is precisely this fact—that allowing factors to directly influence phonological constraint ranking allows for categorical differences in phonological rule application—that speech rate should not be allowed to directly influence phonological constraint ranking or weighting.[5] If it were allowed to, then we would predict the occurrence of phonological alternations or contrasts that occur only at fast rates of speech (or, conversely, only at slow rates of speech). However, as noted above, the small list of speech-rate-driven phonological alternations contain no such difference; all phonological

---

[5] The fact that some register differences do have this categorical profile means that this type of approach might be appropriate for register-conditioned processes, though.

alternations driven by speech rate are independently attested in the languages' phonological grammar, and increased speech rate simply increases the rate of application of the process. In this light, the effect of speech rate on phonological grammar should not be modeled as direct manipulation of constraint ranking or weighting; instead, the effect should be indirect. In the following section, I describe a model that allows speech rate (and other extra-grammatical factors) to influence phonology, but only indirectly.

5.3.2 Phonetically-Informed Candidate Selection (PICS)

As argued throughout this dissertation, any model of the phonology-phonetics interface must incorporate two crucial characteristics: First, it must encode a distinction between two levels of representation broadly corresponding to what is considered phonology and what is considered phonetics. Second, it must allow for interaction between these two levels of representation, such that factors attributed to the phonetic system may influence phonology. However, as argued above, this interaction must be indirect—allowing speech rate to directly influence the process of phonological evaluation predicts that there should be rate-driven alternations that are categorically absent from one rate of speech and present only in another.

In this section, I outline a model with these features and illustrate its characteristics by modeling the interaction of speech rate and mora deletion in SMPM. The basics of the system are that it involves two separate levels of representation for phonology and phonetics, as argued for in Chapter 2, but that it also allows the phonetic system to have a say in which phonological output makes it to the surface. Specifically, the phonological grammar supplies multiple, ranked outputs to the phonetic system (Coetzee, 2006), and the phonetic system determines which of these outputs makes it to the surface by evaluating each phonological output relative to the current speech conditions. I will describe and illustrate the system in more detail below.

The first necessary point is to encode a distinction between phonological and phonetic computation, and to define their respective roles. As argued for in Chapter 2, phonological computation involves units of representation and constraints that are defined at a relatively coarse-grained level and do not make direct reference to the fine-grained phonetic details of the sounds involved. Phonetic computation, on the other hand, is concerned with the minute details of the sounds' realization. I broadly adopt the approach in much phonological literature (e.g., Zsiga, 2000) that the output of phonological computation is the input to the phonetic system, which converts phonological units into fine-grained, physical events that occur in time and space. However, in most feed-forward models of phonology, the phonetic component of the speech system simply converts phonological representations into phonetic representations, and its job stops there. Instead, I argue that the phonetic system is not simply translational or implementational, but that it also acts as a filter over phonological outputs. Specifically, the phonetic system determines which of a set of ranked phonological outputs survives to the surface. In this way, phonetics is more than merely an interpretive system—it also involves a selectional component. This process of phonological computation and phonetic filtering is schematized below in Figure 5.1.

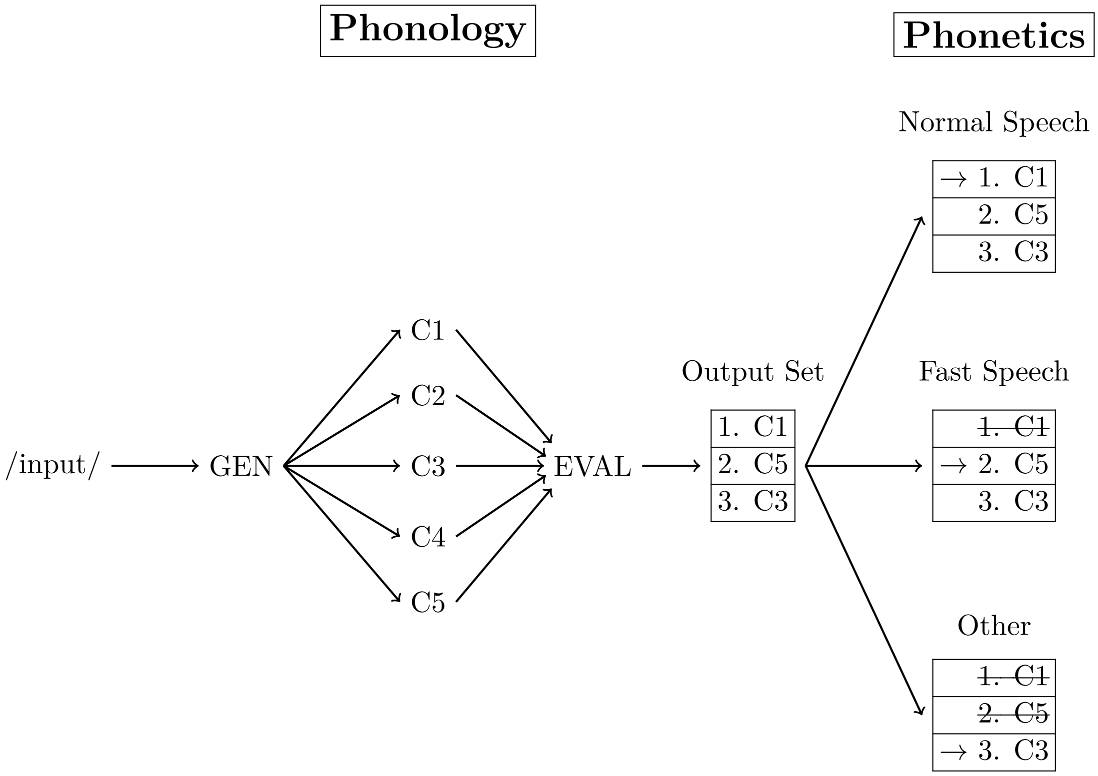


Figure 5.1: Illustration of Phonetically-Informed Candidate Selection (PICS)

The model proposes that phonological and phonetic computation interact in something like the following way: First, phonological computation proceeds as proposed in Optimality Theory (Smolensky and Prince, 1993). For a phonological input, GEN generates a set of output candidates (C1-C5 in Figure 5.1), which are all evaluated by EVAL based on their violation profile relative to set a of ranked (or weighted), violable constraints. However, one point of divergence from classic OT is that this system adopts the proposal that the phonological grammar does not yield just one optimal output candidate. Instead, it provides a set of candidates ranked according to their performance in phonological evaluation, as argued for in Coetzee (2006). Crucially, though, only candidates that are relatively well-formed by virtue of not violating high-ranked constraints are included in this set—others are completely ruled out. This ranked set of potential phonological outputs is passed through the phonetic system, which converts the phonological units into physical configurations, taking into account factors like speech rate and coarticulation. Under most speech conditions, the optimal phonological candidate survives phonetic evaluation and is realized in the speech signal. However, when the physical realization of a phonological output is too marked in the current speech conditions, then the phonetic system rejects that candidate and advances to the second-ranked phonological output. If that candidate is well-formed enough relative to the first candidate, then it survives phonetic evaluation and is implemented. Finally, in yet other conditions, the first two candidates in the list might be ruled out, in which case the next-most-harmonious candidate is evaluated.[6] In this way, phonetics influences which phonological candidate survives to the surface, but does not directly influence the computation by which those candidates are generated in the first place. To illustrate the workings of this model, I will walk through examples of this type of phonological and phonetic evaluation as it is proposed to apply to laryngeal reduction and mora deletion in SMPM.

---

[6] Another implementation could be for phonetic evaluation to take place in parallel, with each phonological output being evaluated at the same time by the same constraints, and the optimal candidate implemented. To maintain the notion of a ranked order of phonological ouputs in this type of implementation, some advantage would be given to candidates according to their ranking in the phonological output set.

5.3.3 Application to mora deletion

The first ingredient is to outline the constraints involved in the phonological evaluation of laryngealized roots. With the understanding that the default case is that in which mora deletion does not occur, what is needed is to weight a constraint prohibiting mora deletion higher than a constraint driving it. A constraint prohibiting mora deletion is MAX[μ], which conflicts with a constraint on the sequencing of modal and non-modal phonation, formalized in terms of an AGREE constraint relativized to phonation type:

- MAX[μ]: Assign one violation for every mora in the input that not have a correspondent in the output.
- AGREE[PHON]: Assign one violation for every pair of adjacent moras of different phonation types.[7]

Weighting the constraint MAX[μ] above AGREE[PHON] results in retention of both moras (29).

---

[7] This constraint is relatively abstract and might be better understood as a stand-in for a family of contextual markedness constraints on the sequencing of modal and non-modal phonation along the lines of the proposed optimal and sub-optimal sequencings described in Silverman (1997).

(29)

/CVˀV/	MAX[μ] WT=8	AGREE[PHON] WT=5	**Harmony score**
☞ a. CVˀV		-1	-5
b. CVˀ	-1		-8

Additionally, the blocking of other phonological changes can be achieved by giving higher weights to constraints that prevent other potential outcomes, such as the deletion of the laryngeal feature or its spread to the adjacent vowel. These constraints are formalized as follows:

- MAX[ˀ]: Assign one violation for every laryngeal feature in the input that is absent in the output.
- IDENT[PHON]: Assign one violation for every segment whose laryngeal specification α in the input is β in the output.

Giving these two constraints a higher weight than those previously discussed blocks other candidates.

(30)

/CV^{ʔ}V/	Max[ʔ] wt=20	Ident[phon] wt=20	Max[μ] wt=8	Agree[phon] wt=5	**Harmony score**
☞ a. CV^{ʔ}V				-1	-5
b. CV^{ʔ}			-1		-8
c. CV	-1		-1		-28
d. CV^{ʔ}V^{ʔ}		-1			-20
e. CVV	-1				-20

So far, this phonological evaluation outputs only one candidate, Candidate A in (30), which retains both moras. And, as argued earlier, there are reasons to prohibit speech rate from influencing phonological constraint ranking in order to make Candidate B, the candidate to which mora deletion has applied, the optimal output. Instead, I adopt Coetzee's (2006) rank-ordering model of EVAL, in which phonological computation outputs not just one optimal candidate, but an ordered list of candidates. In this approach, multiple candidates may survive phonological computation if they violate only lowly-ranked constraints. When implemented in a model with weighted constraints like Harmonic Grammar, this condition can be restated as allowing only candidates with a sufficiently high harmony score to be potential outputs. If, for the purposes of illustration, we set an arbitrary harmony cutoff at -17, then Candidates A and B in (30) would be potential outputs, but Candidates C-E would not.[8]

---

[8] It is likely that only including candidates above a harmony cutoff makes different predictions than

Candidates A and B, then, are the output set. They are ranked based on their harmony score: The one with the highest harmony score is ranked first, and the one with the second-highest harmony score is ranked second.

(31) Output Set of phonological evaluation of /CVˀV/

1. CVˀV
2. CVˀ

This output set is then passed on to the phonetic component, which converts phonological outputs into phonetic units in time and space. Because the phonetic system must be able to evaluate candidates' phonetic markedness, I will model the conversion of phonological units to phonetic units by means of interacting phonetic mapping constraints like those in Boersma (2011), which enforce correspondence between phonological units and their target phonetic values. The model given below is highly simplified and is not meant to be taken as a literal claim about the process of phonetic evaluation; instead, it is meant to show in a simplistic way how the phonetic system can choose between candidates.

The relevant mapping constraints require laryngealized and modal vowels each to correspond to a specific value, and they assign violations if each vowel type does not

---

only including candidates that violate constraints ranked below a certain cutoff point, as proposed in Coetzee (2006). However, I illustrate the process using a harmony cutoff because it enables us to rule out Candidate C in (37), which is not ruled out under a constraint ranking cutoff. The reason for this is that the constraints violated by this candidate are violated by other output candidates, so they would have to be below the ranking cutoff point. This problem could be solved by using strict constraint ranking in classic OT and ranking a conjoined constraint (Smolensky and Prince, 1993) made up of two lower-ranked constraints above the cutoff point, which would rule out Candidate C. However, since this implementation would require porting over Harmonic Grammar analyses from previous chapters in a different format, I proceed with a Harmonic Grammar implementation and a harmony cutoff instead.

reach that value. Though the articulatory, acoustic, and perceptual cues to laryngeal contrasts are multi-dimensional and not always linearly related to each other, I will assign numerical target values to each laryngeal category along a uni-dimensional continuum. Grounding the continuum in the view of phonation types occurring on a continuum of glottal constriction, with breathy voice being less tense, creaky voice being more tense, and modal voice being in the middle of the two (Gordon and Ladefoged, 2001), we can set target values for the phonation types. In this case, -1 is the target value for breathy voice, 0—the default value—is for modal voice, and 1 is the target value for laryngealized voice. These target values can be encoded into the mapping constraints for laryngealized and modal vowels, given below.

- MAP[$V^ʔ$]: A laryngealized vowel corresponds to a glottal constriction of the value 1.
- MAP[V]: A modal vowel corresponds to glottal constriction of the value 0.

The mapping constraints here require laryngealized and modal vowels each to correspond to a specific phonetic value, and they assign violations if each vowel type does not reach that value. Let us suppose that speech rate determines the amount of time in which the output of the phonetic grammar may be produced. In order to produce a faithful realization of a target laryngeal state, the corresponding segment must be allocated a certain amount of time, and allocating any less time to it results in an actual realization that is less accurate. Once again for illustrative purposes only and not as a claim about time requirements in the real world, this time requirement will be set to an arbitrary value of 50 ms.

With these values set, it is now possible to evaluate violations of the mapping constraints. Given that modal voice is the default laryngeal setting for vowels, the default value of any vowel will be 0. In order to produce laryngealized voice, enough time must be devoted to a vowel for its value to raise from 0 to 1. In this simplified example, with the target realization of laryngealized voice set to 1, allocating a laryngealized vowel 50 ms allows its value to reach 1. However, if 40 ms are allocated to it, then its value may only reach 0.8. If 30 ms are allocated to this segment, its value will only be 0.6, and so on and so forth. Given that modal voice is default, it is usually the case that no specific amount of time must be devoted to it in order for its 'target' state to be reached. However, in CV^{ʔ}V words, where a modal vowel is immediately preceded by a laryngealized vowel, it is possible for the laryngeal state associated with the modal vowel to be too constricted because of the preceding vowel—in other words, there must be enough time for the laryngeal state to transition from laryngealized back to the default, modal setting. In this simplified example, we may suppose that the time it takes to get from the default state (0) to the laryngealized target state (1) is the same as the amount of time that it takes to get from the laryngealized target state (1) back to the default state (0).[9] This means that if 50 ms are allocated to the initial laryngealized vowel, the laryngeal value will reach 1, and 50 ms will be required for the value to return to 0. If 40 ms are devoted to the laryngealized vowel, achieving a value of 0.8, then 40 ms will be required to return to 0. If 30 ms are allocated to the laryngealized vowel, achieving

---

[9] This is almost certainly incorrect, since there are directional asymmetries in other glottal changes like pitch excursions—raising pitch takes slightly longer than lowering pitch (Xu and Sun, 2002).

a value of 0.6, then 30 ms will be required to return to 0, and so on and so forth.

The final step is to determine how constraint violations are computed and evaluated. In this case, it makes sense that the farther the actual value of a vowel is from its target value, the stronger the violation that candidate incurs on the relevant mapping constraint. In order to implement this intuition, we may impose another arbitrary value, saying that for every 0.1 units a vowel's actual laryngeal value is from its target value, a violation is assigned according to the constraint's weight. Finally, it is necessary to set a markedness value that will cause the phonetic system to reject a given candidate. Here, we will set this 'threshold' of ill-formedness to a harmony score of -5: If a phonological output cannot be implemented without a harmony score greater than -5, then that output is rejected. All of these values are summarized in Table 5.1.

	Target State	Time to Target	Violations	Threshold
$V^ʔ$	1	50 ms	-1 for every 0.1 deviation	Harmony score ≤ -5
V	0	0-50 ms	-1 for every 0.1 deviation	

Table 5.1: Relevant values for realization of modal and laryngealized vowels

At this point, the assumption that speech rate is taken into account in the phonetic grammar becomes relevant. Let us suppose that at a sufficiently slow rate of speech, the optimal phonological output (Candidate 1 from (31)) is passed through phonetic evaluation. At a sufficiently slow rate of speech, Candidate 1 may be mapped to a phonetic output that does not violate either mapping constraint, as in (32). This is because 50 ms can be allocated to the first vowel, allowing it to reach a target value of 1, there are 50 ms left for the following modal vowel to return to a value of 0. At a slightly faster rate of speech, there are some mapping constraint violations, but are not very severe, as in (33). In both of these cases, the optimal phonological output is implemented because its harmony score is above -5.

