UNIVERSITY OF CALIFORNIA
SANTA CRUZ

VOICE QUALITY AND LARYNGEAL COMPLEXITY IN SANTIAGO LAXOPA ZAPOTEC

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

DOCTOR OF PHILOSOPHY

in

LINGUISTICS

by

Mykel Loren Brinkerhoff

June 2025

The Dissertation of Mykel Loren Brinkerhoff is approved:

Professor Grant McGuire, Chair

Professor Jaye Padgett

Professor Ryan Bennett

Professor Marc Garellek

Peter Biehl
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

List of Tables

Abstract

Voice Quality and Laryngeal Complexity in

Santiago Laxopa Zapotec

by

Mykel Loren Brinkerhoff

This dissertation provides a detailed description and analysis of the SLZ voice quality system, a minority language spoken by about 1000 people in the municipality of Santiago Laxopa, and its interactions with the tonal system of the language. Standard assumptions about the interaction between tone and voice quality in Otomanguean languages proposed by Silverman (1997a,b), where nonmodal phonation is realized in only a portion of the vowel and tone is realized on a modal portion, do not fully hold in Santiago Laxopa Zapotec. Instead, speakers routinely produce nonmodal phonation throughout the entire vowel for breathy vowels. The only time phasing is observed is with the two types of creaky voice that occur in the language: rearticulated and checked. Rearticulated vowels have a period of creakiness in the middle of the vowel, whereas checked vowels have creakiness at the end. Although this creakiness is pronounced in distinct locations, non-modal phonation remains throughout the entire vowel. These results were confirmed through statistical modeling.

…

Dedicated to my family,

Betsy and Maelyn,

I wouldn't be here without you.

Acknowledgments

Acknowledgements in dissertations and theses are always so awkward and sometimes difficult to write. This is because there are no words that can adequately express the gratitude that you feel towards those who have helped you along the way. But I will try my best.

I would like to thank my advisor, Professor Grant McGuire, for his guidance and support throughout my time at the graduate program at the University of California, Santa Cruz. Thanks to him, I have learned and grown as a researcher and as a person. Thanks to his guidance, encouragement, and support, I have been able to complete this dissertation.

This dissertation would also not have been possible without the help of my committee members, Professors Jaye Padgett, Ryan Bennett, and Marc Garellek. Their feedback and advice have been invaluable in shaping my research and my writing. Thank you for your time and effort in reading my dissertation and for your thoughtful comments. I am grateful for your support and encouragement.

Besides my committee members, I would like to also thank the first linguistic Banana slugs I had the pleasure of knowing. Abby and Aaron Kaplan, both alumni from the linguistic program here at UC Santa Cruz, introduced me to the world of phonetics and phonology while I was an undergraduate student at the University of Utah and by their own love of the field, they inspired me to pursue a career in linguistics. I am

grateful for their support while an undergraduate and for helping me take the first steps in my career as a linguist.

Finally, I would like to thank my wife, Betsy, and my daughter, Maelyn, for their patience and understanding. I could not have done this without them. Without their love and support, I would not have been able to complete this dissertation.

Chapter 1

Introduction

1.1 What is Voice Quality

Voice quality refers to the long-term characteristics of an individual's voice (Abercrombie 1967, Laver 1980). However, the term voice quality is often used more narrowly to refer to how the larynx affects the phonetic characteristics of speech sounds. When the larynx is manipulated during speech production, it results in qualities of voice that we describe as modal, breathy, and creaky. Modal voice is the most common type of phonation and is characterized by a regular vibration of the vocal folds, resulting in a clear and full sound. On the other hand, breathy voice is characterized by a partial closure of the vocal folds, allowing more air to escape during voicing, resulting in a breathy or airy quality. Creaky voice is characterized by a tight closure of the vocal folds during voicing, resulting in irregular vibration.

Additionally, phonation is often used interchangeably with voice quality to refer to the same phenomenon (e.g., Keating et al. 2010). However, there are some contexts in which these terms are used with slightly different meanings. Phonation is often used more technically to refer specifically to sound production by the vocal folds. At the same time, voice quality can encompass a broader range of characteristics, including resonance and articulation (Esling et al. 2019). Another way these two terms are used is to make a phonetics (i.e., voice quality) versus phonology distinction (i.e., phonation); Barzilai & Riestenberg (2021) uses this contrast in their work on phonation on San Pablo Macuiltianguis Zapotec.

Languages also make use of voice quality differences to convey paralinguistic information by "indexing the biological, psychological, and social characteristics of the speaker" (Laver 1968, Podesva 2016) or use it for phonemic contrasts (Ladefoged & Maddieson 1996). An example of the former is that female speakers of American English are often described as having a breathier voice quality than males (e.g., Klatt & Klatt 1990). In the latter, most Oto-Manguean languages use voice quality to distinguish between phonemic contrasts (Lillehaugen 2019).

1.2 Measuring voice quality

Acoustic measurements are the most common way of measuring phonation and involve analyzing the sound waves produced during speech (see Garellek 2019 for a detailed discussion). The oldest and most common way this is accomplished is with spectral silt measures, which reflect the open quotient, which is the proportion of the glottal cycle during which the glottis is open (Holmberg et al. 1995). These measures originate in Fischer-Jørgensen's (1968) investigation into breathy vowels in Gujarati. She observed that the amplitude of the first harmonic (H1) is higher for vowels labeled as breathy/murmured than for vowels that were plain/modal. As a way to normalize across the signal and mitigate the effects of the high-pass filter, she subtracted the amplitude of the second harmonic (H2) from the amplitude of the first harmonic (H1) to create a measure that is now commonly referred to as H1−H2.

Subsequent research has shown that this measure is effective at distinguishing not only breathy and modal phonation but creaky phonation as well in a variety of languages (see Garellek 2022 for a history). This has led to the widespread adoption of this measure and for some researchers to claim that "H1−H2 may be a (near-)universal acoustic measure of phonation (Esposito & Khan 2020: p. 8)". However, research has also shown that this measure is not always effective at distinguishing between phonation types in all languages (e.g., Brinkerhoff & McGuire 2025, Chai & Garellek 2022, Esposito 2010b, Simpson 2012).

Based on a multiplicity of research into voice quality showing that voice quality is multidimensional, Kreiman et al. (2014, 2021) have proposed a unified model that incorporates several acoustic measures that capture the spectral slope, the inharmonic source excitation, the time-varying source characteristics, and the vocal tract transfer function. They have that voice quality can best be understood as a combination of these different acoustic measures, and that no single measure is sufficient to capture the complexity of voice quality.

1.3 Interactions between Voice Quality and Tone

Typologically speaking, voice quality can interact with tone in various ways. In Esposito & Khan's (2020) typology of phonation and tone, they describe four different types of interaction between tone and phonation based on whether or not tone and phonation are contrastive in the language. These interactions are summarized in Table 1.1, where each row indicates whether or not tone is contrastive in the language, and each column indicates whether phonation is contrastive. Beginning in the top left cell, the first type of language is one where neither tone nor phonation is contrastive (Type I). The second type of language is one where phonation is contrastive but tone is not (Type II). The third type of language is one where tone is contrastive but phonation is not (Type III). Finally, the fourth type of language is one where tone and phonation are contrastive (Type IV). For types II-IV, the cells are further subdivided based on whether or not tone and phonation are cues for each other. Each of these types of languages is discussed in more detail below.

Table 1.1: Esposito & Khan's (2020) language types based on the contrastiveness of tone and phonation.