(32) Phonetic evaluation of vocalic portion of /CV^{ʔ}V/ in 100ms.

/CV^{ʔ}V/	Map[ʔ] wt=1	Map[V] wt=1	**Harmony score**
a. V^{ʔ}=1, V=0	0	0	0

(33) Phonetic evaluation of vocalic portion of /CV^{ʔ}V/ in 75ms.

/CV^{ʔ}V/	Map[ʔ] wt=1	Map[V] wt=1	**Harmony score**
a. V^{ʔ}=.75, V=0	-2.5	0	-2.5

In a case in which a faster speech rate requires that the vocalic portion of the word be produced in 50 ms, each laryngeal state is allocated 25 ms. Under the hypothetical time constraints established above, this means that the laryngeal state for the laryngealized vowel only reaches 0.5. In this case, the Map[ʔ] constraint is violated to such an extent that it causes the candidate's harmony score to surpass the threshold of ill-formedness.[10]

---

[10] It is worth noting here that this approach is relatively categorical and only sets one threshold—it could be the case that the threshold's value is different for different constructions. Also, considering

(34) Phonetic evaluation of vocalic portion of /CV^{ʔ}V/ in 50ms.

/CV^{ʔ}V/	Map[ʔ] wt=1	Map[V] wt=1	**Harmony score**
☹ a. V^{ʔ}=0.5, V=0	-5	0	-5
b. V^{ʔ}=0.6, V=0.2	-4	-2	-6

At this point, the output of the phonology cannot pass through the phonetic system without reaching a certain level of ill-formedness. As shown in Candidate B in (34), allocating more time (say, 30 ms) to the laryngealized vowel would lower violations of Map[ʔ] by allowing the value corresponding to V^{ʔ} to reach 0.6. However, this would only leave 20 ms for the modal vowel, meaning the value could only lower to 0.2, triggering violations of Map[V]. In a case like this, where the optimal phonological output cannot pass through phonetic evaluation without incurring severe violations, it is rejected outright. When this happens, the next candidate from the output set (Candidate 2 from (31), repeated below) is put through phonetic evaluation.

(35) Output Set of phonological evaluation of /CV^{ʔ}V/

1. CV^{ʔ}V
2. CV^{ʔ}

---

the quantitative nature of phonetic representations, it is also likely that the violations assigned by the mapping constraints are more gradient and noisy than represented here. In either case, the most important point is that it is the avoidance of severe phonetic constraint violations under the pressure of fast speech that is driving the rejection of the phonological output.

When this candidate is evaluated, there is no need to phase two phonation types, and the laryngealized vowel is allocated the entire 50ms, allowing for it to reach its target state and surface faithfully. This means that it does not incur violations of either mapping constraint, and its harmony score does not dip below the threshold. As a result, it survives to the output.

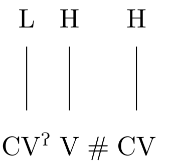
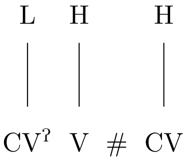
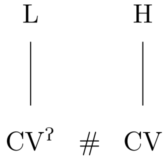
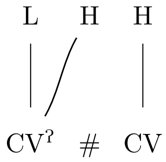
(36) Phonetic evaluation of vocalic portion of /CV^{ʔ}/ in 50ms.

/CV^{ʔ}/	MAP[ʔ] WT=1	MAP[V] WT=1	**Harmony score**
a. V^{ʔ}=1	0	0	0

This ability of the phonetic system to choose between candidates, then, gives it a way to interact indirectly with phonology. It allows for an extra-phonological factor to affect the choice of phonological output, but only through the limited mechanism of determining that a given phonological output is too marked when evaluated relative to the current speech conditions. This limited type of interaction has the convenient property of allowing for phonetic factors like speech rate to indirectly condition the application of an unrelated phonological process, which I will demonstrate below by walking through this framework's modeling of the interaction between mora deletion and tone sandhi in SMPM.

In this case, we may examine the tableau for a laryngealized root with an L-H melody in the conditioning environment of tone sandhi. Using the constraint weights from Chapter 2 and those established earlier in this section, the constraints MAX[μ] and MAX[T] are weighted higher than AGREE[PHON], meaning that the optimal candidate in (37) is Candidate A, the one without deletion of the second mora or modification of any tones.[11] However, Candidate B, to which mora deletion has applied, has a harmony score above the cut-off point of -17, meaning that it is also a possible output. As a result, both Candidates A and B are potential, ranked outputs of phonological evaluation.

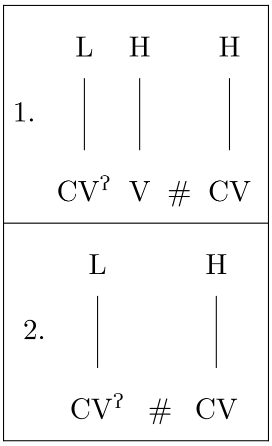
(37) Phonological evaluation

	MAX[μ] WT=8	MAX[T] WT=8	AGR[PHON] WT=5	OCP[H] WT=5	*CONT. WT=5	**Harmony** **score**
☞ a. 			-1	-1		-10
☞ b. 	-1	-1				-16
c. 	-1			-1	-1	-18

---

[11] IDENT[PHON] and IDENT[T] are excluded for reasons of space. However, including them with the weights they receive in previous tableaux successfully rules out candidates that involve the changing of laryngeal or tonal features.

(38) Output Set of phonological evaluation of /$CV^{ʔ}V$/



As discussed earlier, the output set is passed along to the phonetic grammar, which converts each candidate by means of its own constraints. At a sufficiently slow rate of speech like in (33), Candidate 1 from (38) survives to the surface. However, at a fast rate of speech like in (34), the aforementioned conflicts between the values of the laryngealized vowel and the following non-laryngealized vowel mean that there is no potential phonetic output for Candidate A that does not surpass the threshold of ill-formedness. As a result, the phonetic grammar rejects it, and the next-ranked candidate from the output set, Candidate 2 from (38), is passed through phonetic evaluation. Because the phonologically reduced candidate is passed to the phonetics, it survives phonetic evaluation as shown in (36). It is crucial to note here that Candidate B, the candidate that is ultimately the output, has undergone rise flattening. The reason for this is that Candidate 2 was derived by the phonological grammar, where the constraints that drive rise flattening are present alongside the constraints that drive mora deletion. As a result, candidates that undergo mora deletion are also subject to the independent phonological constraints that drive rise flattening.

In this way, the ability of the phonological grammar to provide multiple, ranked outputs, and the ability of the phonetic system to choose between those outputs, provide a straightforward mechanism to understand the role of speech rate in mora deletion in SMPM and its influence on the application of rise flattening: In fast speech, the phonetic realization of phonologically unreduced laryngealized roots is more likely to surpass the threshold of ill-formedness. This means that the phonetic system is more likely to reject bi-moraic phonological outputs at fast rates of speech, and that as a result, candidates to which mora deletion has applied are more likely to be passed through the phonetic module. Because those phonologically reduced candidates are derived in the phonological grammar proper, they are also subject to independent phonological constraints, such as those driving rise flattening.

5.3.4 Interim review

So far in this chapter, I have motivated a view of the role of speech rate in phonological grammar as indirect rather than direct. Specifically, I have proposed that speech rate does not influence phonological computation itself, but instead is taken in to account during phonetic implementation and evaluation. The phonetic system evaluates a ranked set of potential phonological outputs, and it rejects the optimal output if it is too marked in the current speech conditions. In this way, a phonetic factor like speech rate is maintained as extra-phonological, but is given an indirect mechanism by which to influence phonological computation. As a result of this set-up, speech rate does not trigger the application of phonological processes that are completely ‘new’ in a language. Now, there is no mechanism in the system that *forces* this outcome, but it makes sense considering the mechanics of determining a cutoff point: By virtue of the phonetics only being able to influence the choice between a set of outputs that are not too marked, the candidates which it is able to choose from are those that have violated only low-weighted constraints in the phonological grammar. Because these constraints have a low weight, they are very likely to be violated by the optimal phonological output in other configurations, meaning that the phonological change that applies to the second-ranked candidate in an output set is one that almost certainly applies elsewhere in the language’s phonology. For example, the constraint Max[μ] is violated by some shortened function words and also by the first member of a N-N compound in SMPM in slow speech, so violations of it must necessarily not be so drastic that they trigger exclusion of a candidate from the output set, meaning that phonologically-reduced laryngealized roots are possible outputs. In the same way, the constraint Max[V] is violated in slow speech in Italian (e.g., /la elica/ → [l’elica] ‘the propeller;’ Nespor, 1990), and this constraint is also the one violated in the fast speech process of vowel degemination. So, by only allowing outputs that violate low-ranked/low-weighted constraints, this system captures the apparent tendency for rate-dependent processes to exist in a rate-independent form elsewhere in a language’s phonology.

It is important to note that the prediction that phonetic factors like speech rate will not introduce new phonological alternations is not *forced* by the current setup, and the reason for this is that the mechanics of determining the cutoff point in a language’s constraint ranking/weighting is not clearly defined. Consider, for example, the following toy language, in which NoCoda (assign one violation for every coda consonant) is ranked above both Dep, and Dep is ranked above Max. In this case, consonant deletion occurs to avoid a violation of NoCoda, and vowel epenthesis does not occur because Dep is ranked above Max, as shown in (39). So, because of the ranking of faithfulness constraints, this is a language where epenthesis is not an attested phonological process. Now, without a clear definition of where the cutoff point in a language's phonological constraint ranking lies, it is in principle possible for the cutoff point to be made above Dep, as indicated in (39) by the double line between NoCoda and Dep.

(39)

CVC	NoCoda	Dep	Max
☞ a. CV			$*$
b. CVCV		$*!$	
c. CVC	$*!$		

If the cutoff point is set above Dep, then both Candidate A and Candidate B would be included in the output set, as shown below:

(40) Output set from phonological evaluation in (39)

1. CV
2. CVCV

Because Candidate B from (39) is included in the output set in (40), it is in principle possible for it to be selected by the phonetic component in certain circumstances. This is important because it would constitute a case of phonetic pressures triggering the application of epenthesis, a phonological alternation that is otherwise unattested in the language in question, which is something I have argued is avoided by the PICS framework.

There are two potential responses to this point: The first is to say that it is simply a tendency and not a hard and fast rule that phonetic factors do not trigger the application of a phonological alternation otherwise not attested in a language's phonology, since the lack of a definition of where the cutoff point is made leaves open the possibility for otherwise unattested outputs to make it into the output set. The second possibility is to make an attempt to define the cutoff point in a way that ensures the exclusion from the output set of candidates to which otherwise unattested alternations have applied. One way to force this would be to state that the only way a phonological constraint may rank below the cutoff point is if the optimal output in some configuration (that is, the first-ranked, and not the second- or lower-ranked output in an output set) violates it. This condition would have the effect of ruling out Candidate B from (39) from being included in the output set: Because there is no optimal output in this toy language that violates DEP, then DEP cannot rank below the constraint cutoff. This proposal is congruent with Coetzee's (2006:379) note that only constraints ranked above the cutoff point can rule out candidates as ungrammatical (that is, impossible), meaning that candidates that violate them are unattested.

Because the first approach predicts only tendencies for phonetic factors to trigger the application of already-attested alternations, and because the second approach makes the stronger prediction of the two, I adopt the second approach and claim that the cutoff point in a language's constraint ranking is placed between those constraints that are violated by the optimal candidate in some configuration and those that are never violated by the optimal candidate, with the latter ranking above the cutoff point. This definition forces the outcome that the phonetic component may only choose between output candidates to which some extant phonological process has applied, and it precludes the possibility that the phonetic component may trigger entirely 'new' alternations. Of course, the real question here is an empirical one, which concerns whether or not clearly phonetic factors may indeed trigger an otherwise unattested phonological alternation. This question remains an open one.

Even though the modeling of phonetic evaluation in PICS is highly simplified, and even though the mechanism behind determining a cutoff point in a language's constraint ranking is not entirely self-evident, I would like to argue that the PICS model illustrates a way that we might conceive of phonetic influences on phonology. Specifically, the phonetic system is not simply interpretational, converting phonological units into phonetic units. Instead, it also contains at least one extra component, which is the ability to select between potential phonological outputs. There are certainly other ways to model this intuition, and it is not my claim that the PICS model is the be-all and end-all of frameworks of the interface. Instead, my hope is that illustrating how these interactions can be modeled will help to generate testable predictions about other ways in which phonetic pressures might interact synchronically with phonological candidate selection, leading ultimately to a more refined model. However, I do believe that an eventual, successful model will necessarily embody, at least to an extent, the characteristics that I have argued for throughout the dissertation. To that end, I devote the following section to comparing this framework to other proposed models that have the characteristics I have argued are necessary for any account of the phonology-phonetics interface, showing that this type of theory comparison can help to illustrate how different implementations of these features lead to different amounts of empirical coverage and predictions.

5.4 Theory comparison

As mentioned at the end of the last section, the characteristics I have argued must hold of any theory of the phonology-phonetics interface are rather broad: A successful framework must (1) allow different levels of representation for phonology and phonetics, and (2) allow at least some interaction between these two levels of representation. As can be seen in Figure 5.2, these characteristics are embodied by a number of frameworks, including the one just outlined above.

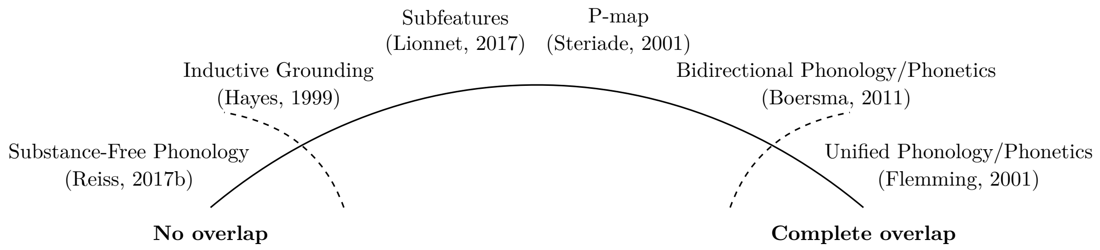


Figure 5.2: Continuum of frameworks of the phonology-phonetics interface

Though determining an end-all-be-all framework of the phonology-phonetics interface is far beyond the scope of this dissertation (hopefully this is evident from the multiplicity and time-depth of proposed models), I would like to spend some time comparing the proposed framework above with several others that embody the necessary characteristics I have argued for, namely Lionnet's (2016, 2017) subfeatural representations, Boersma and Van Leussen's (2017) multi-level, parallel constraint grammar, and the various approaches to constraint induction and/or ranking (Hayes, 1999; Steriade, 2001; Smith, 2004; Flack, 2007; Steriade, 2008). My goal in this comparison is not to argue for one framework over another, but rather to illustrate how theory comparison can help to elucidate recurring themes in and necessary features of an eventual, successful understanding of the phonology-phonetics interface.

5.4.1 Subfeatures

One framework that possesses the characteristics I argue are necessary is Lionnet's (2016, 2017) theory of subfeatural representations. This approach, built to account for multiply-triggered phonological phenomena like the low tone spread process described in Chapter 2, makes use of multiply-valued phonological features to capture the intuition that cumulative coarticulatory pressures can trigger changes in phonological feature values. The basic idea is that segments not only have feature values corresponding to [+F] and [-F], but can also have intermediate feature values. For example, if we redefine [+F] as [[1 F]],[12] and [-F] as [[0 F]], the subfeatural account allows for feature values between 0 and 1, such as [[0.4 F]] (c.f. Chomsky and Halle, 1968:165-170). These subfeatures are a phonological representation of expected phonetic coarticulation—speakers know, for example, that a [+high] vowel coarticulated with a [+low] vowel will be less 'high' than it would be otherwise, and this is represented in the phonological grammar by giving a [+high] vowel in this context the value of [[0.8 high]], for example. As might be extrapolated from this example, subfeatures are not a part of the underlying representation of a sound, but rather are computed in the phonological grammar. Because subfeatures can exist in the phonological grammar, there can be constraints that target them, and therefore alternations that apply only to segments that are already coarticulated with other segments.