	No phonation contrast		**Phonation contrast**	
No tone contrast	No contrasts (Type I)		*f0* not a cue (Type IIa)	*f0* is a cue (Type IIb)
Tone contrast	VQ not a cue (Type IIIa)	VQ is a cue (Type IIIb)	Orthogonal (Type IVa)	Fused (Type IVb)

Type I languages are those languages that lack both tonal contrasts and phonation contrasts. According to Esposito & Khan (2020), type I languages include most Australian Aboriginal languages, most Austronesian languages, most Indo-European languages, most Afro-Asiatic languages, Standard Khmer, and the Turkic languages. Instead, tone and phonation are used for paralinguistic purposes such as "indexing the biological, psychological, and social characteristics of the speaker (Laver 1968)" or racial characteristics (Podesva 2016). Additionally, these languages may use voice quality for pragmatic purposes such as signaling emphasis or emotion.

Type II languages are similar to Type I languages in that they lack tonal contrasts, but instead have phonation contrasts on the vowel; however, these phonation contrasts are further subdivided into two subtypes (IIa and IIb) based on whether or not changes in the fundamental frequency (*f0*) are a cue to the phonation. Type IIa languages are those where *f0* is not a cue. According to Esposito & Khan (2020), these languages are quite rare and only two languages are known to exhibit this behavior: Danish[1] (Grønnum, Vazquez-Larruscaín & Basbøll 2013) and Gujarati (Khan 2012). Type IIb languages are those where *f0* is a cue for the different phonation types. These languages are sometimes called “register languages” and are frequently found in Southeast Asian languages such as Chanthaburi Khmer, Chong, Javanese, Kedang, Mon, Suai, and Wa (e.g., Brunelle & Kirby 2016, DiCanio 2009, Samely & Samely 1991, Wayland & Jongman 2003). For type IIb languages, how *f0* is a cue is language specific. For example, breathy vowels in Wa and Kedang begin with a lower *f0* than their clear/modal counterparts (Samely & Samely 1991). In contrast, breathy vowels in Chanthaburi Khmer have a higher *f0* than their clear counterparts (Wayland & Jongman 2003).

Type III languages are the mirror image of Type II languages in that they have a contrast in tone but lack a phonation contrast. These languages are subdivided into two subtypes based on whether phonation is a cue for the tonal contrasts (Type IIIa vs. Type IIIb). Type IIIa languages are those where tone is contrastive and phonation is not a cue for those tonal contrasts. According to (Esposito & Khan 2020), this includes Japanese, Navajo, Punjabi, Manange, many West African languages, Swedish, and Central Thai. Type IIIb languages, on the other hand, have tone and use phona-

[1] Frazier (2013) and Peña (2022, 2024) contest this claim and show rather convincingly that stød has $f0$ cues.

tion as a cue for one or more tones. For example, Mandarin Chinese's falling-rising tone is frequently accompanied by creak (e.g., Kuang 2017). According to (Esposito & Khan 2020), other examples of this include Cantonese Yue, Khmu' Rawk, Mandarin, Pakphanang Thai, Phnom Penh Khmer, and Yueyang Xiang.

The last type of combination of tone and phonation is Type IV languages. These languages have contrastive tone and phonation and are subdivided into two subtypes (IVa and IVb). Type IVa languages are those where tone and phonation are allowed to combine freely with no restrictions; Esposito & Khan (2020) call these languages *orthogonal* and are exemplified by languages such as Dinka, Mazatec, Mpi, Yalálag Zapotec, and Yi languages (see Esposito & Khan 2020 for references).

In comparison, Type IVb languages also use contrastive tone and phonation, but they are fused in such a way that it is impossible to state if they have tonal or phonation contrasts. Esposito & Khan (2020) call these languages *fused* but are also commonly called "register languages" and are primarily found in Southeast Asia (Brunelle & Kirby 2016, Enfield 2005, Masica 1976). Furthermore, Esposito & Khan (2020) also claim that most Zapotec languages also fall into this type of language because there are specific tones that only arise with specific phonation types. For example, Santa Ana del Valle Zapotec's rising tone can only occur with modal phonation and its falling tone can only happen with breathy phonation (Esposito 2010b); additionally, Isthmus Zapotec is notable for its high tone and falling tone being unable to appear with any phonation in monosyllables (Pickett, Villalobos & Marlett 2010).

Type IV languages are also collectively referred to as *laryngeally complex* languages by Silverman (1997a,b) because tone and phonation are both produced in the larynx and the interactions between the two are complex. The complexity in these languages arises from the variety of ways that tone and phonation are produced. For Silverman, laryngeal complexity most frequently manifests as a form of phasing between tone and phonation. Phasing is the idea that the most optimal way for tone and phonation to be produced sequentially in a vowel. This often manifests as a portion of the vowel being produced with nonmodal phonation and another portion produced with modal phonation. The rationale is that tone is easily perceived with modal phonation, while nonmodal phonation makes it difficult to perceive tone. However, Silverman also states that in laryngeally complex languages, tone and phonation can also be produced simultaneously if the nonmodal phonation is produced more weakly; this is argued to be the distinction between the laryngeally complex languages of Jalapa Mazatec and Mpi (Ladefoged & Maddieson 1996, Silverman 1997a).

1.4 Research question

This dissertation investigates the acoustics of voice quality in Santiago Laxopa Zapotec and its interactions with tone. Specifically, it aims to answer the following questions: (i) does the recently proposed residual H1* acoustic measure more effective captures the phonation contrasts between Santiago Laxopa Zapotec's four phonations than the traditional H1*−H2* acoustic measure; (ii) how is the acoustic landscape of voice quality in Santiago Laxopa Zapotec structured; (iii) which acoustic measures most effectively capture and classify the voice quality contrasts; (iv) and how do those acoustic measures help explain Santiago Laxopa Zapotec's laryngeal complexity?

These four questions form a cohesive program of study and are essential for several reasons. First, if new acoustic measures are proposed, we need to determine to what extent they are effective at capturing the phonation contrasts in a language compared to more established measures. If the new measure is shown to be more effective than a more established one, then it should be considered when performing our acoustic investigations. Second, understanding the acoustic landscape of voice quality in a language can provide insights into how voice quality is structured. Third, identifying the most effective acoustic measures for classifying voice quality contrasts can help researchers develop more accurate and reliable methods for analyzing voice quality. Fourth, investigating the interactions between tone and voice quality can shed light on the complexities of these interactions and their implications for our understanding of phonetics and phonology. Finally, answering these questions concerning Santiago Laxopa Zapotec, a language that has not been extensively studied, can contribute to our theoretical understanding by allowing us to test claims and verify the robustness of the proposed measures and theories across languages.

1.5 Outline of the dissertation

The rest of this dissertation is organized as follows. Chapter 2 provides a detailed description of the vowel, voice quality, and tone system of Santiago Laxopa Zapotec, an Oto-Manguean language spoken in Oaxaca, Mexico. This chapter describes the phonetic and phonological properties of the vowels, the phonation types, and the tonal contrasts in the language. Chapter 3 presents the results of an acoustic analysis of the voice quality contrasts in SLZ, focusing on the recently proposed residual H1* acoustic measure from Chai & Garellek (2022) and its effectiveness its effectiveness in distinguishing between the phonation types while comparing its results to the more traditional H1*−H2* acoustic measure. The results show that this measure is more effective than the conventional H1*−H2* measure in distinguishing the phonation types in Santiago Laxopa Zapotec, adding credence to this acoustic measure and its adoption by researchers.

Chapter 4 presents the results of an acoustic analysis investigating the acoustic landscape that Santiago Laxopa's voice quality occupies using multidimensional scaling. This chapter demonstrates that voice quality in Santiago Laxopa Zapotec occupies a 3-dimensional space where the first and third dimensions correlate with spectral slope. In contrast, the second dimension correlates with the periodicity and noise of the signal. This chapter also shows that Santaigo Laxopa Zapotec's 3-dimensional space is consistent with Keating et al.'s (2023) findings that cross-linguistically voice quality occupies a primarily two-dimensional space that is also defined by spectral slope and periodicity/noise. These findings suggest that the acoustic landscape of voice quality is consistent across languages and that similar acoustic properties define the dimensions of this space. Furthermore, the findings also show that residual H1* is highly correlated with the spectral slope dimensions of the acoustic landscape. This further suggests that this measure effectively captures the spectral properties of voice quality and is important for defining the acoustic space.