An illustrative example is the one used to motivate and model the subfeatural account, namely doubly-triggered rounding harmony in Laal, an endangered isolate spoken by about 800 people in Chad (Lionnet, 2017). In this process, a round vowel triggers rounding of a preceding vowel if (1) they are of the same height and backness, and (2) the preceding vowel is also flanked by a labial consonant. This is shown in (41), where underlying /ɨ/ becomes [u] when it follows a labial consonant and precedes

---

[12] As in Lionnet (2017), subfeatural values are differentiated from binary features by double brackets.

a round vowel.

(41) /ɓɨ̀r-ú/ → [ɓùr-ú]
fish.hook-PL
'Fish hooks." (Lionnet, 2017:526)

If any of these characteristics are not met, then rounding harmony does not apply. For example, a final round vowel on its own does not trigger harmony (42-a), nor does a labial consonant on its own (42-b). Additionally, if the two vowels mismatch in height (42-c) or backness (42-d), then the rounding harmony does not take place.

(42) a. sə̀g-ó
tree.species-PL
'Trees (of a particular species)'

b. kə̀ə̀m-ə́
tree.species-PL
'Trees (of a particular species)'

c. ɓə̀r-ú
plant.species-PL
'Plants (of a particular species)'

d. píl-ù
mat-PL
'Mats' (Lionnet, 2017:527)

The logic of the subfeatural account is roughly as follows: when a [[0 round]] (that is, [-round]) vowel is adjacent to a labial consonant, it takes on a feature value of [[x round]] because of expected coarticulation with the labial consonant. Then, there is a separate rounding harmony process that targets vowels with an [[x round]] feature.[13] The [[x round]] vowel undergoes this harmony process, and becomes [[1 round]], which is the subfeatural equivalent of [+round]. This is illustrated in Figure 5.4, where ɨ^{b} represents the vowel [ɨ] with an [[x round]] feature.

---

[13] This is different from another, singly-triggered rounding harmony that applies in the context of a

Input	**Coarticulation**	**Harmony**
/ɨ/ = [[0 round]]	[[0 round]] → [[x round]]	[[x round]] → [[1 round]]

/ɓɨ̏r-ú/ ⟶ [ɓɨ̏^{b}r-ú] ⟶ [ɓɨ̏r-ú]

Figure 5.3: Visualization of doubly-triggered rounding harmony

Because an intermediate [[x round]] feature value is required for harmony to apply, both triggers must be present: If only the labial consonant is present, then coarticulation will result in an [[x round]] feature value, but there is no following round vowel to target the coarticulated [[x round]] vowel and trigger harmony.

This process can be derived by the interaction of faithfulness-like constraints and markedness-like constraints. Specifically, there are constraints requiring a [+F] segment to have a [[1 F]] feature value, and a [-F] segment to have a [[0 F]] feature value. These constraints, those that follow, and the OT tableaux are all minimally modified from Lionnet (2017:537-541).

- *[[>0rd]]/[-rd]: A vowel must not be higher than [[0 round]] on the subfeatural scale if it is specified as [-round].
- *[[<1rd]]/[+rd]: A vowel must not be lower than [[1 round]] on the subfeatural scale if it is specified as [+round].

These constraints, which serve to associate binary phonological features with subfeatural values at the end of a numerical scale, conflict with another constraint that penalizes vowels without a subfeatural value between 0 and 1 in specific contexts:

---

separate set of suffixes (Lionnet, 2017:532).

- *¬Coarticulated$_{\mathrm{B}\to ə/ɨ}$: A vowel ə/ɨ must be assigned a value of [[x round]] if it is adjacent to a labial consonant.

When the constraint driving coarticulatory [x round] values is ranked higher than the quasi-faithfulness constraints, then the vowels [ə/ɨ] receive an [x round] value when preceded or followed by a labial consonant:

(43)

/Bɨ/ ([-round])	*¬C$_{\mathrm{B}\to ə/ɨ}$	*[[>0rd]]/[-rd]	*[[<1rd]]/[+rd]
☞ a. Bɨ [[x round]]		*	
b. Bɨ [[0 round]]	*!		
c. Bɨ [[1 round]]	*!	*	

With these constraints and their ranking deriving [x round] values, another constraint can be introduced to drive vowel harmony. This constraint penalizes sequences of [x round] and [1 round] vowels of identical height and backness. It drives rounding harmony of vowels that would receive an [x round] feature, so long as it is ranked above IDENT[ROUND].

- *[[[$\geq$ x round]] -rd] [[[1 round]] +rd]/[+syll, α height, β front]:

  A [-round] segment whose subfeatural [[round]] value equals or exceeds x may not directly precede a [+round] segment whose subfeatural value is [[1 round]] in the ordered set of output segments that are [+syllabic] and share the same [height] and [front] specifications.

- Ident[round]:

  Assign one violation for every segment whose feature [α round] in the input is realized as [β round] in the output.[14]

When these two constraints are incorporated into the ranking as below, they derive doubly-triggered rounding harmony: The [0 round] vowel in Candidate A is ruled out by the constraint driving [x round] features, and the [x round] feature in Candidate B is ruled out by the constraint driving harmony of [x round] vowels. As a result, Candidate C, to which rounding harmony has applied, is the optimal candidate.

(44)

/ɓɨ̀r-ú/	*¬C$_{\text{B→ə/ɨ}}$	*[[$\geq$ x round]] [[1 round]]	Ident[round]	*[[>0rd]]/[-rd]
a. ɓɨ̀r-ú	*!			
b. ɓɨ̀$^{\text{b}}$r-ú		*!		*
☞ c. ɓùr-ú			*	*

This account, then, is able to capture the intuitions that Laal rounding harmony, and many other multiply-triggered alternations, are abstract and phonological, but nonetheless driven by cumulative coarticulatory effects. They are abstract because [x F] values do not make reference to fine-grained phonetic detail, unlike the direct pho-

---

[14] Note that this is not violated if a [+/-round] segment is realized as [x round], where x is greater than 1 and less than 0.

netics models argued against in Chapter 2. Instead, they represent an abstraction over the physical phonetic structures involved in coarticulation (Lionnet, 2016:48-49). That being said, they also capture the intuition that these multiply-triggered processes are driven by cumulative coarticulatory pressures by encoding expected coarticulation in an [x F] feature value, and allowing those feature values to be targeted for phonological processes. These two characteristics—the separation between phonology and phonetics, and the interaction between the two—mean that this model bears the necessary characteristics argued for in Chapters 2 and 3. However, as I will show, a subfeatural account struggles to derive the interaction between laryngeal reduction and mora deletion.

The difficulty that mora deletion in SMPM presents for a subfeatural account comes from two inter-related sources: First, a subfeatural account relies on phonological features of the context surrounding the segment in question. Second, by virtue of being an abstraction over the physical phonetic structures involved in coarticulation, it does not take speech rate into account. As argued in Chapter 3, laryngeal reduction and mora deletion are processes that cannot be given a clear, phonologically-defined conditioning environment, meaning that there are not consistent phonological features in the environment that can be appealed to in triggering mora deletion. Additionally, by virtue of its strong tie to laryngeal reduction, one of the main driving factors behind mora deletion is speech rate, which cannot be codified in a subfeatural approach. Said another way, the account of doubly-triggered rounding harmony requires a [+labial] consonant and a [+round] vowel in the local context—one to derive the [x round] feature, and the other to target it for harmony. The difference in SMPM is that the pressures involved appear not to codifiable as two separate phonological features, but rather one feature [+constricted glottis] and a separate, phonetic pressure of increased speech rate. Said another way, laryngeal reduction and, by extension, mora deletion are not driven by *multiple* coarticulatory pressures, but rather by a single coarticulatory pressure that is exacerbated by speech rate. Because speech rate has no place in the subfeatural model, an unaugmented version of it cannot derive the rate-dependent application of mora deletion. To illustrate why this is, I will first outline a basic derivation of mora deletion in SMPM using a subfeatural account.

The first point, as with Laal, is to introduce constraints that enforce correlation between [+/-F] feature values and extremes on the subfeatural scale. In this case, I will proceed under the assumption that laryngealized vowels have the feature [+constricted glottis] ([+cg]), and modal vowels the feature [-constricted glottis] ([-cg]).

- *[[>0 cg]]/[-cg]: A vowel must not be higher than [[0 cg]] on the subfeatural scale if it is specified as [-cg].
- *[[<1 cg]]/[+cg]: A vowel must not be lower than [[1 cg]] on the subfeatural scale if it is specified as [+cg].

As in Laal, these constraints would conflict with another constraint driving the assignment of intermediate [x F] values:

- *¬Coarticulated$_{\text{ʔ}\rightarrow\text{V}}$: A [-cg] vowel must be assigned a value of [[x cg]] if it is adjacent to a laryngealized vowel.

As before, ranking the constraint driving [xF] feature values above the constraints penalizing them derives outputs with subfeatural values. Additionally, ranking the constraint penalizing subfeatures corresponding to [+ cg] over the constraint penalizing subfeatures corresponding to [- cg] blocks modification of the laryngealized vowel, as in Candidate D below.

(45)

/vˀ + v/	*¬$C_{ʔ\rightarrow v}$	*[[<1cg]]/[+cg]	*[[>0cg]]/[-cg]
☞ a. vˀ [[1 cg]] + v [[x cg]]			∗
b. vˀ [[1 cg]] + v [[0 cg]]	∗!		
c. vˀ [[1 cg]] + v [[1 cg]]	∗!		∗
d. vˀ [[y cg]] + v [[x cg]]		∗!	∗

The optimal candidate in (45), then, is one with a subfeatural [x cg] value, reflecting the fact that the modal vowel is coarticulated with the preceding laryngealized vowel. The next step is to derive mora deletion. Following the process for Laal rounding harmony, we might posit a constraint that prohibits sequences of [+cg] and [x cg] vowels. Ranking this constraint above MAX[V] would derive mora deletion:

- *[[[1 cg]] +cg] [[[≥ x cg]] -cg]:

  A [-cg] vowel whose subfeatural [[cg]] values equals or exceeds x may not directly follow a [+cg] vowel whose subfeatural value is [[1 cg]].

- Max[v]: Assign one violation for every vowel in input that does not have a correspondent in the output.

In this case, Candidate A is ruled out because it violates the newly-introduced markedness constraint. Candidates B and C violate the constraint driving [x cg] values, meaning that they, too are ruled out. Candidate D avoids violating either of these constraints by deleting the modal vowel, which violates low-ranked Max[v], and it is the optimal candidate.

(46)

/v^{ʔ} + v/	*¬Coarticulated$_{ʔ\to v}$	*[[1 cg]] [[$\geq$ x cg]]	Max[v]	*[[>0cg]]/[-cg]
a. v^{ʔ} + v [[x cg]]		*!		*
b. v^{ʔ} + v [[0 cg]]	*!			
c. v^{ʔ} + v [[1 cg]]	*!			*
☞ d. v^{ʔ}			*	

It is here that the analysis runs into trouble: The constraint definitions and ranking above will derive categorical mora deletion in all cases where there is a laryngealized vowel followed by a modal vowel—that is, it will derive mora deletion in all laryngealized roots, regardless of the phonological environment. The reason for this is that the segment triggering the [x cg] feature and the segment triggering mora deletion are one in the same—there is no way, with a consistent constraint ranking, to cause the derivation of just a [x cg] feature and not mora deletion in the context of a laryngealized vowel.[15] This is because coarticulation exacerbated by to speech rate—analyzed earlier as the driving factor behind mora deletion—is not a part of the model. Instead, what is needed for a subfeatural account to work here is for coarticulation with one phonological unit to cause the derivation of the subfeature [x cg], and then another, separate phonological unit to trigger the phonological alternation of mora deletion. This is visualized below, where $\underset{\sim}{\mathrm{v}}$ represents a modal vowel coarticulated with a laryngealized vowel.

Input	**Coarticulation**	**Mora deletion**
/v/ = [0 cg]	[0 cg] → [x cg]	[x cg] → ∅

$$/\mathrm{v}^{ʔ}\mathrm{v}/ \xrightarrow{\qquad\qquad} [\mathrm{v}^{ʔ}\underset{\sim}{\mathrm{v}}] \xrightarrow{\qquad\qquad} [\mathrm{v}^{ʔ}]$$

Figure 5.4: Visualization of hypothetical doubly-triggered mora deletion

The first phonological unit triggering the derivation of an [x cg] feature can somewhat easily be identified as the preceding [+ cg] vowel. However, the second trigger is not phonological, but rather the phonetic factor of speech rate. This raises significant difficulty for the subfeatural account because abstraction away from fine-grained phonetic detail is a central piece of the philosophy of the framework. Specifically, subfeatures are a abstract reification of speakers' and listeners' phonetic knowledge of coarticulation (Lionnet, 2017:543-544). This phonetic knowledge is abstracted away from contextual variation caused by external factors such as speech rate, and instead constitutes a relatively static, abstract generalization in the phonological grammar. This is empirically motivated by the fact that the doubly-triggered rounding harmony in Laal is

---

[15] One might appeal to partially-ordered constraints (Anttila, 1995) to drive variation here, but this would fail to capture the fact that speech rate is a principle driving factor in the process.

insensitive to speech rate and occurs even in slow speech (Lionnet, 2017:532). To incorporate the fine-grained phonetic factor of speech rate, which is necessarily concrete and contextually-varying, into the grammar would go against this philosophy of abstracted phonetic knowledge.[16] In addition to contradicting the philosophical approach of the subfeatural framework, there may also be independent reasons to exclude speech rate from the phonological grammar. As discussed earlier in the chapter, allowing speech rate to influence phonological constraint ranking, for example, predicts alternations at one rate of speech that are absent at another, and vice versa. The limited typological data discussed earlier do not support this conclusion.

Now, it is worth pointing something out here: Showing that the subfeatural account does not derive the rate-conditioned process of mora deletion might be (not wholly inaccurately) perceived as a bit unfair—after all, this account is designed for cases of cumulative coarticulatory effects that are categorical and unaffected by speech rate. This means that, by definition, it does not take speech rate into account and therefore cannot model phonological alternations that depend on it. However, I think the discussion here is apt because coarticulation, in this account and in many others, is considered a phonetic factor, and when two or more of these phonetic factors 'gang up,' they can trigger a phonological change. In laryngeal reduction and mora deletion in SMPM, it also seems like multiple phonetic factors, namely laryngeal coarticulation and

---

[16] It is possible that, in keeping with the same philosophical approach, the grammar could make reference to an abstraction of speech rate. However, it is unclear how such an abstraction of speech rate in the phonological grammar would work—it necessarily varies from situation to situation. At least some processes of coarticulation, on the other hand, have been shown to have some consistent characteristics regardless of contextual changes (Recasens, 2015).

time pressures, gang up to trigger a phonological change. From this perspective, alternations driven by cumulative coarticulatory effects like Laal's doubly-triggered rounding harmony are not so different from mora deletion in SMPM—both involve an apparent cumulative influence of purportedly phonetic factors in the application of a phonological alternation. The difference between them is that the doubly-triggered rounding in Laal appears to be driven by an *abstraction* of coarticulation that is removed from its physical realization in time and space, but mora deletion appears to be driven by a combination of coarticulation as well as variation along the very physical, non-abstracted factor of speech rate. In this light, it is important to consider the effectiveness of the subfeatural framework relative to the framework proposed above because their comparative strengths and weaknesses are able to clarify necessary features of a an eventual, satisfactory account of the phonology-phonetics interface. To this end, the next section details how the PICS model described earlier, which is able to derive the speech-rate-conditioned process of mora deletion in SMPM, is also able to derive doubly-triggered rounding harmony in Laal.

5.4.2 Applying PICS to multiply-triggered processes

In this section, I show that a framework of the type outlined in this chapter is able to derive cases in which cumulative coarticulatory effects trigger phonological neutralization. Now, it is crucial to note here that I am not arguing that all apparent cases of cumulative coarticulatory effects *should* be modeled in the way I show below, but simply that at least some of then *can* be modeled in this way. For example, the low tone spread process of Chapter 2 is likely not one that should be derived in this way, as I discuss later. However, in the case that some multiply-triggered processes should be modeled as the direct result of cumulative coarticulatory pressures, this section outlines how the PICS model would do so.