Chapter 5 presents a random forest analysis about which acoustic measures are most effective at distinguishing the phonation types in Santiago Laxopa Zapotec. This chapter shows that the most effective acoustic measures for characterizing the voice quality in Santiago Laxopa Zapotec are very similar to those correlated with the dimensions of the acoustic landscape in Chapter 4, with the addition of duration.

Based on the findings in Chapters 4 and 5, Chapter 6 discusses the implications of these findings for our understanding of voice quality and its interactions with tone. This chapter investigates explicitly Silverman's (1997a) claims about how tone and voice quality must be phased. Three generalized additive mixed models were assessed on the acoustic measures of $f0$, Strength of Excitation, and HNR $<$ 1500 Hz to examine these claims about phasing. The findings show that despite Herrera Zendejas's (2000) claims that Zapotec languages do not show phasing between tone and voice quality there is evidence that suggests that there is a phasing between tone and voice quality in Santiago Laxopa Zapotec. Specifically that the breathy and checked vowels are associated with a modal portion followed by a nonmodal portion of the vowel, being consistent with Silverman's (1997a) postvocalic phasing pattern. Rearticulated vowels, on the other hand, are associated with a nonmodal portion followed by a modal portion of the vowel, being consistent with Silverman's (1997a) prevocalic phasing pattern. These findings suggest that the interactions between tone and voice quality in Santiago Laxopa Zapotec are more complex than previously thought and that phasing plays a significant role. Furthermore, the findings suggest that the implicational hierarchy for the interactions between the phasing relationships between tone and voice quality proposed by Silverman (1997a) does hold for Santiago Laxopa Zapotec.

Finally, Chapter 7 concludes the dissertation by summarizing this research's main findings and contributions to our understanding of voice quality. It also discusses potential avenues for future research on voice quality and its interactions with tone. It highlights the importance of continued research to further our understanding of the complexities of voice quality and its role in language.

Chapter 2

Vowels and suprasegmentals in Santiago Laxopa Zapotec

2.1 Introduction

Santiago Laxopa Zapotec (SLZ; *Dille'xhunh Laxup* [di˩ʒeˀ˩ʐun˩ lːa˩ʂupʰ˥]) is a Northern Zapotec language spoken by approximately 1000 people in the municipality of Santiago Laxopa, Ixtlán,Oaxaca, Mexico and in diaspora communities throughout Mexico and the United States (Adler & Morimoto 2016, Adler et al. 2018, Foley, Kalivoda & Toosarvandani 2018, Foley & Toosarvandani 2022). According to Smith-Stark (2007), SLZ is part of the macro variety of Cajonos Zapotec, which also includes Zoogocho Zapotec, Yatzachi Zapotec, Yalálag Zapotec, Tabaá Zapotec, Lachirioag Zapotec, and several other varieties spoken in the Sierra Norte of Oaxaca, Mexico.

Figure 2.1: Photo of Santiago Laxopa taken by Beto Diaz, a resident.

2.2 Vowels in Santiago Laxopa Zapotec

SLZ exhibits a four-vowel inventory; see Table 2.1. This type of vowel inventory is very common among Sierra Norte Zapotecs. Most varieties have the vowels /i/, /e/, /a/, and /o/ (Nellis & Hollenbach 1980, Jaeger & Van Valin 1982, Butler H. 1997, Avelino 2004, Long & Cruz 2005, Sonnenschein 2005).

The vowel /o/ is marginal in SLZ's lexicon, only appearing in a few lexical items

Table 2.1: Vowel qualities in Santiago Laxopa Zapotec.

	front	central	back
high	i		u∼o
mid	e		
low		a	

such as the diminutive classifier *do'*. Instead, this vowel is replaced by /u/ in most cases. However, this difference is not universal among all speakers in the community. For the most part older speakers exhibit the vowel /o/ in their speech, while younger speakers tend to replace it with /u/. Most speakers, when asked, classify the two back rounded vowels as the same phoneme and view them as a dialectal feature between the different pueblos. For example, in neighboring San Bartolomé Zoogocho the /u/ vowel is very marginal and has led Sonnenschein (2005) to describe the language as having only four vowels.It is interesting to note that everywhere that SLZ has the vowels /u/ or /o/, Zoogocho only has /o/. Further evidence for this comes from plotting the vowels along the first two formants. As shown in Figure 2.2, the vowels /o/ and /u/ occupy nearly identical vowel spaces.

Additional evidence for the overlap of /o/ and /u/ can be measured with a combination of Pillai scores (Pillai 1955, Hay, Warren & Drager 2006, Nycz & Hall-Lew 2014) and Bhattarcharyya's Affinity (Bhattacharyya 1943, Johnson 2015, Warren 2018, Strelluf 2018). Both of these measures show what degree of overlap exists between

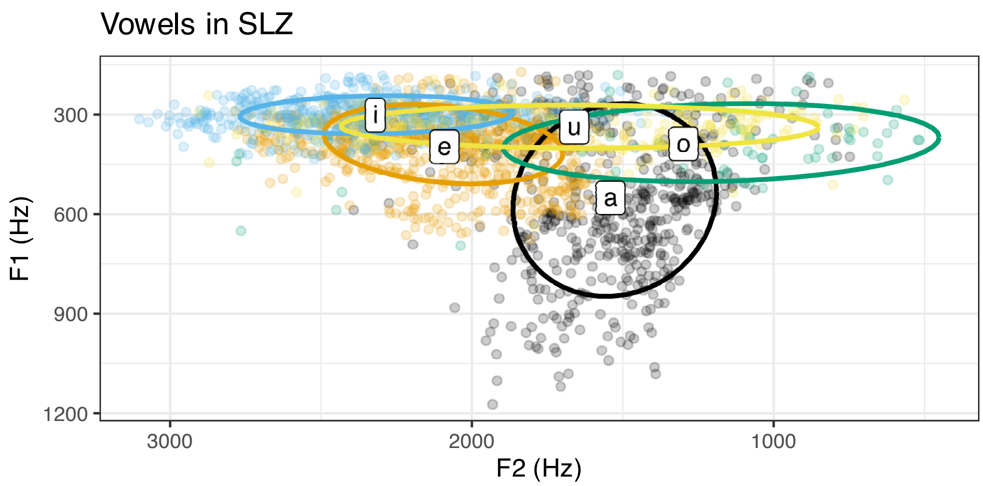
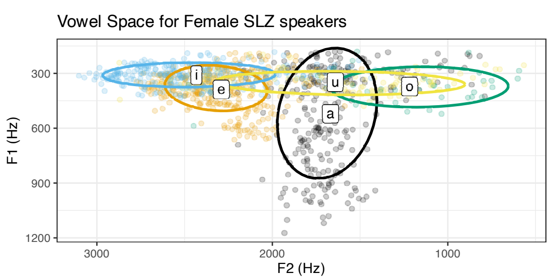
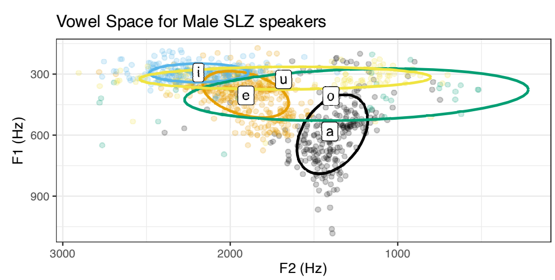


Figure 2.2: Vowel space of Santiago Laxopa Zapotec. The ellipses around each vowel mean represents 1 standard deviation. The scale of the axes are in barks with their corresponding Hz values.



(a) Female vowel space



(b) Male vowel space

Figure 2.3: Vowel space for SLZ speakers by gender. The ellipses around each vowel mean represents 1 standard deviation. The scale of the axes are in barks with their corresponding Hz values.

two different items in some space. Their use in linguistics has been used mainly to show the process of complete and partial mergers between vowels, such as the NEAR-SQUARE vowel merger in New Zealand English (Hay, Warren & Drager 2006). The Pillai scores and Bhattacharyya's Affinity show that the vowels /o/ and /u/ are nearly identical in their vowel space; see Table 2.2.

Table 2.2: Pillai scores and Bhattacharyya's Affinity for /o/ and /u/ in SLZ.

	Pillai score	Bhattacharyya's Affinity
All speakers	0.157	0.892
Females	0.138	0.890
Males	0.224	0.858

In interpreting these results, Pillai scores range from 0 to 1, with 0 indicating overlap and 1 indicating complete no overlap. Bhattacharyya's Affinity ranges from 0 to 1, with 0 indicating no overlap and 1 indicating complete overlap. The results show that the overlap between /o/ and /u/ is not complete, but it is also not completely separating. This is consistent with the observations made by myself and other researchers for this variety (Toosarvandani, p.c.).

In summary, we can conclude that SLZ is similar to other Northern Zapotec varieties in having a four-vowel inventory. The vowel /o/ is marginal in the lexicon and is often replaced by /u/ in younger speakers. The vowels /o/ and /u/ occupy nearly identical vowel spaces, and the overlap between the two vowels is not complete but is also not completely separating.

2.3 Phonation in Santiago Laxopa Zapotec

One of the defining characteristics of the Oto-Manguean languages is their use of contrastive phonation (Campbell 2017a,b). This fact is so common and widespread that it is has been reconstructed for the proto-language (see Rensch 1976). Additionally, the use of contrastive phonation in Oto-Manguaean languages led Silverman (1997a,b) to create the term *laryngeal complexity* to describe how phonation is realized in these languages and the development of theories of how phonation interacts with tone.

This use of contrastive phonation is prevalent in the Zapotec languages (see Ariza-García 2018 for an overview and typology of the phonation contrasts in the Zapotec language family), with SLZ being no exception. SLZ has a four-way voice quality contrast: modal, breathy, and two types of laryngealization which are checked and rearticulated. These contrasts are exemplified in the near-minimal quadruple in (1).

(1) Four-way near minimal phonation contrast

a. *yag* [çax˩] 'tree; wood; *almúd* (unit of measurement ∼4kg)'

b. *yah* [ça̤˩] 'metal; rifle; bell'

c. *cha'* [tʃaˀ˩] 'cooking pot'

d. *ya'a* [çaˀa˩] 'market'

In representing the checked and rearticulated vowels, I follow Avelino (2010) and Uchihara (2016) in representing laryngealization as a superscript glottal stop in the IPA transcriptions (i.e., [aˀ] or [aˀa]). This allows for a standardized way to discuss these phonations without referencing their variable pronunciation, see Sections 2.3.2 and 2.3.3 for a more detailed discussion of checked and rearticulated phonations respectively.

2.3.1 Breathy vowels in SLZ

SLZ is also unique in regards to its voice quality contrasts because it is a Northern Core Zapotec that has developed breathy voice, which has not been described in any of the neighboring Sierra Norte varieties (Nellis & Hollenbach 1980, Jaeger & Van Valin 1982, Butler H. 1997, Avelino 2004, Sonnenschein 2005, Long & Cruz 2005). However, breathy voice does appear in other Zapotec languages most frequently in Central Valley Zapotecs (Munro & Lopez 1999, Esposito 2004, 2010b, Uchihara 2016, Ariza-García 2018). Breathy voice is characterized by aperiodicity in the harmonic structure of throughout the vowel or a portion thereof depending on the speaker.

An example of breathy voice is shown in Figure 2.4, which is the spectrogram and waveform for the word *yah* [ça̤˩]‘metal; rifle; bell’ produced by a female speaker in their 50s. In this figure, we see that the breathy voice on the vowel shows a large amount of noise as evidenced by the aperiodicity in the waveform and the haziness seen in the spectrogram.

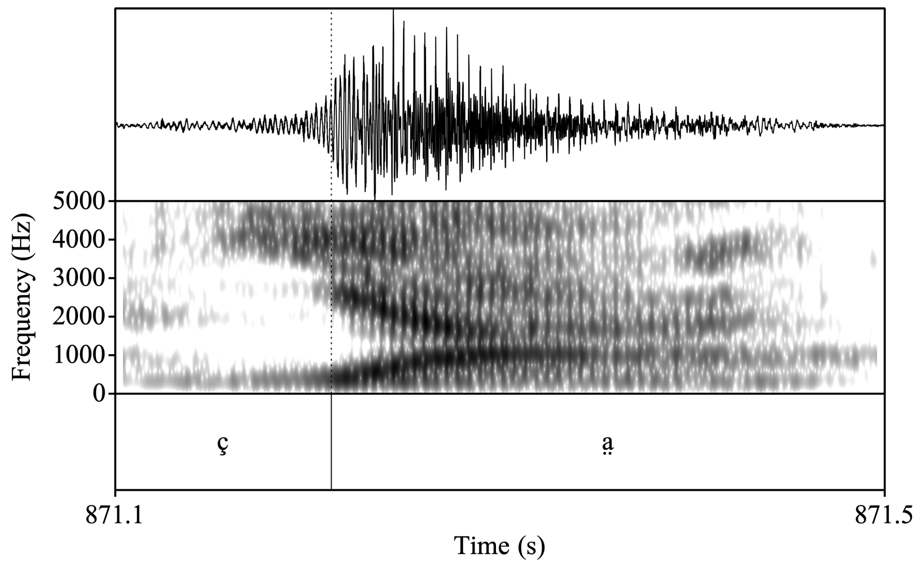


Figure 2.4: Breathy vowel in the word *yah* 'metal; rifle'

As will be discussed in Chapters 5 and 6, breathy voice is also characterized as having a longer duration than modal voice and having a lower fundamental frequency ($f0$) than the other phonation types with the same tone. This behavior is consistent with the behavior of breathy voice in other languages (e.g., Blankenship 1997, Brunelle 2012, Brunelle & Kirby 2016, Hillenbrand & Houde 1996a).

2.3.2 Checked vowels in SLZ

SLZ shares with other Zapotec variates two types of laryngealization: checked and rearticulated. These two types of laryngealization are characterized by the presence of a glottal occlusion or some form of creaky voice that is produced in a portion of the vowel. Checked vowels are typically produced with a glottal stop at the end of the vowel (i.e., /aʔ/) or a period of creakiness at the end of the vowel (i.e., /aa̰/).

As seen in Figure 2.5, which was produced by a male speaker in his mid-twenties, the checked vowel in the word *yu'* [çuˀ˨] 'earth' is characterized by a rapid decrease in amplitude and aperiodicity at the end of the vowel.