The basic process is very similar to the speech-rate-conditioned process of mora deletion, though different in several ways. The starting assumption is that coarticulation between neighboring segments is determined by the phonetic system, not the phonological grammar. As in the discussion of how the model accounts for mora deletion in SMPM, the phonological grammar outputs a ranked set of candidates, and the phonetic system evaluates each candidate. If the optimal phonological candidate is so ill-formed that it surpasses a threshold of ill-formedness, then the next candidate in the output set is evaluated. If that candidate is different in a way that ameliorates the issues presented by the optimal output, then it survives to the surface.

Let us take once again as an illustrative example the process of doubly-triggered rounding harmony in Laal. In cases like (47)-(48) below, where only one coarticulatory trigger is present, no change happens from input to phonological output.

(47) kə̀ə̀m-ə́
tree.species-PL
'Trees (of a particular species)' (Lionnet, 2017:527)

(48) sə̀g-ó
tree.species-PL
'Trees (of a particular species)' (Lionnet, 2017:527)

The lack of application of vowel harmony in (48), can be derived by giving a constraint driving rounding harmony a lower weight than the faithfulness constraints prohibiting changes to or deletion of a vowel:[17]

(49)

/sə̀g-ó/	MAX[V] wt=7	IDENT[V] WT=3	AGREE[RD] WT=2	**Harmony** **score**
☞ a. sə̀g-ó			-1	-2
b. sòg-ó		-1		-3
c. sə̀g	-1			-7

Candidate A, the fully faithful candidate, is the optimal output. Additionally, let us assume that IDENT[RD] is weighted low enough that violation of it is not severe enough to exclude a candidate from the output set. In this case, the set of potential outputs produced by the phonological grammar is the ranked set in (50), with the candidate to which rounding has applied occupying a lower slot in the set, but nonetheless being included.

---

[17] Note that a candidate with progressive harmony is analyzed in Lionnet (2017:541) as being ruled out by the positional faithfulness constraint IDENT[ROUND]$_{\sigma 2}$, which penalizes changing the [+/-round] feature of the vowel of the second syllable.

Note also that I use weighted constraints here once again. This is to make the transition between phonological and evaluation easier to follow. The previous section used strict constraint ranking to accord with the presentation in Lionnet (2016, 2017)

(50) Output Set of phonological evaluation in (50)

1. [sə̀g-ó]
2. [sòg-ó]

This set of candidates is then passed along to the phonetic system, which converts each into phonetic representations and evaluates it. As in the SMPM example, we will assume a set of mapping constraints. Here, I will set the targets for [-round] and [+round] vowels at the ends of a scale from 0 to 1 as a uni-dimensional abstraction of the articulatory, acoustic, and perceptual measures associated with rounding.[18]

- MAP[-RD]: A [-round] vowel corresponds to a target value of 0 (assign one violation for every 0.05 distance between target value and actual value).
- MAP[+RD]: A [+round] vowel corresponds to a target value of 1 (assign one violation for every 0.05 distance between target value and actual value).

These constraints require [-round] vowels to be realized with a value of 0 along the relevant dimension, and for [+round] vowels to be realized with a value of 1 along this same dimension. When [+/- round] vowels are coarticulated with consonants or vowels that have a distinct labial or round feature (for example, a [-round] vowels is followed by a [+round] vowel), then that coarticulation has an effect on the ability of the vowel to reach its target specification.[19] Let us suppose that a [-round] vowel followed by

---

[18] This scale is very similar to the scale used for subfeatures, which is an interesting similarity between the two accounts.

[19] I should note here that 'reaching a target' is a highly simplified stand-in for reaching a certain point

a [+round] vowel deviates by a value of 0.3 from its target, and as result is realized with a value of 0.3 instead of its target value of 0. Giving the MAP[+RD] constraint a higher weight than the MAP[-RD] constraint means that anticipatory coarticulation (V1 being affected by V2) in this case is preferred to perseverative coarticulation (V2 being affected by V1). As a result, Candidate A, in which the non-round vowel deviates from its target value but the round vowel does not, is the preferred realization of a sequence of a [-round] and [+round] vowel. Candidate B, which shares deviation from the target value across both vowels, and Candidate C, in which the round vowel deviates from its target but the non-round vowel does not, are less harmonious than Candidate A.

(51) Anticipatory rounding

/ɨ...u/	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony** **score**
☞ a. [ɨ]=0.3, [u]=1	0	-6	-6
b. [ɨ]=0.15, [u]=0.85	-3	-3	-9
c. [ɨ]=0, [u]=0.7	-6	0	-12

Now let us assume that labial consonants cause neighboring [-round] vowels to deviate from their target as well. Let us arbitrarily set the magnitude of deviation that

---

on a collapsed multi-dimensional space, defined either articulatorily or perceptually, and that simply reaching a target is too simplistic. Instead, what is evaluated here also likely takes into consideration the amount of time that target is held, whether the surrounding context conveys any other consequences of or cues to the sound in question, etc. However, I use the notion of reaching a target as a way to illustrate, at the simplest level, how this type of system could work.

they cause at 0.1.[20] This means that, when a [-round] vowel is preceded or followed by a labial consonant, its actual realization is 0.1, instead of the target realization of 0. Allowing the MAP[+RD] constraint to apply also to labial consonants, the result is that Candidate A, in which the labial consonant is faithfully realized but the vowel deviates from its target, is optimal. Candidate B, in which the two split deviation from their targets evenly, and Candidate C, in which the labial consonant deviates from its target but the vowel does not, are less harmonious.

(52) Consonant-vowel coarticulation

/Bɨ.../	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☞ a. [B]=1, [ɨ]=0.1	0	-2	-2
b. [B]= 0.95, [ɨ]=0.05	-1	-1	-3
c. [B]=0.9, [ɨ]=0,	-2	0	-4

Having established the coarticulatory effects of anticipatory vowel-vowel coarticulation and of consonant-vowel coarticulation, it is possible to model how the two pressures combined trigger vowel deletion. Here, it is necessary to set the threshold of ill-formedness past which a phonological output is rejected. Here, I will set it at the

---

[20] There is acoustic evidence in Laal that the coarticulatory changes caused by labial consonants are greater than that caused by round vowels (Lionnet, 2017:553), so these values do not accurately convey the magnitude of coarticulatory pressures. However, setting them at this value allows the model to capture the fact that [-round] vowels do not undergo rounding when flanked by two labial consonants but not followed by a [+round] vowel across all speech rates, as shown later.

arbitrary value of -7: Any phonological candidate that cannot be implemented without reaching that threshold of ill-formedness is rejected, and the next candidate from the output set is chosen. In both cases above, there was a possible candidate whose harmony score was -6 or higher, so no candidate reached the threshold of ill-formedness. However, as shown below, when both coarticulatory triggers are present, there is no candidate with a harmony score of -6 or higher. This is because the round vowel causes the non-round vowel to deviate from its target by 0.3, and the labial consonant causes the non-round vowel to deviate by 0.1 more. As a result, all possible candidates surpass the threshold of ill-formedness, and the optimal phonological output is thus rejected.

(53) Rejection of phonological output

/Bɨ...u/	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☹ a. [ɨ]=0.4, [u]=1	0	-8	-8
b. [ɨ]=0.2, [u]=0.8	-4	-4	-12
c. [ɨ]=0, [u]=0.6	-8	0	-16

In the case that the optimal phonological output is rejected, the second-ranked phonological output is put through the phonetic system. This is the candidate to which rounding harmony has applied. As shown below, this output does not reach the threshold of ill-formedness and therefore is implemented. The reason it does not reach the threshold of ill-formedness is because the coarticulatory rounding pressures of the following round vowel and flanking labial consonant do not cause it to deviate from its target value because it is already round.

(54) Implementation of second phonological output

/Bu...u/	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☞ a. [u]=1, [u]=1	0	0	0

In this way, the PICS model can derive multiply-triggered processes by allowing the cumulative effects of coarticulatory pressures to trigger the rejection of the optimal phonological output. Because neither coarticulatory pressure on its own causes the optimal phonological candidate to reach the threshold of ill-formedness, cases with only one trigger are properly implemented. However, when the combined triggers are present, their cumulative effect is enough to trigger rejection of the phonological candidate and the selection of the next-ranked candidate. However, there is an additional complication which this model must take into account, and that is the effect of speech rate.

5.4.3 Modeling rate-insensitivity in PICS

Speech rate has an effect on the extent of coarticulation, such that the extent of coarticulation between segments tends to be greater at fast rates of speech and lesser at slow rates of speech (Gay, 1981; Li and Kong, 2010; Recasens, 2015, 2018). This suggests that, as speech rate increases, the strength of coarticulation with neighboring segments should increase. We might model this by saying, for example, that a [-round] vowel will be realized further from its target value when coarticulated with a labial consonant in fast speech than in slow speech. Conversely, we might expect that, as speech rate decreases, the effect of physical coarticulatory pressures decreases. If this pressure decreases enough, we might expect that some alternations purportedly driven by cumulative coarticulatory pressures should not happen at slow rates of speech, since the pressures driving the alternation are alleviated. Such considerations have caused proponents of direct phonetics accounts to posit that coarticulatory pressures are abstracted across speech rates (e.g., Flemming, 2001:30-31). However, given that the PICS model necessarily takes speech rate into account, this same abstraction is not a possible way to capture the insensitivity of many multiply-triggered processes to speech rate. However, this is not necessarily a fatal flaw. The influence of speech rate on coarticulation is not always linear—it is not the case, for example, that all segments are more coarticulated with each other in fast speech, and less coarticulated with each other in slow speech. For example, Hertrich and Ackermann (1995) found that while the magnitude of perseverative vowel-vowel coarticulation (the effect of V1 on V2) was smaller in slow speech than fast speech in German, anticipatory vowel-vowel coarticulation (the effect of V2 on V1) was the identical between slow and normal speech. That is, slowing down speech did not lessen anticipatory vowel-vowel coarticulation. Additionally, DiCanio (2014) found that that Triqui speakers expanded their pitch range in fast speech relative to normal speech, producing non-neutralizing dissimilatory effects between higher and lower tones. That is, speeding up speech did not increase tonal coarticulation. So, while some coarticulatory effects are straightforwardly and somewhat linearly affected by speech rate, not all of them are. This fact is consistent with a view of at least some coarticulation as planned and controlled (Whalen, 1990), as opposed to its being an automatic, mechanical by-product of producing segments in a string.

All of these points are important because doubly-triggered rounding in Laal is rate-insensitive: It is not the case, for example, that it is attested only at normal and fast rates of speech but does not occur in slow speech. Instead, it occurs even across pauses (Lionnet, 2017:532). This is also a characteristics of low tone spread in SMPM, the vowel fronting process in Cantonese discussed in Chapter 2 (Flemming, 2001:30), and many other multiply-triggered phonological alternations. In order for the PICS model, which necessarily takes speech rate into account, to adequately model these processes, it needs to derive them regardless of speech rate. I will show here that this framework is still able to model rate insensitivity so long as at least one of the coarticulatory pressures driving the alternation is of the relatively rate-insensitive type described in Hertrich and Ackermann (1995), meaning that reducing speech rate does not reduce the coarticulatory pressures and cause the alternation to fail to apply.

To illustrate this, let us consider the two coarticulatory pressures involved in doubly-triggered rounding harmony in Laal. The first is the effect of a labial consonant on an adjacent (or near-adjacent) [-round] vowel, and the second is the effect of a [+round] vowel on a preceding [-round] vowel. What is needed is to posit that one of these coarticulatory patterns is maintained even in slow speech. Since it was anticipatory vowel-vowel coarticulation that did not decrease in German (Hertrich and Ackermann, 1995), we might propose that anticipatory vowel-vowel coarticulation in Laal is also maintained in slow speech, perhaps enforced in some way by the phonetic system. This means that the coarticulatory effect of a round vowel on a preceding non-round vowel is just as strong in slow speech as it is in fast speech. Porting over the proposed effect size from the preceding section, the effect can be set at 0.3, such that a [-round] vowel coarticulated with a following [+round] vowel is realized with a value of 0.3, instead of its target of 0. So, the tableau in (65), reproduced below, shows the output of a sequence of a [-round] and [+round] vowel in normal and slow speech.

(55) Anticipatory rounding in slow, normal, and fast speech.[21]

/ɨ...u/	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony** **score**
☞ a. [ɨ]=0.3, [u]=1	0	-6	-6
b. [ɨ]=0.15, [u]=0.85	-3	-3	-9
c. [ɨ]=0, [u]=0.7	-6	0	-12

Now, what is needed is for at least one of the coarticulatory pressures to not be lessened in slow speech. That is, a certain degree of physical coarticulation beyond that expected by a purely automatic view of coarticulation must be enforced, even in slow speech for at least one of the triggers involved. However, the other pressure(s) can be speech-rate-sensitive. For example, the effect of flanking labial consonants on

---

[21] It is also possible that a coarticulatory pattern will not *decrease* in slow speech but may *increase* in fast speech. If that were the case here, then this tableau would only hold of normal and slow speech.

the realization of a [-round] vowel can be sensitive to speech rate, such that there is more coarticulation at fast rates of speech, and less coarticulation at slower rates of speech. Once again porting over the proposed effect size from earlier, the effect of a labial consonant in normal speech on an adjacent [-round] vowel is to cause it to deviate by 0.1 from its target realization. So, the tableau in (52), repeated below, shows the effect of consonant-vowel coarticulation at normal speech rates.

(56) Consonant-vowel coarticulation in normal speech.

/Bɨ.../	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☞ a. [B]=1, [ɨ]=0.1	0	-2	-2
b. [B]=0.9, [ɨ]=0,	-2	0	-4
c. [B]= 0.95, [ɨ]=0.05	-1	-1	-3

Having established a baseline of normal speech, let us assume that the deviation that is triggered by a labial consonant increases to 0.15 in fast speech, and decreases to 0.05 in slow speech.[22] The following results would obtain, and in each case, Candidate A is optimal:

---

[22] Notably, it doesn't completely go away in slow speech—even in very slow speech, there is *some* coarticulation between adjacent segments.

(57) Consonant-vowel coarticulation in fast speech.

/Bɨ.../	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☞ a. [B]=1, [ɨ]=0.15	0	-3	-3
b. [B]=0.85, [ɨ]=0,	-3	0	-6
c. [B]= 0.95, [ɨ]=0.1	-1	-2	-4

(58) Consonant-vowel coarticulation in slow speech.

/Bɨ.../	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☞ a. [B]=1, [ɨ]=0.05	0	-1	-1
b. [B]=0.95, [ɨ]=0,	-1	0	-2

Finally, these values mean that even two labial consonants flanking a [-round] vowel in fast speech do not trigger rounding harmony, as shown below. This is consistent with the fact that multiple labial consonants do not trigger rounding in Laal, even though they interact additively in the coarticulation they trigger (Lionnet, 2017:554).

(59) Double consonant-vowel coarticulation in fast speech.

/BɨB.../	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☞ a. [B$_1$]=1, [ɨ]=0.3, [B$_2$]=1	0	-6	-6
b. [B$_1$]=0.85, [ɨ]=0.15, [B$_2$]=1	-3	-3	-9
c. [B$_1$]= 0.85, [ɨ]=0, [B$_2$]=0.85	-6	0	-12

Now, with the threshold of ill-formedness set to -7 as before, it is possible to show that the cases with both coarticulatory triggers present will always surpass the threshold, even in slow speech. First, as in (53), the optimal phonological output is rejected in normal speech because it cannot be implemented without surpassing the threshold.

(60) Rejection of phonological output at normal speech rate.

/Bɨ...u/	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☹ a. [ɨ]=0.4, [u]=1	0	-8	-8
b. [ɨ]=0.2, [u]=0.8	-4	-4	-12
c. [ɨ]=0, [u]=0.6	-8	0	-16

At fast and slow rates of speech, the result is the same. In fast speech, the labial consonant exerts an even larger coarticulatory influence. In slow speech, the labial consonant has less of an effect. However, because anticipatory coarticulation is the same across speech rates, then the lowered effect of the labial consonant is not enough to save the derivation.

(61) Rejection of phonological output at fast speech rate.

/Bɨ...u/	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☹ a. [ɨ]=0.45, [u]=1	0	-9	-9
b. [ɨ]=0.25, [u]=0.8	-4	-5	-13
c. [ɨ]=0, [u]=0.55	-9	0	-18

(62) Rejection of phonological output at slow speech rate.