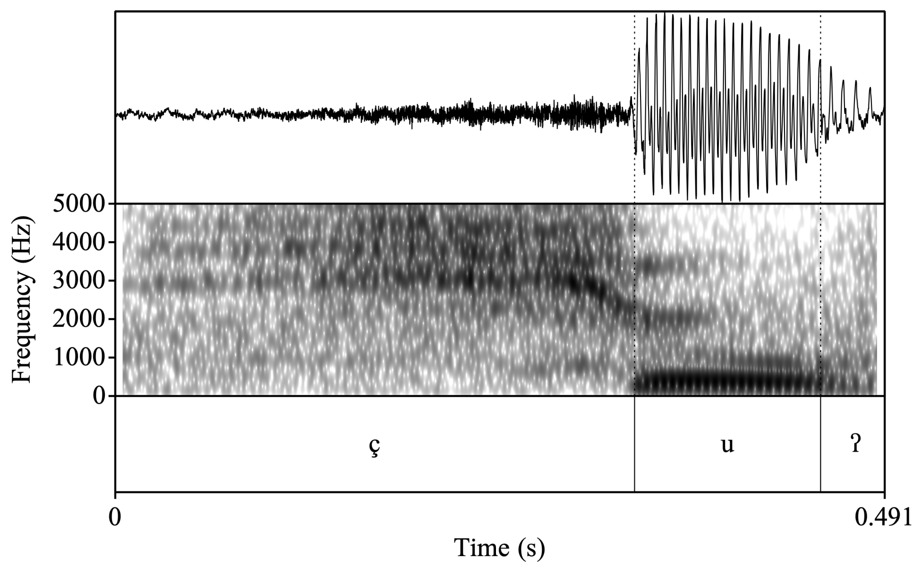


Figure 2.5: Checked vowel in the word *yu'* 'earth'

As will be discussed in Chapters 5 and 6, checked vowels are also characterized as having a longer duration than modal voice and the second half of the vowel is characterized as having a decrease in voicing amplitude and the presence of creaky voice.

2.3.3 Rearticulated vowels in SLZ

Rearticulated vowels are the term used by Oto-Mangueanists to describe the class of laryngealized vowels which have a glottal occlusion or some sort of glottalization in the middle of the vowel (Ariza-García 2018) and are distinct from other uses of the term such as rearticulated vowels in Mayan languages (Baird 2011). Linguists have also used other terms such as interrupted, laryngealized, broken, creaky, et cetera (Adler & Morimoto 2016, Ariza-García 2018, Avelino 2010, Long & Cruz 2005, Sonnenschein 2005). I will be using the term rearticulated to refer to these vowels throughout the dissertation.

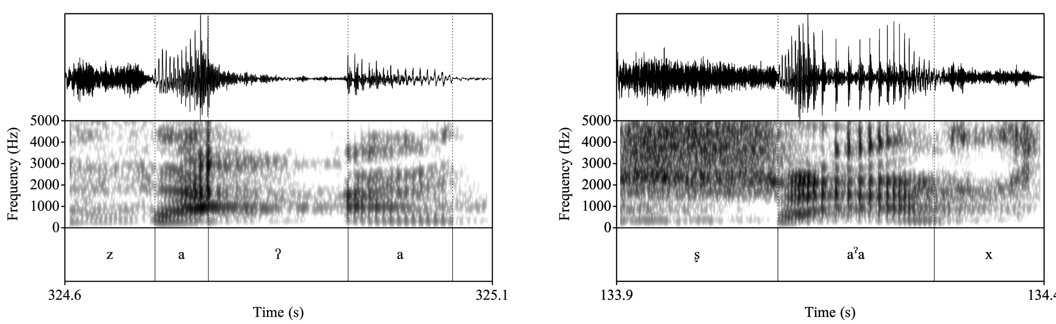
These vowels are also highly variable in their production, hence the large number of terms used to refer to these vowels. For some speakers, they are produced with a full glottal stop in the middle of the vowel (i.e., /aʔa/), while others produce a vowel with apparent modal voice but with a drop in amplitude (similar to what Gerfen & Baker 2005 found for some Mixtec varieties), while others produce a creaky voice throughout the entire vowel. Some speakers produce a combination of these unique productions. This is also consistent with the findings of Avelino (2010) who found that the rearticulated vowels in Yalálag Zapotec had at least four different pronunciations among his consultants, as shown in Table 2.3.

Figures 2.6 and 2.7 show how two different speakers produce these vowels. Figure 2.6 shows how a female speaker in her 50s produces this vowel. Figure 2.6a shows

Table 2.3: Layngealized Vowels in Yalálag Zapotec

/Vˀ V/	[VʔV]
	[VV̰V]
	[VV̰ːV̆]
	[VV̰V̰]

the rearticulated vowel in the word *za'a* [zaˀa˩˥] 'corncob'. In this spectrogram and waveform FSR produces a modal vowel that is clearly divided by a glottal stop. This is the typical pronunciation used by this speaker. However, when the rearticulated vowel cooccurs with a low tone, like in the word *xa'ag* [ʂaˀax˩] 'topil', FSR produces the vowel with creaky voice, as seen in the spectrogram in Figure 2.6b.

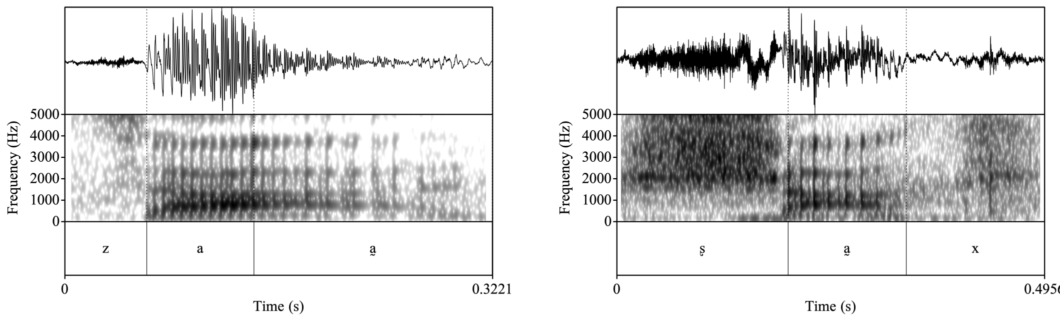


(a) *za'a* [zaˀa˧˥] 'corncob'

(b) *xa'ag* [ʂaˀax˩] 'topil'

Figure 2.6: FSR's rearticulated vowels in *za'a* 'corncob' and *xa'ag* 'topil'

However, RD, a male in their mid-20s, produces the rearticulated vowels consistently with creaky voice. This is shown in Figures 2.7a and 2.7b which are the same words as found in Figure 2.6.



(a) *za'a* [zaˀa˧˥] 'corncob'

(b) *xa'ag* [ʂ̡aˀax˩] 'topil'

Figure 2.7: RD's rearticulated vowels in *za'a* 'corncob' and *xa'ag* 'topil'

As will be discussed in Chapters 5 and 6, rearticulated vowels are also characterized as having a longer duration than modal voice and the first half of the vowel is characterized as having a decrease in voicing amplitude and the presence of creaky voice.

2.4 Tonal contrasts in Santiago Laxopa Zapotec

One of the most well known features of all Oto-Manguean languages is the fact that they are tonal languages and exhibit a large range of tonal systems (Pike 1948, Rensch 1976, Josserand 1983, Silverman 1997a, Beam de Azcona 2007, DiCanio 2010, 2012a, Elliott, Edmondson & Cruz 2016, Campbell 2017a,b, Lillehaugen 2019, Eischens 2022). SLZ has a five-way tonal contrast which consists of three level tones (high, mid, low) and two contour tones (rising and falling).

Table 2.4: Examples of the five tonal patterns observed in the Santiago Laxopa Zapotec monosyllables.

High	*xha*	[ʐ̨a˥]	‘clothing.POSS’
Mid	*lhill*	[ɾiʒ˧]	‘house.POSS’
Low	*yu’*	[çuˀ˩]	‘earth’
Rising	*yu’u*	[çuˀu˧˥]	‘quicklime (Sp. cal)’
Falling	*yu’u*	[çuˀu˥˩]	‘house’

Following discussion from Brinkerhoff, Duff & Wax Cavallaro (2022) these tones appear to be limited in their distribution. It is true that all five patterns can surface on a syllable but there is a restriction in what tonal patterns are allowed to surface on words that are larger than bimoraic. The patterns that we observe on bimoraic nominals are: HL, MH, and LL.

2.5 Interactions between tone and voice quality

Most previous work on the interaction of tone and phonation has been focused on the languages of East and Southeast Asia (e.g., Masica 1976, Thurgood 2002, Yip 2002, Enfield 2005, Michaud 2012, Brunelle & Kirby 2016). What has been found in these descriptions is that certain tones and phonations are codependent. For example, Smalley (1976) and Ratliff (1992) both describe White Hmong’s *-g* tone as being a mid-low tone with breathy phonation, and Mandarin’s tone 3 is often associated with creaky phonation (Hockett 1947). Brunelle (2009) found that creaky phonation plays an important role in producing certain tones. Additionally, work on S'gaw Karen has found that two tones are only differentiated by some form of non-modal phonation (Boehm p.c.).

However, there have been some observations–especially in Mesoamerica–that tone and phonation are independent of each other (e.g., Silverman 1997a, Garellek & Keating 2011). This means that tone can independently occur with any phonation type. This has also been extensively described in multiple Zapotecan languages (e.g., Avelino 2004, 2010, Chávez-Peón 2010, Campbell 2011, Villard 2015, López Nicolás 2016)

Chávez-Peón (2010) describes the tone and phonation interactions in San Lucas Quiaviní Zapotec (SLQZ), a central valley variety of Zapotec. The distribution of tone and phonation is found in Table 2.5. We see that in SLQZ, both low- and falling-tones have the full range of possible combinations. However, we see gaps in the high-tone for breathy and rising tones that can only occur with modal phonation.

Table 2.5: SLQZ tone and phonation interactions (Chávez-Peón 2010).

	Modal	**Breathy**	**Creaky**	**Interrupted**
High	✓	–	✓	✓
Low	✓	✓	✓	✓
Falling	✓	✓	✓	✓
Rising	✓	–	–	–

Based on elicitation data collected from 2020-2022, SLZ has a more expansive distribution of tone and phonation when compared to SLQZ but seems to be very similar to other Northern Zapotec varieties (e.g., Avelino 2004). The distribution of SLZ tonal and phonation combinations are given in Table 2.6.

Table 2.6: SLZ tone and voice quality combinations.

	Modal	**Breathy**	**Checked**	**Rearticulated**
High	✓	–	✓	✓
Mid	✓	✓	✓	✓
Low	✓	✓	✓	✓
High-Low	✓	✓	✓	✓
Mid-High	✓	✓	–	✓

One of the striking things in this is the lack of high tone with breathy phonation. This gap is interesting because of the long-time association of high pitch with breathiness (for an overview, see Esling et al. 2019). This gap is common across the Zapotecan languages that have breathy voice (Campbell p.c.). Regarding breathy phonation in SLQZ, one of the Valley Zapotec varieties, Uchihara (2016) offers some convincing evidence that the phonation originated in syllables with low tone and then spread to other tones via analogy. Investigating the origin of breathy voice in SLZ would be important in understanding how breathy vowels originated in the Zapotecan family where breathy voice is rare typologically (Ariza-García 2018). Such a study is beyond the scope of this paper.

Chapter 3

On using Residual H1* for voice quality research

3.1 Introduction

Languages use voice quality distinctions to convey phonemic distinctions (Garellek 2019) and to convey paralinguistic information by "indexing the biological, psychological, and social characteristics of the speaker" (Laver 1968, Podesva 2016). Voice quality contrasts have been studied extensively by examining their correlates in the acoustic signal (e.g., Esposito & Khan 2020), resulting in a large and complex literature on acoustic correlates of phonation differences (see Garellek 2019).

[0] The majority of the chapter was previously published as Brinkerhoff & McGuire (2025). This chapter adds to Brinkerhoff & McGuire by including discussion of a generalized additive mixed model (Hastie & Tibshirani 1986, Wood 2017) analysis of H1*−H2* and residual H1*.

One measure that Fischer-Jorgensen established, the fundamental's relative strength-ive amplitudes of the first harmonic and second harmonic. As established by Fischer-Jorgensen, the relative strength of the fundamental is a correlated measure of breathy voice in contrast with modal voice Fischer-Jørgensen (1968). In order to normalize the amplitude of the fundamental and counteract some of the effects of high-pass filtering and differences in sound pressure in the signal, she proposed that you could subtract the amplitude of a higher harmonic, in this case, the second harmonic (H2), from the amplitude of the fundamental (H1). Since its introduction, H1*−H2* has been used in many studies to measure not only breathy voice but other voice quality contrasts as well (Garellek 2019, Chai & Garellek 2022).

Despite the large amount of evidence in support of H1*−H2*, it is not without its problems. At the fundamental level, it is not clear that H1*−H2* adequately measures the strength of the fundamental. Sundberg (2022) found that H1 and H2 are affected differently by subglottal pressure, compromising some of the original reasoning behind the use of H1*−H2* from Fischer-Jørgensen (1968). Similarly, in a comprehensive overview of the main concerns with H1*−H2* as a phonation type measure, Chai & Garellek (2022) found that in addition to the issues mentioned above, errors in measuring H1*−H2* are uncomfortably high. This is mainly due to the need to precisely measure two different harmonic amplitudes; when there are errors in calculating H1, this, in turn, leads to errors in calculating H2 (Arras 1998). An example of this type of error propagation is that errors in measuring the fundamental frequency, which is especially common with non-modal phonation, are introduced into measuring harmonics because they are based on the fundamental. Despite algorithms correcting for vowel height, a common error that occurs is when a high fundamental frequency co-occurs with a low first formant (Chai & Garellek 2022). This situation causes errors in tracking the fundamental frequency and the first formant. A final issue that can occur when measuring the harmonics is in contexts where the vowel is nasalized. Simpson (2012) shows that in these nasalized contexts, the first nasal pole (P0) can increase the amplitude of H2 and, when the fundamental frequency is high, H1 increases instead.

This collection of errors leads Chai & Garellek (2022) to propose a new measure, residual H1*. This measure is calculated by first regressing H1 on energy and then subtracting the product of energy and the energy factor from H1. Chai and Garellek argue that this measure better reflects the initial purpose of using H1*−H2*. Furthermore, they find that residual H1*: (i) provides better differentiation between phonation types in !Xóõ; (ii) was more robust for measuring creak in Mandarin with respect to different utterance positions; and (iii) has a stronger relationship to the open quotient than H1*H2* based on a comparison using electroglottogram data.

The contributions Chai & Garellek (2022) make are very intriguing and have the potential to alter how spectral analyses of voice quality are performed. However, they are not convincing on their own. If residual H1* is to be widely adopted in linguistics, speech pathology, and other speech sciences, then it requires considerable evidence of its effectiveness in acoustic studies. This paper offers additional evidence for residual H1*'s effectiveness.

Given the promising nature of this measure, we tested residual H1* with data from Santiago Laxopa Zapotec. Although Chai & Garellek (2022) evaluated residual H1* for !Xóõ and Mandarin, both of which contain tone and voice quality, they do not factor tone into their analysis.[1] This is where testing the effectiveness of residual H1* in Zapotec languages is beneficial. Zapotec languages are often described as laryngeally complex (Silverman 1997a, Ariza-García 2018). Laryngeal complexity is defined as a language that allows contrastive tone and contrastive voice quality that are unrestricted in their interactions. This laryngeal complexity between tone and phonation types presents a unique challenge for acoustic analysis and testing of the residual H1* measure.

Other Zapotec languages have also been studied for voice quality research. For example, Esposito (2010b) found that in Santa Ana del Valle Zapotec there was a biological sex difference in which acoustic measures best capture the voice quality contrasts similar to observations from Klatt & Klatt (1990). Arellanes Arellanes (2010) found that the realization of laryngealization in a variety of San Pablo Güilá Zapotec is highly variable and depends on the position of laryngealization in the phrase. This

[1] !Xóõ is most like Zapotec in that it has three phonologized phonation categories. It also has tone, which to our understanding is not restricted by phonation and which is not analyzed in Chai & Garellek (2022). The Mandarin case is less relevant as its phonation is prosodic and tonally linked.

was also found to be the case in Betaza Zapotec (Crowhurst, Kelly & Teodocio 2016). Barzilai & Riestenberg (2021) found that tone and phrasal position also played a role in how voice quality is realized in San Pablo Macuiltianguis Zapotec. These studies show that Zapotec languages have much to contribute to our understanding of voice quality.