/Bɨ...u/	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☹ a. [ɨ]=0.35, [u]=1	0	-7	-7
b. [ɨ]=0.2, [u]=0.85	-3	-4	-10
c. [ɨ]=0, [u]=0.65	-7	0	-14

Because the optimal phonological output may not be passed through phonetic evaluation and implementation without reaching the threshold of ill-formedness, it is rejected, and the next-ranked phonological output is put through the phonetic system. This is the candidate to which rounding harmony has applied. As shown below, at all rates of speech, this output does not reach the threshold of ill-formedness and therefore is implemented. As before, the reason it does not reach the threshold of ill-formedness is because the coarticulatory rounding pressures of the following round vowel and flanking labial consonant do not cause it to deviate from its target value.

(63) Implementation of second phonological output at all rates of speech.

/Bu...u/	MAP[+RD]=1 WT=2	MAP[-RD]=0 WT=1	**Harmony score**
☞ a. [u]=1, [u]=1	0	0	0

In this way, the PICS framework is able to model rate-insensitive, doubly-triggered processes like Laal rounding harmony. This is important because it is able to account for both the rate-sensitive process of mora deletion in SMPM as well as the rate-insensitive process of harmony in Laal. The subfeatural framework originally proposed to account for this harmony in Laal, though, is unable to account for the rate-sensitive process of mora deletion in SMPM. What we have found, then, is that a framework that does not take rate into account is able to account for some processes that are apparently driven by cumulative phonetic effects, but not all. What is more, in order to account for rate-driven processes, this type of approach has to be augmented in ways that are both against the philosophy behind it, as well as unmotivated by the limited typological patterns seen in this chapter. On the contrary, a model like PICS, which is built to account for phonological processes affected by speech rate, is also able to account for multiply-triggered phonological alternations. It does not need to be augmented with a separate mechanism to account for rate insensitivity; instead, it is able to derive these processes via its regular mechanisms, so long as some pre-requisite conditions hold of the coarticulatory pressures involved. Specifically, at least one of the pressures must be of the apparently rate-invariant, intentional type described in (Hertrich and Ackermann, 1995).

However, if this really is the right type of model for these processes, and if cumulative coarticulatory pressures work in something like the way described here, then several characteristics should hold of at least some multiply-triggered processes. First, given that rate-insensitivity is only possible in this model with coarticulatory effects that do not change in slow speech, then rate-insensitive, multiply-triggered phenomena should all involve at least one coarticulatory effect that is maintained in slow speech. Additionally, multiply-triggered processes whose individual coarticulatory effects all decrease with speech rate should, at least in some cases, not apply in slow speech. This is because, if the rejection of the optimal phonological candidate is directly due to the effects of physical coarticulation, then a decrease in that physical coarticulation might be expected to allow the optimal phonological candidate to survive phonetic evaluation and implementation without reaching the threshold of ill-formedness. For similar reasons, we might also expect to see some multiply-triggered phenomena that require, for example, two triggers in slow or normal speech, but only one trigger in fast speech.

5.4.4 Further considerations

As mentioned earlier, it is not clear that all multiply-triggered processes should be modeled in this way, though I have shown that they can. For example, the type of analysis presented here, which necessarily relies on the physical phonetic structures, would not work for the low tone spread process discussed in Chapter 3. The reason for this is the same reason that the direct phonetics account was inadequate: Because this type of analysis relies directly on the physical structures involved and has no recourse to abstractions like derivation, it is unable to tell the difference between two identical physical structures that have different underlying representations. That is, it cannot handle opacity. So, it is clear that at least some multiply-triggered phenomena should decidedly *not* be analyzed as directly triggered by the coarticulatory pressures underlying them, but rather by a much broader abstraction of the physical pressures in the terms of inductively-grounded phonological constraints along the lines of Hayes (1999).[23]

However, it may well be the case that there are some multiply-triggered processes that should be analyzed in a model like PICS. This is because phonological alternations that are sensitive to ganging-up effects are not a homogeneous class. For example, in addition to the cumulative markedness effects discussed at various points throughout this dissertation, there are also cumulative faithfulness effects, in which vi-

---

[23] In the same way, a subfeatural account would likely have trouble in modeling the low tone spread process. The reason is that subfeatures are calculated based on knowledge about expected coarticulation, and if there are no coarticulatory differences between derived and underlying low tones, then there should be no difference in subfeatures between them. As a result, we would expect low tone spread to apply to derived lows as well as underlying lows.

olation of one high-ranked faithfulness constraint is preferred to violation of more than one lower-ranked faithfulness constraint (see Farris-Trimble, 2008 for discussion of these cases). Additionally, alternations apparently driven by cumulative coarticulatory pressures also vary in whether they require the triggers to be phonologically distinct or not. For example, in Laal, two labial consonants are not enough to trigger harmony—there must be both a labial consonant *and* a round vowel. However, Cantonese vowel fronting is different in that it requires multiple triggers, but those triggers can have the same phonological features. That is, a back vowel adjacent to one coronal consonant is not fronted, but a back vowel flanked by coronal consonants does.

So, if there are multiply-triggered alternations that should be analyzed in a model like PICS, then we should also expect to see some multiply-triggered alternations that display some sensitivity to speech rate in the ways outlined earlier: Either they do not apply in very slow speech, or they require fewer triggers in fast speech. If there is a multiply-triggered process that is rate insensitive whose its component coarticulatory pressures *are* sensitive to speech rate, both in slow speech and in fast speech, this would suggest that this alternation should not be modeled in PICS. Serious difficulty for the PICS model would be presented by a process with these characteristics for which there is also some compelling reason to believe that the physical phonetic pressures involved are truly the driving pressures, and reanalysis via a model like inductive grounding is impossible. It is not immediately apparent to me what these characteristics might be, since sensitivity to speech rate seems to be one of the most compelling pieces of evidence for the relevance of physical phonetic structures influencing phonological computation, but that is not to say that there are none. For example, one might find a multiply-triggered alternation that has the characteristics of a phonological alternation (i.e., is neutralizing and rate-independent), but which appears to be the result of merely additive coarticulatory effects. In this case, the physical pressures could be argued to be directly relevant to the phonological alternation, and this would be difficult to model in a framework like PICS. However, as it stands, it appears that the PICS model has a decently wide range of empirical coverage, accounting not just for rate-conditioned processes but also for some coarticulation-driven processes.

5.4.5 The derivational timing of PICS

I have argued so far that a model like PICS is capable of deriving multiply-triggered alternations that are apparently driven by cumulative coarticulatory pressures. However, before moving on to comparisons with other frameworks, it is worth considering the derivational timing of the process of phonetic filtering, since it is relevant to the framework's ability to model some multiply-triggered alternations. Specifically, I have argued independently for two pieces of the PICS approach that, when put together, create a more complicated picture of the derivation of multiply-triggered processes. The first piece is that the phonetic component of the grammar acts on the output of the phonological component only after all phonological computation is completed. In other words, the process of phonetic filtering is an interface phenomenon—it applies at the interface between phonology and phonetics, but not during the phonological derivation itself. The second piece is the claim in Chapter 3 that a derivational model of phonology is necessary in SMPM to account for the opaque interaction between low tone spread and rise flattening. Under a derivational model of phonological computation, the output of one level, such as lexical phonology, is the input to another level, such as post-lexical phonology (Kiparsky, 1982). Putting together these two pieces—a purely post-phonological phonetic system and a derivational model of phonology—makes a prediction: The phonetic filtering process proposed above should not apply to the output of intermediate levels of phonological derivation, only to the output of the final level of derivation. The reason for this is that the output of early stages of phonology must go through additional levels of derivation before being fed into the phonetic system. Because derivationally-intermediate outputs are not subject to phonetic filtering, then phonetic factors like speech rate should not be able to affect which output candidate is fed through to later levels of phonological derivation. This can be seen in Figure 5.5 below, which shows that the phonetic system only operates on the output set of post-lexical phonology, not on the output of lexical phonology.[24]

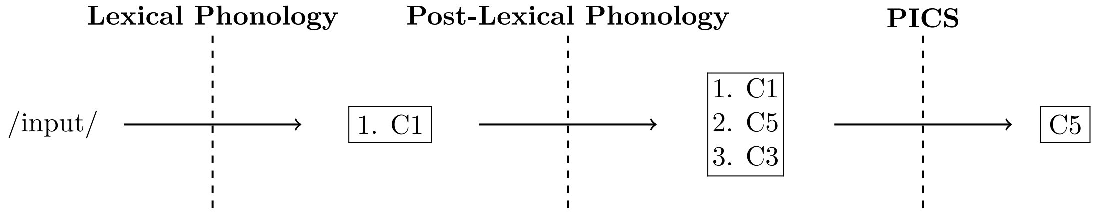


Figure 5.5: Purely post-phonological PICS

However, this is not the only logically-possible implementation of phonetic

---

[24] I have only listed one output of the lexical phonology for simplicity's sake, though it is in principle possible for a multiple-output framework like Coetzee's (2006) to be combined with a derivational phonological model to produce output sets at every level of representation.

filtering. For example, it could be the case that something like PICS applies to the output of every level of phonological derivation. Under this approach, the output of lexical phonology could be a set of ranked output candidates, which is then filtered by the phonetics, as schematized in Figure 5.6. What is more, this type of derivation would be needed if a process like PICS were truly responsible for the derivation of multiply-triggered rounding harmony in Laal in a way like that described above. The reason for this is that the harmony process in Laal is lexically-specific and word-bounded, meaning it almost certainly falls into the realm of lexical, or 'early,' phonology (e.g., Coetzee and Pater, 2011:402). So, in order for phonetic factors to influence the choice of phonological output in this case, they would have to apply to the output of lexical phonology. However, this type of iterative implementation of phonetic filtering also introduces several complications that are not faced by a purely post-phonological instantiation of PICS. These complications concern typological predictions and empirical motivation as well as phonological and phonetic units of representation.

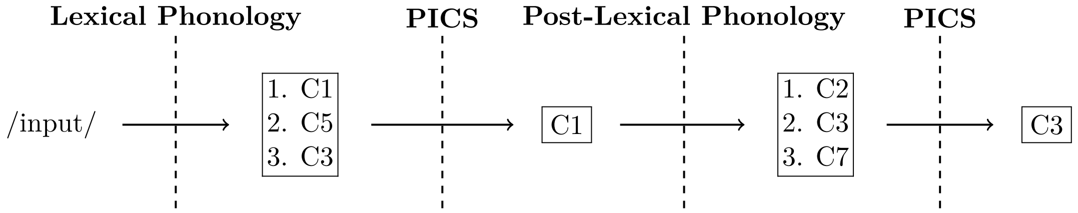


Figure 5.6: Iterative PICS

The first complication, albeit a small one, is that this implementation predicts that speech rate should be able to influence lexical phonology, but the limited typology of rate-conditioned alternations outlined above does not include a lexical phonological alternation. Instead, all of the alternations influenced by speech rate are post-lexical, applying between words. Of course, the limited nature of the typology does not preclude the existence of rate-conditioned lexical phonology, so this piece of evidence is merely suggestive. The second complication that an iterative PICS model would introduce is SMPM-internal: I have argued that a process like PICS should not be used to model low tone spread in SMPM because its reliance on physically-defined structures means that it is unable to distinguish between physically-identical but derivationally-distinct configurations, and as a result it cannot properly derive the opaque interaction between low tone spread and rise flattening. Under a completely post-phonological model like that in Figure 5.5, the inability of PICS to properly model low tone spread is straightforwardly understood, since it simply cannot influence early levels of phonological derivation. However, if PICS can apply to the output of early derivational levels of phonology, its inability to account for low tone spread is less straightforward.

The final complication introduced by an iterative implementation of PICS is more conceptual than the previous two, and it involves the difference between phonological and phonetic units of representation. I have argued in this dissertation that phonological and phonetic units of representation are distinct, with phonological representations being more coarse-grained and abstract than phonetic representations. This point is important because in an approach like PICS, the phonetic component is tasked not only with the filtering of phonological outputs, but also with the conversion of phonological units into phonetic units. As currently articulated, the input to the phonetic component is a set of candidates defined in terms of phonological units of representation, and the output of the phonetic component is a single output defined in terms of phonetic units of representation. This point raises an issue for an iterative PICS model: Let us suppose that the output of the lexical phonology is a ranked set of output candidates, which is then input to the phonetic component to undergo PICS. This process converts phonological units to phonetic units, and also evaluates the viability of the resulting phonetic configurations in the current speech conditions. At the end of this evaluation process, there is a single output defined in terms of phonetic units of representation. However, the issue is that this output, defined in phonetic units, must then be the input to the next level of phonological evaluation. If the distinction between phonological and phonetic units of representation holds at all levels of phonological derivation, as I believe that it should,[25] then the output of the phonetic component would necessarily be converted back into phonological units to undergo phonological evaluation at the post-lexical level. This process of multiple conversions can be seen in the following schematization of the process of phonological and phonetic evaluation of a simple high tone, where the step labeled 'Conversion' would be necessary to maintain the distinction between phonological and phonetic units of representation.

This conversion of phonological units to phonetic units, and then back again to phonological units introduces a level of complexity into the model that is not present in the purely post-phonological implementation of PICS. The introduction of additional

---

[25] Though see McCarthy (2011) for a discussion of differing amounts of non-contrastive information at different levels of derivation in Harmonic Serialism. Additionally, the fact that a commonly-cited difference between lexical and post-lexical phonology is whether or not it is structure-preserving (i.e., whether the outputs must consist of segments already in the phonemic inventory of the language in question) could be understood as resulting from a difference in units of representation across derivational levels.

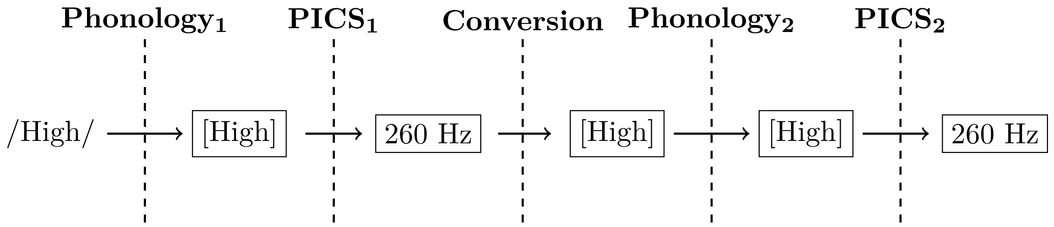


Figure 5.7: Conversion of units of representation in iterative PICS

complications is not fatal to a hypothesis, but it does require clear empirical support. Motivation for this type of iterative approach would come from a phonological alternation in a given language that must be analyzed as derivationally early but is nonetheless influenced by phonetic factors like speech rate in a way similar to the effect of speech rate on vowel degemination in Italian or mora deletion in SMPM. In the face of such evidence, the PICS model would likely need to be iterative in order to account for the data.

In the absence of such a phonological alternation, though, I adopt the purely post-phonological implementation of PICS in Figure 5.5. The reasons for this are that it involves fewer complications than the iterative model, it straightforwardly captures the facts that none of the rate-conditioned phonological alternations outlined earlier are clearly phonologically 'early,' and it straightforwardly predicts that the PICS model should not be used derive low tone spread in SMPM, a conclusion for which I argued earlier. The final reason to prefer the purely post-phonological implementation of PICS is that it makes easily-falsifiable predictions: It predicts that there should be no clearly phonetically-influenced phonological alternations that must necessarily precede other phonological alternations. This prediction can be falsified by the finding of such an alternation, while the prediction of the iterative model, namely that there *should* be such an alternation, could not be definitively falsified without a vast amount of typological research showing a lack of such alternations cross-linguistically. For these reasons, I opt for a purely post-phonological implementation of PICS.

Having adopted this stance, though, there is one unfortunate consequence, which is that doubly-triggered rounding harmony in Laal is not an entirely appropriate test case for the ability of PICS to model alternations driven by cumulative coarticulatory pressures. The reason for this is that, under the set of assumptions outlined above, a lexical phonological like harmony in Laal should not be derived in PICS. Importantly, though, this is a point against the adequacy of the specific analysis of harmony in Laal outlined earlier, but not against the ability of PICS to model multiply-triggered alternations in the first place. The larger point, which is that the empirical coverage of PICS extends beyond rate-dependent alternations to alternations ostensibly driven by coarticulatory pressures, remains.