We find that residual H1* can adequately capture differences in voice quality and is a more robust measure of voice quality than H1*−H2*, adding credence to the use of this measure instead of H1*−H2* in voice quality research. The remainder of this paper is organized as follows. Section 3.2 provides a brief overview of the Santiago Laxopa Zapotec language. Section 3.3 describes the methods used in the data collection, data processing, and statistical modeling used in this study. Section 3.4 presents the results of the study. Section 3.5 concludes the paper.

3.2 Santiago Laxopa Zapotec

Santiago Laxopa Zapotec is a Northern Zapotec language of the Oto-Manguean language family (Adler & Morimoto 2016, Adler et al. 2018, Foley, Kalivoda & Toosarvandani 2018, Foley & Toosarvandani 2022, Sichel & Toosarvandani 2020a,b, Brinkerhoff, Duff & Wax Cavallaro 2021, 2022). It is spoken by 981 people in the municipality of Santiago Laxopa, Ixtlán, Oaxaca, Mexico and a small number of other speakers from the diaspora in Mexico and the United States.

Santiago Laxopa Zapotec exhibits a four-vowel inventory (that is, /i/, /e/, /a/, and /u/), which is further distinguished by a four-way contrast in voice quality. This variety is unique because it is a Northern Zapotec that has developed a breathy voice (/V̤/) in addition to the two types of laryngealization that characterize the rest of the Zapotec languages, namely checked and articulated. Checked vowels are defined as a modal vowel that ends with a period of creaky voice or a glottal closure (/VV̰/ or /Vˀ/). Rearticulated vowels are also defined as a modal vowel that also has a period of creakiness or glottal closure but aligned to the middle of a vowel (/VV̰V/ or /VˀV/). This makes an otherwise modal vowel appear as if it is interrupted by this laryngealization.[2] This difference in laryngeal timing is one of the key differences between these phonations (see Ariza-García 2018 for a detailed typological study of voice quality distinctions in Zapotec languages).

Santiago Laxopa Zapotec is also tonal with three level tones (H, M, and L) and two contours (MH and HL) appearing in nominals (Brinkerhoff, Duff & Wax Cavallaro 2022).[3] The language has a complex interaction between tone and phonation types. Every tone can appear with every phonation type, with two exceptions being that breathy voice cannot appear with the high tone and checked voice cannot appear with the rising contour tone. It is unclear whether these are accidental gaps or have

[2] This is different from how rearticulated vowels are defined in other languages, where they are a modal vowel followed by an "echo vowel" which may or may not have a glottal closure intervening between the modal and "echo" vowel (see Baird 2011). A fuller description of how these vowels are realized in one variety of Zapotec and the terms used are found in Avelino (2010).

[3] The tonal system of Santiago Laxopa Zapotec for verbs and other lexical categories is still being evaluated.

a phonetic underpinning.

These interactions between voice quality and tone present a rich environment for testing the reliability of voice quality measures in laryngeally complex languages.

3.3 Methods

3.3.1 Elicitation

Ten native speakers of SLZ (five female; five male) participated in a wordlist elicitation. Elicitation was performed in the pueblo of Santiago Laxopa, Ixtlán, Oaxaca, Mexico during the summer of 2022 using the built-in microphone of a Zoom H4n handheld recorder (16-bit, 44.5 kHz).

The wordlist consisted of 72 items repeated three times in isolation and the carrier sentence *Shnia' X chonhe lhas* [ʃnːiaˀ X tʃone ɾas] "I say X three times". This sentence was chosen to minimize any effect of the phrasal position (see Crowhurst, Kelly & Teodocio 2016: for a study about phrasal effects on voice quality in Zapotec). Additionally, there are no co-articulatory effects on voice quality found with this frame and the items in the wordlist. Between these 72 words, there were 11 words with breathy voice, 9 with rearticulated voice, 10 with checked voice, and 42 with modal. Thirteen of the 72 words were disyllabic and eight of the thirteen contained the same voice quality in each syllable. Of those 13 disyllabic word, only five contained different voice qualities in each of the syllables. This resulted in 85 different syllables.[4]

The token selection for the wordlist was done in consultation with a native speaker. Similarly to Barzilai & Riestenberg (2021), we did not balance the tones with the different voice qualities. The frequency in which certain tones co-occur with certain voice qualities is uneven in the language, making it difficult to control for. The distribution of tones and voice quality across the 85 syllables used in this study are presented in Table 3.1. This imbalance was taken into account by including Tone as a fixed effect in our statistical models in Section 3.3.3.

Table 3.1: Distribution of tone and voice quality in the wordlist

	High	Mid	Low	Rising	Falling
Modal	14	9	15	2	10
Breathy	—	—	11	—	2
Checked	1	—	9	—	
Rearticulated	1	—	4	—	4

3.3.2 Data Processing

Each vowel of the target words in the carrier sentence condition was labeled following Garellek (2020) for where the vowel began and ended. Each vowel in the word list was annotated for speaker, word, vowel, tone, voice quality, and utterance number.

[4]There is ongoing debate about the status of strong syllables in Zapotec. The majority of evidence for the presence of stress comes from non-Northern Zapotec varieties (e.g., Chávez-Peón 2008, Mock 1988). At this time, evidence for stress in Northern Zapotec remains to be seen.

This labeling was conducted for each vowel in the target word from the elicitation list of the carrier sentences.

These vowels were then extracted and fed into VoiceSauce for acoustic measurement (Shue et al. 2011). The formants were measured using Snack (Sjölander 2004), while the fundamental frequency (*f0*) was measured using the STRAIGHT algorithm (Kawahara, Cheveigne & Patterson 1998). Spectral slope measures were corrected for formants and bandwidths (Hanson 1997, Iseli, Shue & Alwan 2007). Each vowel was measured with ten equal time intervals, resulting in 22890 data points in total.

The data were cleaned of outliers following the same steps as taken by Chai & Garellek (2022) in their study. The H1*, H1*–H2*, and *f0* values were z-scored by speaker to reduce the variation between the speakers and provide a way to directly compare the different measures on the same scale. Data points with an absolute z-score value greater than 3 were considered outliers and excluded from the analyses. Within each vowel category, we calculated the Mahalanobis distance in the F1-F2 panel. Each data point with a Mahalanobis distance greater than 6 was considered an outlier and excluded from the analysis. Using the Mahalanobis distance allows us to compare the data points to the mean of the F1-F2 panel for each vowel category. The larger the Mahalanobis distance is the more deviant the data point is from the mean which in turn means that the data point was improperly tracked. This is comparable to what was done in Seyfarth & Garellek (2018), Chai & Ye (2022), and Garellek &

Esposito (2023).

Time points whose $f0$, F1, or F2 values were outliers were also excluded from H1* and H1*−H2* analyzes because H1* and H1*−H2* are calculated based on *f0*, F1, and F2. Energy was excluded if it had a zero value and then logarithmically transformed to normalize its right-skewed distribution. Afterward, the resulting logarithmically transformed data was z-scored, and any data point with a z-score greater than 3 was excluded. This outlier removal resulted in 1918 datapoints being removed.

After removing the outliers, we calculated residual H1* for the remaining data points following Chai & Garellek (2022). First, a linear mixed effects model was generated with the z-scored H1* as the response variable and the z-scored energy as the fixed effect. The uncorrelated interaction of the z-scored energy by speaker was treated as random. The energy factor resulting from this linear mixed-effects model was extracted. Finally, the z-scored H1* had the product of the z-scored energy and the energy factor subtracted from it, giving us the residual H1* measure.