5.4.6 Multi-level parallel constraint grammar

So far in this chapter, I have compared the PICS model with a subfeatural account (Lionnet, 2016, 2017), arguing that PICS has greater empirical coverage in that it can model both speech-rate-conditioned alternations and multiply-triggered alternations, while subfeatures can handle the latter but not the former. I have also considered multiple implementations of PICS, arguing for a one-time, post-phonological phonetic evaluation. In the following two sections, I briefly outline several other frameworks that have the characteristics I argued for in Chapters 3 and 4 and discuss their similarities and differences to PICS. The first relevant framework is Boersma and Van Leussen's (2017) multi-level parallel constraint grammar. This model has the characteristics argued to be necessary of any framework of the interface, in that it incorporates distinct levels of representation for phonology and phonetics, but also allows them to interact indirectly. The basic idea is that phonological candidates are not input-output pairs, as has been assumed throughout this dissertation, but rather that phonological candidates are 'paths' from an input at the most abstract level of representation to an output at the most concrete level of representation. The optimal candidate path is the one that incurs the least violations, with violations determined by constraints that evaluate correspondence between levels of representation as well as conditions imposed at a given level of representation. This can be visualized in Figure 5.8 below, which shows possible derivations for underlyingly laryngealized roots in SMPM.[26] As seen in the figure, phonological faithfulness and markedness constraints govern the mapping between phonological input and phonological output, and phonetic mapping and output constraints govern the mapping between phonological output and phonetic output.

The phonological input may be realized with or without mora deletion, and each of these phonological outputs may be realized in several distinct ways as a phonetic output. In this model, candidates are evaluated in parallel across all levels of representation, instead of only at one level of representation at a time. As a result, each 'path'

---

[26] The 'phonetic output' representations here are just illustrative placeholders for more phonetically-detailed forms. For example, 'CVV̰V' would represent a form with model-creaky-modal voicing, while 'CVʔV' would represent a form with modal voice, glottal closure, and then modal voice.

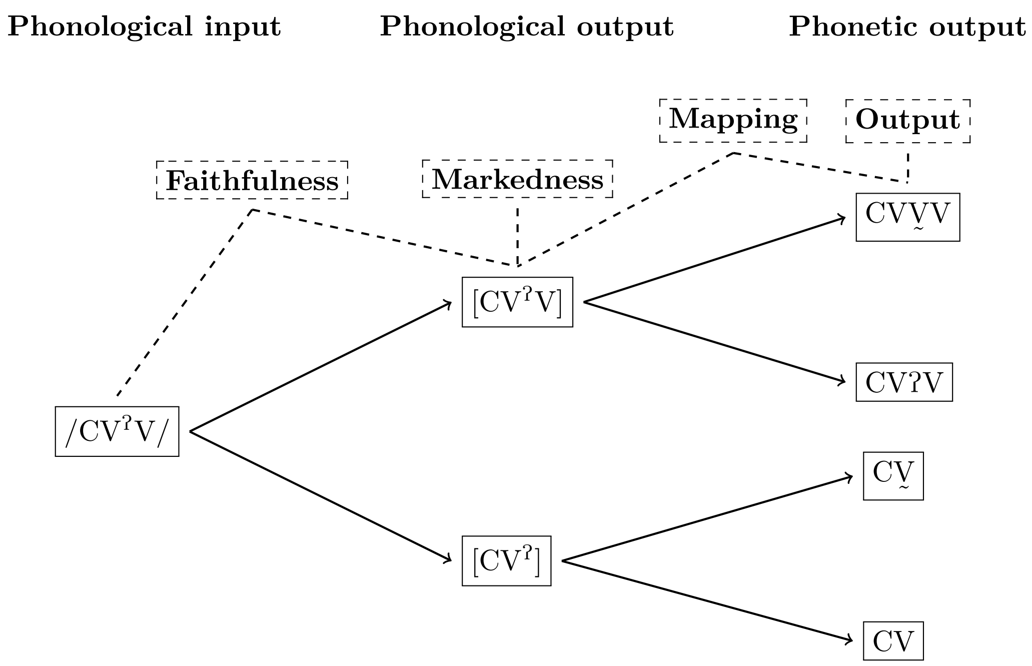


Figure 5.8: Candidate paths for laryngealized roots in SMPM.

from the phonological input to a single phonetic output in Figure 5.8 is a candidate. In this figure, there are four potential paths that start from the phonological input, so there are four potential candidates.

- Candidate A: /CVˀV/ → [CVˀV] → [CVV̰V]
- Candidate B: /CVˀV/ → [CVˀV] → [CVʔV]
- Candidate C: /CVˀV/ → [CVˀ] → [CV̰]
- Candidate D: /CVˀV/ → [CVˀ] → [CV]

Because candidates are paths and not forms at a given level of representation, it is possible for constraints at a later level of representation to influence the path taken at an earlier level of representation. For example, if phonetic mapping or output constraints heavily disfavor the phonetic outputs of Candidates A and B ([CVV̰V] and [CVʔV], respectively), then these phonetic constraint violations might override the preference for the phonological output [CVˀV]. In this case, then, the phonetic factors involved in rejecting certain phonetic outputs could lead to the favoring of paths involving a different phonological output. This is very similar to the PICS model, since the phonetic constraints are not a part of the phonological grammar, but they do have the ability to, in a sense, 'override' the preferences of the phonological grammar.

While both models are able to handle low tone spread and mora deletion in SMPM, they do make slightly different typological claims that are similar to the differences between predictions of the post-phonological and iterative implementations of PICS discussed above. A characteristic of Boersma and Van Leussen's (2017) model is that it involves parallel evaluation not just between phonological and phonetic levels of representation, but also mapping from meaning to morphemes, from morphemes to underlying phonological form, and then from phonological forms to phonetic forms. As a result, the model allows for any level of representation to affect any other level of representation.[27] However, as it is currently designed, PICS only explicitly allows phonetic filtering to apply to the last output of the phonological grammar, as discussed earlier. In other words, the phonetic system chooses between outputs of post-lexical phonology, but not between outputs of lexical phonology, or between two morphemes

---

[27] Note that this framework avoids the 'multiple conversions' issue of the iterative PICS model, since its parallel evaluation means that it does not need to convert phonological representations into phonetic representations and then back again. Instead, each level of the candidate path contributes to an overall ranking.

that share a lexical-semantic meaning. In this way, PICS is more constrained than a multi-level parallel constraint grammar, which does allow for these cross-level effects. As noted before, this implementation of PICS makes predictions, among which is the prediction that phonetic factors like speech rate should not influence a derivationally early phonological alternation, and also that similar phonetic factors should not influence the choice between two suppletive allomorphs. Once again, whether these predictions hold is ultimately an empirical question.

Another difference between the two models is that, by adopting Coetzee's (2006) rank-order model of EVAL, the PICS framework is constrained to choose from a limited set of phonological outputs. Specifically, only those candidates that violate low-ranked or low-weighted constraints are possible outputs, and it is only from these outputs that the phonetic system may draw. Phonetics is not able to obligate the phonological grammar to supply more outputs than are in the output set. The multi-level parallel constraint grammar does not have this characteristic: There is, in principle, no restriction on what phonological output may be chosen. Of course, this might turn out to be inconsequential, since phonological well-formedness is part of the overall computation of the optimal candidate path. However, it is in principle possible for constraints on phonetic form to have an outsize impact on the derivation that results in rejection of many phonological outputs. Once again, the PICS model is slightly more restrictive in this regard, having to choose from a constrained set of phonological outputs.

5.4.7 Constraint-formulation approaches

Another class of models that display the characteristics that I argue are necessary are those that impose a difference between phonological and phonetic levels of representation, but allow for language users' phonetic experience to inform the content and/or ranking of phonological constraints. One such proposal of inductive constraint grounding (Hayes, 1999) was outlined in Chapter 2 in order to illustrate the phonetic grounding of the phonological constraints involved in deriving low tone spread in SMPM. There are, in fact, a number of proposals that argue that phonological constraints can be more or less directly influenced by the substantive patterns behind them.

For example, in order to account for the fact that markedness constraints relativized to phonologically strong positions usually serve to augment the perceptual prominence of those positions, Smith (2004) argues that a filter on CON allows markedness constraints specific to phonologically strong positions if and only if they serve to enhance the perceptual prominence of that position. If a formally-possible constraint does not augment prominence in the position, then it is ruled out as a possible member of the universal constraint set. In this way, knowledge of perceptual prominence plays a direct role in defining the space of possible constraints. In another proposal, Flack (2007) argues that not all constraints in the universal constraint set of CON are innate, but that some constraints are consistently induced by all learners of all languages through their articulatory and perceptual experience. In this approach, too, phonetic knowledge has a hand in defining the possible set of phonological constraints.

Yet another approach is Steriade's (2001, 2008) P-Map, which proposes that faithfulness constraints in phonological grammars are ranked according to speakers' and listeners' knowledge of the relative similarity of phonological items in a given environment, such that some faithfulness constraints that prohibit perceptually large changes are universally ranked above faithfulness constraints that prohibit perceptually smaller changes in the same environment. Here, phonetic knowledge is able to play a role in constraint ranking, whereas the aforementioned proposals limited phonetic knowledge's role to constraint formulation or filtering.

These models vary in multiple ways, such as the in nature of the effect that phonetic knowledge has on the shape of phonological grammar (e.g., in constraint formulation or in constraint ranking) as well as whether they allow for language-specific phonological constraints (for Hayes (1999:26), some constraints may be language-specific, but for Flack (2007:8), all constraints are universal regardless of whether they are innate or induced). However, they all advance the insight that language learners' experience with speaking and listening can be used to directly influence phonological grammar, either by having an effect on the space of possible constraints, or on the space of possible constraint rankings. In that sense, these approaches embody the characteristics I have argued for throughout the dissertation: They separate phonetics and phonology, since that the constraints involved are formulated in terms of formal, phonological units, and they allow for some indirect interplay between phonetics and phonology, since they describe a way in which phonetic knowledge is abstractly encoded in phonological constraints and their ranking. However, since they allow for direct manipulation of either the contents of CON or the ranking of constraints, they allow for a more direct interaction between phonetics and phonology.

In the PICS model, phonetics and phonology constitute two separate systems, with phonetics acting on the outputs of phonological computation and not directly influencing how candidate are computed or evaluated. In the constraint-formulation/ranking approaches, phonetics plays a role in defining the constraints used in phonological evaluation, or even in establishing their ranking. Given that their effect is to produce static, phonological constraints, the way that they would account for a speech-rate-conditioned process like mora deletion in SMPM would be to establish phonological constraints relativized to fast speech. However, including rate-specific constraints in the phonological grammar predicts, via constraint permutation, languages in which a phonological process applies at one rate but never at another. Take, for example, the hypothetical constraints below, one relativized to fast speech and one not:

- $*\textsc{v}_\alpha\textsc{v}_\beta$[FAST]: Assign one violation for every pair of non-identical adjacent vowels in the output in fast speech.
- $*\textsc{v}_\alpha\textsc{v}_\beta$: Assign one violation for every pair of non-identical adjacent vowels in the output.

These constraints would conflict with a faithfulness constraints like IDENT[V], which penalizes changing features of vowels. The ranking $*\textsc{v}_\alpha\textsc{v}_\beta$[FAST] » IDENT[V] » $*\textsc{v}_\alpha\textsc{v}_\beta$ would derive phonological assimilation between adjacent, non-identical vowels in fast speech, but maintenance of the contrast in normal speech:

(64) Fast speech

/uo/	*$\text{V}_\alpha\text{V}_\beta$[FAST] WT=5	IDENT[V] WT=3	*$\text{V}_\alpha\text{V}_\beta$ WT=1	**Harmony score**
☞ a. [uu]	0	-1	0	-3
b. [uo]	-1	0	-1	-6

(65) Normal speech

/uo/	*$\text{V}_\alpha\text{V}_\beta$[FAST] WT=5	IDENT[V] WT=3	*$\text{V}_\alpha\text{V}_\beta$ WT=1	**Harmony score**
a. [uu]	0	-1	0	-3
☞ b. [uo]	0	0	-1	-1

In this hypothetical language, which is the implementation of a version of constraint induction that is sensitive to speech rate, there is a phonological alternation at one rate of speech that is not attested at another. This type of pattern runs contrary to the findings of the limited typological survey earlier in the chapter, precisely because it incorporates speech rate directly into the grammar, though in an abstracted form. As a result, using these proposals to model speech rate effects in phonology makes a different prediction than that made by the PICS model: They would necessarily predict that there are languages in which phonological changes may occur at one rate of speech but be categorically absent at another, whereas the PICS model does not. If the typological claim that speech rate only indirectly influences phonology through the filtering of outputs is correct, then using frameworks like those outlined in this section to model rate-sensitive phonology would be undesirable on typological grounds.

This point should be considered alongside the fact that these models are meant to be relativized to phonological objects only, and to abstract over factors like speech rate (Hayes, 1999:12)—rate-conditioned phonology is not what they were designed to account for. However, it is not the case that PICS can do all of the work that these models do, and I do not propose they be done away with. The reason for this is that PICS provides no mechanism by which phonetic grounding is incorporated into phonological constraints—it only aims to model how phonetic factors can influence the choice of output from an already-developed phonological grammar. In fact, Chapter 2 argued that direct reference to physical phonetics is not what drives low tone spread in SMPM, but that inductive grounding of phonological constraints is useful in analyzing the process. So, it is likely the case that the filtering of phonological outputs is just one of the ways that phonetic knowledge influences phonology, and processes like constraint induction constitute another. So, though approaches like those outlined here are not necessarily the best tool with which to analyze rate-driven phonology, this is not to their detriment—this class of proposals aims to account for another, likely separate way in which phonetic factors may influence phonology.

5.4.8 Review

In this section, I have outlined several frameworks of phonetic influence on phonology that have the characteristics I have argued are necessary, namely a separation between phonology and phonetics, and an interaction between them. I argued that, because a subfeatural account like that in Lionnet (2016, 2017) abstracts away from speech rate, it is unable to derive mora deletion in SMPM, and that augmentation to incorporate speech rate both goes against the philosophical approach embodied in the framework, and is also undesirable from a typological lens. An approach like the multi-level parallel constraint grammar of Boersma and Van Leussen (2017) is able to model the phonetic 'filtering' proposed to occur in the PICS model, though it differs in the proposed scope of phonetic influence on other components of the grammar. Finally, proposals that phonetic knowledge defines and/or ranks phonological constraints are able to roughly capture rate-dependent phonological alternations if they are allowed to reference speech rate, though this is, once again, against the philosophy of such accounts and also against the typological trend. However, given that models like inductive grounding are helpful in understanding alternations not fully accounted for in the PICS model, like low tone spread in SMPM, their empirical merits mean that they are independently desired in any model of the phonology-phonetics interface.

5.5 Conclusion

This chapter began by arguing that, though speech rate is able to influence phonology, it appears that the scope of its effect may be rather limited: In an investigation of rate-driven phonological alternations in the handful of languages for which these alternations could be shown to be truly phonological, it was shown that speech rate does not appear to introduce 'new' phonological alternations into a language's grammar. Instead, it seems to 'recycle' existing alternations, either across larger boundaries, or in different environments. The indirectness of this interaction is not predicted by approaches that model speech rate in the phonological grammar, since these necessarily predict that there should be rate-conditioned alternations that are present at one rate of speech but categorically absent from another.

On the basis of this observation, as well as the evidence from the previous two chapters that phonology and phonetics are distinct but nonetheless interact, a model of Phonetically-Informed Candidate Selection (PICS) was proposed, which broadly states that the phonetic system acts as a filter over a ranked set of phonological outputs. Specifically, if the optimal phonological output cannot be implemented without reaching a threshold of ill-formedness, then it is rejected, and the second-ranked output from the output set is passed through phonetic evaluation. If its phonological differences from the optimal output mean that it avoids the phonetic issues the first candidate faced, then it is implemented. In this way, the phonetic system does not merely translate phonological units into phonetic units, but it also has a selectional component. However, given that this selectional component may only act on a pre-defined set of potential phonological outputs, it is unable to fundamentally alter phonological computation.

A PICS-type model was compared to other frameworks of phonetic influence on phonology that broadly embody the characteristics I argued in chapters 3 and 4 to be necessary for any framework of phonetic influence on phonology, namely that the two systems be separate but able to interact. In each of these cases, certain advantages and disadvantages of each model were found. For example, PICS is able to model multiply-triggered alternations, but it must make certain assumptions about the interaction between speech rate and the coarticulatory pressures involved in order to do so. My goal throughout this theory-comparison section has not been to argue for the superiority of a PICS-type model, but rather to illustrate how comparison between various implementations of the same intuitions about the influence of phonetics on phonology can be used to generate testable hypotheses, the investigation of which are all but certain to shed more light on the complex and still-controversial nature of the relationship between phonetics and phonology. My hope is that this chapter has served to push this debate forward in a meaningful way.

Chapter 6

Conclusion

The driving question throughout the course of this dissertation has been the following: How is language users' knowledge of abstract sound categories related to their knowledge about the production and perception of those sounds? As is hopefully evident by this point, the answer to this question is anything but straightforward and has engendered significant debate, with a range of models that make varying and contradictory claims about the relationship between abstract sound categories (phonology) and their concrete, physical properties (phonetics). One contribution of this dissertation is that it has outlined two empirical phenomena in SMPM that are not amenable to analysis in terms of two extremes of the continuum of frameworks of the phonology-phonetics interface. Specifically, the process of low tone spread, described and analyzed in Chapter 3, requires phonology and phonetics to constitute two distinct levels of representation, ruling out models that do not allow a separation between the two. On the other hand, the interaction of laryngeal reduction and mora deletion outlined in

Chapter 4 requires phonology and phonetics to interact with each other, and specifically requires the phonetic component of a language's sound to be able to influence the phonological component in some way. This requirement rules out models that do not allow any interaction between phonology and phonetics. When considered together, these requirements necessitate a model of the phonology-phonetics interface that (1) differentiates between phonological and phonetic levels of representation, and (2) allows for some interaction between these two levels of representation. The resulting picture is represented in Figure 6.1, where only the frameworks between the two dotted lines may model both low tone spread and mora deletion in SMPM.

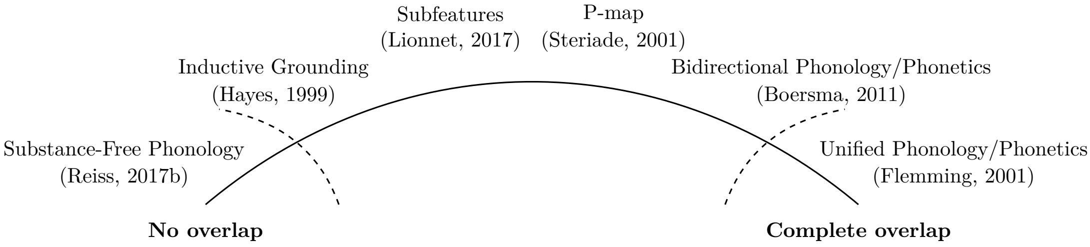


Figure 6.1: Continuum of frameworks of the phonology-phonetics interface

The remaining potential models frameworks represent a wide and disparate range of proposals, some of which make contradictory claims. For example, constraint induction models (Hayes, 1999; Smith, 2004; Flack, 2007) propose that phonetic knowledge is at least partially responsible for the formation of phonological constraints, but not their ranking, while Steriade (2001, 2008) proposes that phonetic knowledge is directly involved in the ranking of phonological constraints. Though they make distinct claims about the nature of the interaction between phonetics and phonology, they required characteristics argued for in Chapter 3-4 do not necessarily favor one of these approaches over the other—both differentiate phonological and phonetic levels of representation, and both allow for some interactional mechanism between the two. Because of the wide range of remaining proposals, Chapter 5 more deeply investigated the nature of the interaction between phonetics and phonology, arguing that the influence of phonetics on phonology is indirect. The reason for this is that the interaction of the phonetic factor of speech rate with phonological processes appears to be fairly constrained. Specifically, rate-conditioned phonological alternations appear to either (1) widen the domain of application of an extant phonological process, or (2) apply a family of processes (e.g., vowel deletion) in a new environment in fast speech. Crucially, there appear to be no cases of a phonological alternation that occurs in fast speech but has no phonological analogue in slow speech. An alternation of this type would, for example, involve phonological vowel harmony applying in fast speech but never in normal speech, or a complexifying sandhi process applying in normal speech but never in fast speech. However, alternations of this type are apparently unattested.[1]

Based on the indirect nature of the relationship between phonology and phonetics, a model of Phonetically-Informed Candidate Selection (PICS) was proposed, which argues that the phonetic system is not merely interpretational or implementational, but also involves a selectional component. Specifically, PICS adopts the proposal in Coetzee (2006) that the phonological grammar outputs not a single, optimal candi-

---

[1] Of course, the number of languages investigated is quite small because of the complexity involved in determining the phonological status of a fast speech alternation, so it is certainly possible that such a process exists. However, I have proceeded under the assumption that the sample investigated is largely representative of the types of fast speech alternations cross-linguistically.

date, but rather a set of potential output candidates that are relatively well-formed by virtue of violating only low-ranked/weighted constraints. This output set is evaluated by the phonetic system, which is able to reject the optimal phonological output when it is too phonetically marked given the speech conditions under which it is evaluated, which include factors like rate and coarticulation. The process, schematized in Figure 6.2, allows phonetics to influence phonology, but only indirectly through the filtering of potential output candidates.

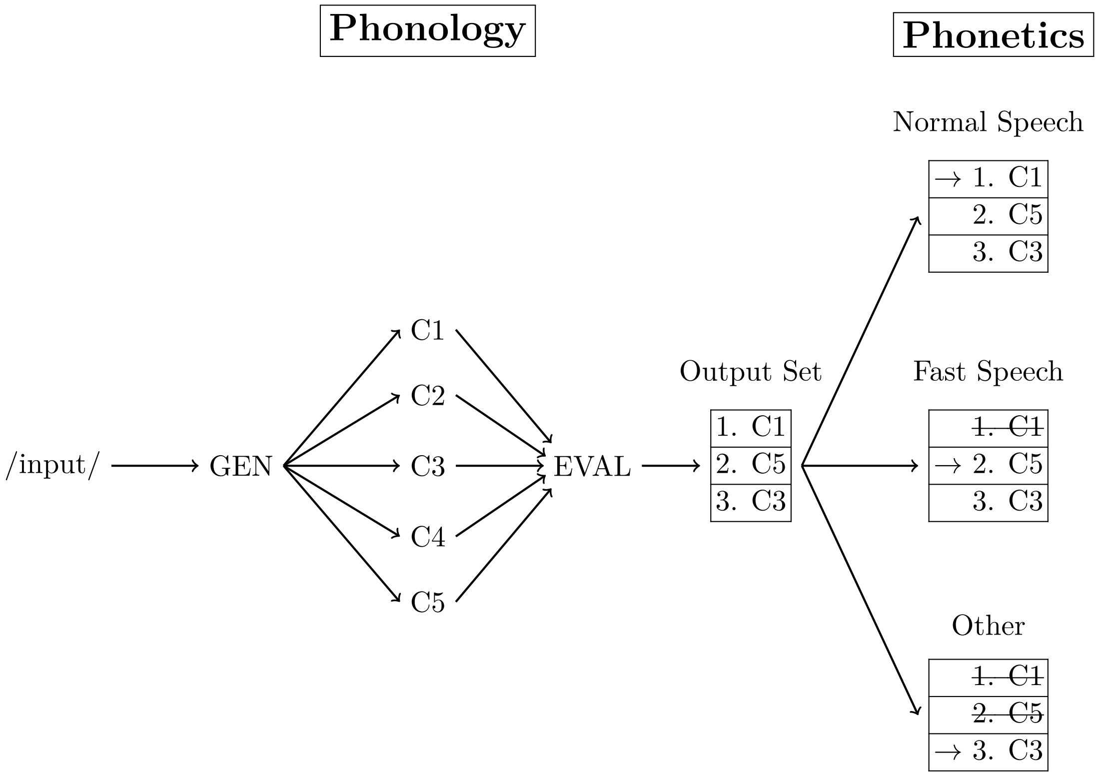


Figure 6.2: Illustration of Phonetically-Informed Candidate Selection (PICS)

The PICS model was compared with other models of the interface, showing that it has a relatively wide range of empirical coverage. For example, it is able to model at least some multiply-triggered alternations of the type described in Lionnet (2016, 2017), so long as certain assumptions about the interaction between speech rate and coarticulation hold. Ultimately, though, the PICS model is not put forward as a definitive model of the phonology-phonetics interface, but rather is meant to embody the intuitions developed throughout the dissertation, namely that phonology and phonetics are distinct levels of representation, but that phonetics does more than merely implement phonological outputs. The fact that PICS is incomplete is especially clear when one considers that the only mechanism of interaction between phonetics and phonology that it allows is output filtering. Because of this, other phenomena like the phonetic grounding of phonological constraints are left unexplained. It is likely, then, that the intuitions embodied in the PICS model are simply one piece of the range of phonology-phonetics interactions, and other pieces like inductive grounding (Hayes, 1999; Smith, 2004; Flack, 2007) are needed. However, I hope that the development and explicit articulation of these empirically-grounded intuitions has served to illuminate some directions of future research and framework comparison, with the ultimate goal of pushing forward the field's understanding of the relationship between abstract and concrete linguistic knowledge.

6.1 Future directions

There are many areas for future research left open in this dissertation. The most immediate involve the typological predictions made by the PICS model, namely that phonetically-conditioned phonological alternations should always have an analogue elsewhere in a language's phonology, and that phonetic factors should not influence the output of lexical phonology. While the discussion in Chapter 5 relies on a relatively informal definition of what it means to for an alternation to be a part of a larger family of alternations (that is, it involves violation of low-ranked faithfulness constraint), this intuition merits further development. Additionally, the question of what extra-phonological factors may influence phonology, and whether their interactions with phonology show the same indirect nature as that of speech rate, is worth further investigating. The strongest form of the PICS model predicts that all phonetic effects on phonological alternations should have this indirect nature, and that other apparently extra-phonological effects that do not have this indirect nature must arise by other means. For example, since the effect of speech style on phonological alternations can be categorical in that a phonological contrast in one register is absent in another (§5.3.1), this proposal makes the claim that speech style be considered a phonological factor. Finally, another prediction that PICS makes is that, if it is in fact an appropriate model for coarticulation-driven alternations, then there should be at least some multiply-triggered alternations that show sensitivity to speech rate. This sensitivity might mean that a multiply-triggered alternation does not apply in slow speech, or, conversely, requires fewer triggers in order to apply in fast speech.

Another area of future research concerns the number and nature of ways that the phonetic level of representation may influence phonological grammar. This dissertation has argued that at least two mechanisms are necessary, namely a sort of phonetic filtering like that in the PICS model, and phonological constraint induction (Hayes, 1999; Smith, 2004; Flack, 2007). However, there is no clearly-apparent upper limit on the ways that phonetics may influence phonology. While defining this limit is far beyond the scope of this dissertation, there exist several possibilities.

The first possibility is that phonetic influence on phonology is relatively limited, and that there are not many other mechanisms of interaction between phonetics and phonology than those outlined in this dissertation. It is important to note that, given the limited scope of PICS and constraint induction, there are almost certainly other necessary pieces that concern issues such as whether and how phonological features are related to language users' phonetic knowledge (e.g., Lin and Mielke, 2008), or how languages' phonological inventories are related to pressures on production and perception (Lindblom, 1990; Flemming, 1995). Nevertheless, it is possible that a phonetic system enriched with a fairly small number of interactional mechanisms is appropriate.

Alternatively, it might be the case that the phonetic component is much more articulated and complex than what is commonly assumed in many phonological theories, and that it is even more complex than what I have argued for. There are several ways that this complexity could take shape. The first is that the phonetic component could have a relatively large number of mechanisms that allow it to indirectly influence phonology, and as a result that there are a multitude of ways that phonetics can influence phonology. The second way is that the phonetic component could undertake a type of computation similar to phonology, in which it is able to directly make non-trivial and even categorical changes to its input. For example, it could be the case, as is proposed in some phonological frameworks, that rather categorical but non-contrastive alternations like allophony are computed in the phonetic component rather than the phonological component (e.g., Anderson, 1975). To accomplish this, the phonetic system would need direct access to units of abstraction like distinctive features, syllables, and prosodic words, since these are often relevant domains for allophony. Additionally, the phonetic system would need to be able to manipulate these representations in non-trivial ways. This idea is not without some precedent, since more ‘performance’-oriented processes like real-time sentence processing are sensitive to rather abstract, hierarchically-defined units like phrasal categories, which are also used in theoretical syntax (e.g., Pickering and Van Gompel, 2006). However, the inclusion of abstract structure in the phonetic component would need to be reconciled with the argument in Chapter 3 that phonology and phonetics operate over distinct units of representation. One way to do this would be to claim that phonology has access only to abstract units, and phonetics has access to both abstract units and fine-grained, physical units.

Ultimately, the number and nature of interactional mechanisms between phonology and phonetics is an empirical question, and one that can only be addressed by considering multiple sources of convergent evidence from distinct empirical sources. For example, considering both typological tendencies and behavioral data can help us to understand the ways that operations like language processing can influence the structure of languages’ grammars over time. This dissertation only explicitly argues for two ways that phonetics can influence phonology, and does not make any overt claims about other types of interactions or complexifications of the phonetic component. That being said, it does not preclude the possibility of other mechanisms of phonetic influence on phonology, nor of a more highly-articulated and internally-complex phonetic system.

As I hope is evident from the discussion in this section and throughout this dissertation, the question of how language users' abstract knowledge of the sound systems of their languages influences and is influenced by the physical processes of speech production and perception is anything but settled. However, as demonstrated in the previous chapter, the explicit formulation of a model and its comparison with competing models serves to create testable predictions, which in turn leads to advancement of the field's knowledge. Ultimately, the goal of this dissertation has been exactly that: To use empirical evidence to formulate and test hypotheses about the nature of the relationship between abstract and concrete linguistic knowledge, with the end result of advancing our understanding of the nature of language users' knowledge of the sound systems of their languages.

Bibliography

Anderson, S. R. (1975). On the interaction of phonological rules of various types. *Journal of Linguistics*, 11(1):39–62.

Anttila, A. (1995). Deriving variation from grammar: A study of Finnish genitives.

Ao, B. X. (1993). *Phonetics and phonology of Nantong Chinese*. PhD thesis, The Ohio State University.

Avelino, H. (2010). Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation. *Journal of Voice*, 24(3):270–280.

Bach, E. and Harms, R. T. (1972). How do languages get crazy rules? *Linguistic change and generative theory*, 1:21.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, 42(1):1–22.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1):1–48.

Becerra Roldán, B. (2015). Un estudio fonológico del mixteco de Santo Domingo Huendio, Oaxaca. Master's Thesis.

Becerra Roldán, B. (2017). Análisis sincrónico y consideraciones diacrónicas sobre la fonológia del mixteco de San Pedro Tulixtlahuaca. *MA thesis, Universidad Nacional Autónoma de México*.

Benua, L. (1995). Identity effects in morphological truncation. In *Papers in Optimality Theory*, chapter 18, pages 131–203. University of Massachusetts Occasional Papers in Linguistics.

Bermúdez-Otero, R. (1999). *Constraint interaction in language change: quantity in English and Germanic*. PhD thesis.

Bermúdez-Otero, R. (2011). Cyclicity. *The Blackwell companion to phonology*, pages 1–30.

Bermúdez-Otero, R. (2015). Amphichronic explanation and the life cycle of phonological processes. In *The Oxford handbook of historical phonology*, pages 374–399. Oxford University Press.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. *Bidirectional optimality theory*, 180:33–72.

Boersma, P. and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic inquiry*, 32(1):45–86.

Boersma, P. and Van Leussen, J.-W. (2017). Efficient evaluation and learning in multilevel parallel constraint grammars. *Linguistic Inquiry*, 48(3):349–388.

Boersma, P. and Weenink, D. (2020). Praat: Doing phonetics by computer (version 6.1.27).

Bolozky, S. (1977). Fast speech as a function of tempo in natural generative phonology. *Journal of Linguistics*, 13(2):217–238.

Bolozky, S. and Schwarzwald, O. (1990). On vowel assimilation and deletion in casual Modern Hebrew.

Borowsky, T. J. (1986). *Topics in the lexical phonology of English*. PhD thesis, University of Massachusetts Amherst.

Brodkin, D. (2022). Minimality, movement, and existential match. Unpublished manuscript.

Browman, C. P. and Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180.

Cambier-Langeveld, T. and Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics*, 27(3):255–280.

Chen, Y. (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics*, 34(2):176–201.

Cheng, L. L. S. (1991). Feature geometry of vowels. In *Proceedings of the 9th West Coast Conference on Formal Linguistics*, volume 9, page 107. Center for the Study of Language (CSLI).

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

Chomsky, N. and Halle, M. (1968). The sound pattern of English.

Coetzee, A. W. (2006). Variation as accessing ‘non-optimal’ candidates. *Phonology*, pages 337–385.

Coetzee, A. W. (2016). A comprehensive model of phonological variation: Grammatical and non-grammatical factors in variable nasal place assimilation. *Phonology*, 33(2):211–246.

Coetzee, A. W. and Pater, J. (2011). The place of variation in phonological theory. *The handbook of phonological theory*, page 401.

Cohn, A. C. (2007). Phonetics in phonology and phonology in phonetics. *Working Papers of the Cornell Phonetics Laboratory*, 16:1–31.

Davidson, L. (2006). Schwa elision in fast speech: Segmental deletion or gestural overlap? *Phonetica*, 63(2-3):79–112.

DiCanio, C. (2014). Triqui tonal coarticulation and contrast preservation in tonal phonology. In *Proceedings from Sound Systems of Mexico and Central America.*

DiCanio, C., Benn, J., and Castillo García, R. (2020). Disentangling the Effects of Position and Utterance-Level Declination on the Production of Complex Tones in Yoloxóchitl Mixtec. *Language and Speech*, pages 516–557.

DiCanio, C., Benn, J., and García, R. C. (2018). The phonetics of information structure in Yoloxóchitl Mixtec. *Journal of Phonetics*, 68:50–68.

Du, N. and Durvasula, K. (2020). Phonetically incomplete neutralization can be phonologically complete.

Duranti, A. (1981). Speechmaking and the Organisation of Discourse in a Samoan Fono. *The Journal of the Polynesian Society*, 90(3):357–400.

Eischens, B. (2020). Negative features in San Martín Peras Mixtec. *Proceedings of 50th Annual Meeting of the North East Linguistic Society.*

Eischens, B. (to appear). Polar question formation in San Martín Peras Mixtec. *Proceedings of the 25th Workshop on Structure and Constituency in the Languages of the Americas.*

Farris-Trimble, A. W. (2008). *Cumulative faithfulness effects in phonology.* PhD thesis.

Flack, K. G. (2007). *The sources of phonological markedness.* PhD thesis.

Flemming, E. (1995). *Auditory representations in phonology.* PhD thesis.

Flemming, E. (2001). Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology*, 18(1):7–44.

Fletcher, J. (2010). The prosody of speech: Timing and rhythm. *The handbook of phonetic sciences*, pages 521–602.

Garellek, M. (2019). The phonetics of voice. In *The Routledge handbook of phonetics*, pages 75–106. Routledge.

Garellek, M., Chai, Y., Huang, Y., and Van Doren, M. (2021). Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association.*

Garrapa, L., Cristiano, A., and Nagy, N. (2021). Vowel Elision in Italian as spoken in Florence and Lecco. *Languages.*

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, 38(1-3):148–158.

Gerfen, C. and Baker, K. (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics*, 33(3):311–334.

Gerfen, H. (2013). *Phonology and phonetics in Coatzospan Mixtec*, volume 48. Springer Science & Business Media.

Goldsmith, J. (1976). *Autosegmental phonology.* PhD thesis, MIT Press London.

Gordon, M. and Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of phonetics*, 29(4):383–406.

Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *The annals of Statistics*, pages 70–84.

Hasegawa, N. (1979). Casual speech vs. fast speech. In *Papers from the Fifteenth Regional Meeting of the Chicago Linguistic Society*, volume 15, pages 126–137.

Hauser, I., Hughto, C., and Somerday, M. (2016). Faith-UO: Counterfeeding in harmonic serialism. In *Proceedings of the Annual Meetings on Phonology*, volume 2.

Hayes, B. (1990). Precompiled phrasal phonology. *The phonology-syntax connection*, 85.

Hayes, B. and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic inquiry*, 39(3):379–440.

Hayes, B. P. (1999). Phonetically driven phonology. *Functionalism and formalism in linguistics*, 1:243–285.

Hedding, A. (2019a). New information and the grammar of Focus: Evidence from San Martín Peras Mixtec.

Hedding, A. (2019b). Two tone sandhi processes in San Martín Peras Mixtec.

Hertrich, I. and Ackermann, H. (1995). Coarticulation in slow speech: Durational and spectral analysis. *Language and Speech*, 38(2):159–187.

Holt, L. L. and Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, 72(5):1218–1227.

Instituto Nacional de Estadística y Geografía (2010). Censo de población y vivienda.

Iverson, G. K. and Salmons, J. C. (1996). Mixtec prenasalization as hypervoicing. *International Journal of American Linguistics*, 62(2):165–175.

Josserand, J. K. (1983). *Mixtec dialect history.* PhD thesis, UMI Ann Arbor.

Jun, S.-A. and Beckman, M. (1993). A gestural-overlap analysis of vowel devoicing in Japanese and Korean. In *67th annual meeting of the Linguistic Society of America, Los Angeles.*

Jun, S.-A. and Jiang, X. (2019). Differences in prosodic phrasing in marking syntax vs. focus: Data from Yanbian Korean. *The Linguistic Review*, 36(1):117–150.

Kaisse, E. M. (1985). *Connected speech: The interaction of syntax and phonology.* Academic Pr.

Keating, P. A. (1996). The phonology-phonetics interface. *UCLA Working Papers in Phonetics*, pages 45–60.

Keating, P. A., Garellek, M., and Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. In *ICPhS.*

Kilbourn-Ceron, O. (2017). *Speech production planning affects variation in external sandhi.* PhD thesis.

Kilbourn-Ceron, O., Wagner, M., and Clayards, M. (2016). The effect of production planning locality on external sandhi: A study in/t. In *The proceedings of the 52nd Meeting of the Chicago Linguistics Society.*

Kingston, J. (2007). The phonetics-phonology interface. *The Cambridge handbook of phonology*, pages 401–434.

Kingston, J. (2011). Tonogenesis. *The Blackwell companion to phonology*, pages 1–30.

Kiparsky, P. (1982). Lexical morphology and phonology. *Linguistics in the morning calm: Selected papers from SICOL-1981*, pages 3–91.

Kiparsky, P. (2000). Opacity and cyclicity. 17(2-4):351–366.

Kirchner, R. (2000). Geminate inalterability and lenition. *Language*, pages 509–545.

Kirchner, R. (2004). Consonant lenition. In Hayes, B., Kirchner, R., and Steriade, D., editors, *Phonetically based phonology*, chapter 10.

Kreiman, J. and Gerratt, B. R. (2010). Perceptual sensitivity to first harmonic amplitude in the voice source. *The Journal of the Acoustical Society of America*, 128(4):2085–2089.

Kreiman, J., Gerratt, B. R., et al. (2010). Effects of native language on perception of voice quality. *Journal of Phonetics*, 38(4):588–593.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13):1–26.

Leben, W. R. (1973). *Suprasegmental phonology.* PhD thesis, Massachusetts Institute of Technology.

León Vásquez, O. (2017). Sandhi tonal en el mixteco de Yucuquimi de Ocampo. *MA thesis, Centro de Investigaciones y Estudios en Antropología Social.*

Li, Y.-h. and Kong, J.-p. (2010). Effect of speech rate on inter-segmental coarticulation in Standard Chinese. In *2010 7th International Symposium on Chinese Spoken Language Processing*, pages 44–49. IEEE.

Lin, Y. and Mielke, J. (2008). Discovering place and manner features: What can be learned from acoustic and articulatory data. *University of Pennsylvania Working*

Papers in Linguistics, 14(1):19.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The journal of the Acoustical society of America*, 35(11):1773–1781.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling*, pages 403–439. Springer.

Lionnet, F. (2016). *Subphonemic teamwork: A typology and theory of cumulative coarticulatory effects in phonology*. PhD thesis.

Lionnet, F. (2017). A theory of subfeatural representations: The case of rounding harmony in laal. *Phonology*, 34(3):523–564.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and speech*, 29(1):3–11.

Lisker, L. and Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and speech*, 10(1):1–28.

Macaulay, M. and Salmons, J. C. (1995). The phonology of glottalization in Mixtec. *International Journal of American Linguistics*, 61(1):38–61.

Macaulay, M. A. (1996). *A grammar of Chalcatongo Mixtec*, volume 127. Univ of California Press.

Maechler, M. (2021). *Hartigan's Dip Test Statistic for Unimodality - Corrected*.

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, pages 207–263.

McCarthy, J. J. (2000). Harmonic serialism and parallelism. In *North East Linguistics Society*, volume 30, page 8.

McCarthy, J. J. (2005). Taking a free ride in morphophonemic learning. *Catalan Journal of Linguistics*, 4:19.

McCarthy, J. J. (2007). Derivations and levels of representation. In de Lacy, P., editor, *The Cambridge Handbook of Phonology*, pages 99–117. Cambridge University Press.

McCarthy, J. J. (2011). Perceptually grounded faithfulness in Harmonic Serialism. *Linguistic Inquiry*, 42(1):171–183.

McCarthy, J. J. and Pater, J. (2016). *Harmonic grammar and harmonic serialism*. Equinox Publishing Limited.

McCollum, A. (2019). Gradient morphophonology: Evidence from Uyghur vowel harmony. In *Proceedings of the Annual Meetings on Phonology*, volume 7.

McKendry, I. (2013). Tonal association, prominence and prosodic structure in south-eastern nochixtlán mixtec.

Mendoza, G. (2020). Syntactic sketch of San Martín Peras Tu'un Savi. BA Thesis.

Mendoza Ruiz, J. (2016). Fonología segmental y patrones tonales del Tu'un Savi de Alcozauca de Guerrero. Master's Thesis.

Mo, Y. (2007). Temporal, spectral evidence of devoiced vowels in Korean. In *Proceedings of ICPhS*, pages 445–448.

Moon, S.-J. and Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *The Journal of the Acoustical society of America*, 96(1):40–55.

Mosel, U. and Hovdhaugen, E. (1992). Samoan reference grammar.

Myers, S. (2000). Boundary disputes: The distinction between phonetic and phonological sound patterns. *Phonological knowledge: Conceptual and empirical issues*, pages 245–272.

Nespor, M. (1987). Vowel degemination and fast speech rules. *Phonology*, 4(1):61–85.

Nespor, M. (1990). Vowel deletion in italian: the organization of the phonological component. *The Linguistic Review*.

Nespor, M. and Vogel, I. (1986). Prosodic phonology.

Nicenboim, B., Roettger, T. B., and Vasishth, S. (2018). Using meta-analysis for evidence synthesis: The case of incomplete neutralization in german. *Journal of Phonetics*, 70:39–55.

Ostrove, J. (2018). *When phi-agreement targets topics: the view from San Martín Peras Mixtec*. PhD thesis.

Pankratz, L. and Pike, E. V. (1967). Phonology and morphotonemics of ayutla mixtec. *International Journal of American Linguistics*, 33(4):287–299.

Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive science*, 33(6):999–1035.

Penner, K. (2019). Prosodic structure in Ixtayutla Mixtec: Evidence for the foot.

Peters, S. and Mendoza, G. (2020). Morphophonological processes in Piedra Azul Tù'un Ndá'vi (Mixtec, San Martín Peras). *Winter Meeting of the Society for the Study of the Indigenous Languages of the Americas*.

Peters, S. L. (2018). The inventory and distribution of tone in Tù'un Ndá'vi, the Mixtec of Piedra Azul (San Martín Peras), Oaxaca. Master's Thesis.

Pickering, M. J. and Van Gompel, R. P. (2006). Syntactic parsing. In *Handbook of* *psycholinguistics*, pages 455–503. Elsevier.

Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, 2:33–52.

Pike, E. V. and Small, P. (1974). Downstepping terrace tone in Coatzospan Mixtec. *Advances in tagmemics*, pages 105–134.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rabadán, L. E. and Salgado, G. R. (2018). Festivals, Oaxacan immigrant communities and cultural spaces between Mexico and the United States: The Guelaguetzas in California. *Migraciones Internacionales*, 9(3):37–65.

Recasens, D. (2015). The effect of stress and speech rate on vowel coarticulation in Catalan vowel–consonant–vowel sequences. *Journal of Speech, Language, and Hearing Research*, 58(5):1407–1424.

Recasens, D. (2018). Coarticulation. In *Oxford Research Encyclopedia of Linguistics*.

Reiss, C. (2017a). Substance free phonology. In *The Routledge handbook of phonological theory*, pages 425–452. Routledge.

Reiss, C. (2017b). Substance free phonology. In *The Routledge handbook of phonological theory*, pages 425–452. Routledge.

Sande, H., Jenks, P., and Inkelas, S. (2020). Cophonologies by ph(r)ase. *Natural Language & Linguistic Theory*, pages 1–51.

Sapir, S. (1989). The intrinsic pitch of vowels: Theoretical, physiological, and clinical considerations. *Journal of Voice*, 3(1):44–51.

Selkirk, E. (1984). Phonology and syntax: The relationship between sound and structure. *Cambridge, Mass.*

Shue, Y.-L. (2010). *The voice source in speech production: Data, analysis and models.* University of California, Los Angeles.

Silverman, D. (1997). Laryngeal complexity in Otomanguean vowels. *Phonology*, 14(2):235–261.

Simpson, A. P. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics*, 40(3):477–490.

Smith, J. (2005). Phonological constraints are not directly phonetic. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 41, pages 457–471. Chicago Linguistic Society.

Smith, J. L. (2004). *Phonological augmentation in prominent positions.* Routledge.

Smolensky, P. and Legendre, G. (2006). *The harmonic mind: From neural computation to optimality-theoretic grammar.* MIT press.

Smolensky, P. and Prince, A. (1993). Optimality theory: Constraint interaction in generative grammar. *Optimality Theory in phonology*, page 3.

Stanley, R. (1967). Redundancy rules in phonology. *Language*, pages 393–436.

Steriade, D. (2000). Paradigm uniformity and the phonetics-phonology boundary. *Papers in laboratory phonology*, 5:313–334.

Steriade, D. (2001). The phonology of perceptibility effects: The p-map and its consequences for constraint organization. *Ms., UCLA.*

Steriade, D. (2008). ‘the phonology of perceptibility effects: The P-map and its consequences for constraint organization'. The Nature of the Word: Essays in Honor of Paul Kiparsky.

Stremel, S. (2022). /i/ deletion in San Martín Peras Mixtec. Unpublished manuscript.

Towne, D. (2011). *Gramática popular del tacuate (mixteco) de Santa María Zacatepec, Oaxaca*. Instituto Lingüístico de Verano, Mexico.

Tsuchida, A. (1997). *Phonetics and phonology of Japanese vowel devoicing*. PhD thesis.

Uchihara, H. and Mendoza Ruiz, J. (2021). Minimality, maximality and perfect prosodic word in Alcozauca Mixtec. *Natural Language & Linguistic Theory*, pages 1–51.

Van Oostendorp, M. (1997). Style levels in conflict resolution. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, pages 207–230.

Wagner, M. (2012). Locality in phonology and production planning. *McGill working papers in linguistics*, 22(1):1–18.

Warner, N. (2011). Reduction. *The Blackwell companion to phonology*, pages 1–26.

Warner, N. and Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, 130(3):1606–1617.

Whalen, D. H. (1990). Coarticulation is largely planned. *Journal of Phonetics*, 18(1):3–35.

Whalen, D. H. and Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of phonetics*, 23(3):349–366.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95(4):2240–2253.

Xu, Y. and Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, 111(3):1399–1413.

Yip, M. (2002). *Tone*. Cambridge University Press.

Zendejas, E. H. (2014). *Mapa fónico de las lenguas mexicanas: Formas sonoras 1 y 2*, volume 19. El Colegio de Mexico AC.

Zsiga, E. (1995). An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. *Haskins Laboratories Status Report on Speech Research No. SR-117/118*, pages 67–79.

Zsiga, E. C. (2000). Phonetic alignment constraints: consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, 28(1):69–102.

Zwicky, A. (1972). Note on a phonological hierarchy in English. *Linguistic change and generative theory*, 275:301.