The measures were then orthogonally coded according to their position in the vowel (first, middle, and third) and for tone (H, M, L, R, F) for statistical modeling.

3.3.3 Statistical modeling

Two linear mixed-effects regression models were fitted, one for the normalized H1*−H2* and residual H1*. Each model had the tone and the interaction between voice quality and position in the vowel as fixed effects. Vowel and interaction between the speaker, the word, and the repetition were treated as random intercepts.[5]

The tone and the interaction between voice quality and position in the vowel were selected as fixed effects for several reasons. The first is that five unique tones appeared in the data and it is well established that tone interacts with voice quality in different ways (see Esposito & Khan 2020 and Garellek 2019 for discussion). By treating tone as a fixed effect in our model, we can account for these interactions. The interaction between voice quality and position in the vowel as a fixed effect was included to account for the temporal differences between the two different laryngealizations, checked and rearticulated vowels. Checked vowels in Zapotec languages have a glottal occlusion or a short period of creaky voice located at the right edge of the vowel. This is in contrast to rearticulated vowels, where there is a glottal occlusion or creaky voice in the middle of the vowel. Because this difference between checked and rearticulated vowels is temporal in nature, we can account for this difference through the interaction of voice quality and position in the vowel.[6]

The interaction between speaker, word, and repetition was treated as a random intercept because this allows us to consider that each speaker said each word on the elicitation list three times. This intercept accounts for not only the intra-speaker variability, but also the inter-speaker variability during each time the word was uttered.

[5] $Measure \sim Phonation * Position + Tone + (1|Speaker : Word : Repetition) + (1|Vowel)$

[6] Tone and voice quality are closely linked. By including only the positional interaction with voice quality, we can avoid collinear interactions that appear when we try to include tone in the interaction.

Treating a vowel as a random intercept allows us to capture the fact that each voice quality occurred with different vowels during elicitation.

Additional statistical modeling was performed with two generalized additive mixed models (GAMM) using the `mgcv` package in R (Wood 2017) and plotted using the `tidygam` package (Coretta 2024). Phonation was treated as a fixed effect to model the main effect of phonation on the acoustic measures. A smooth term for time (i.e., measurement number) was included to model the nonlinear relationship between time. A second smooth term was included to allow time to vary by phonation type. Random effects for speaker and the interaction between speaker and phonation were included to account for variability in the data due to individual differences between speakers and individual differences in how they produce the phonation types. A tensor product interaction between time and repetition was also included as a random effect to account for the variability in the data due to the different repetitions the speakers were asked to complete. The inclusion of tone and word did not improve the model fit and, therefore, were excluded from the final model.

3.4 Results

In interpreting the results, there are certain expectations about how these measures should capture the voice quality contrasts. As discussed in Garellek (2019), it is expected that breathy vowels should have a higher spectral slope than modal vowels across the two measures in keeping with observations about breathy vowels in other languages. Because both checked and rearticulated vowels make use of creaky voice, they should have a lower spectral slope than modal vowels. Additionally, because there is a temporal distinction between checked and rearticulated with where the creakiness appears, this should also be captured in the two measures.

3.4.1 H1*−H2*

Figure 3.1 shows the mean H1*−H2* values for each voice quality at each of the ten vowel intervals. We see that the breathy, checked, and rearticulated values are lower than the modal values at each of the first nine intervals. In the final intervals, breathy and rearticulated are essentially equal to the modal value. In contrast, checked's value remains lower than the modal's value throughout the entire vowel. This measure does not match the expectations discussed above.

3.4.2 Residual H1*

Figure 3.2 shows the mean residual H1* values for each voice quality at each of the ten vowel intervals. In contrast to Figure 3.1, we see that breathy has a higher residual H1* measure than modal throughout the duration of the vowel, which is consistent with other observations for breathy voice (Fischer-Jørgensen 1968). Checked and rearticulated both have lower values than the modal at each of the 10 intervals.

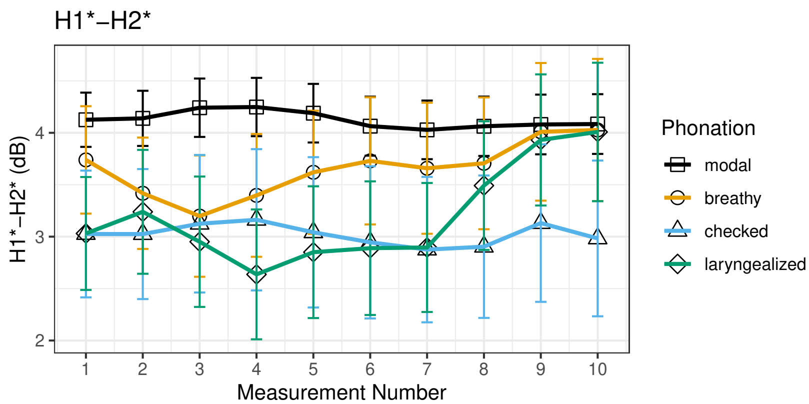


Figure 3.1: H1*—H2* across the duration of the vowel. Points represent the mean of each measure across the ten intervals. The error bars around each point represent a 95% confidence interval. A line was plotted over each to show how the acoustic measure functions across the ten intervals.

In addition, it shows that the checked voice has a lower residual H1 * value than the rearticulated voice at intervals 8 through 10. The rearticulated voice has a lower residual H1 * value than the checked voice at intervals 1 through 7, showing the temporal distinction between these two voice qualities. This measure complies with the expectations discussed above.

3.4.3 Model Comparison

To assess the robustness of the models, we compared the residual H1* linear mixed-effects model to the H1*—H2* linear mixed-effects model. This was done using two methods: direct comparison of the outputs of the two models in the same way as

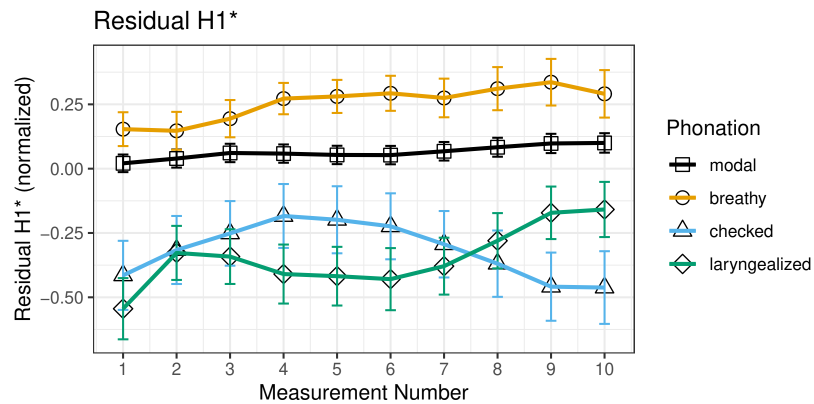


Figure 3.2: Residual H1* across the duration of the vowel. Points represent the mean of each measure across the ten intervals. The error bars around each point represent a 95% confidence interval. A line was plotted over each to show how the acoustic measure functions across the ten intervals.

Chai & Garellek (2022) and the Akaike Information Criterion (AIC).

Table 3.2 compares the linear mixed effects models for H1*–H2* and residual H1*. In comparing these models, we find that the residual H1* model performed better than the H1*–H2* model in distinguishing voice quality contrasts in Santiago Laxopa Zapotec. This is supported by the larger absolute value of the coefficient estimate, the lower standard error, and the higher t-value of the residual H1* to distinguish breathy, checked, and rearticulated vowels from modal vowels.

Table 3.3 shows the results of the AIC comparison between the H1*–H2* and residual H1* models. The residual H1* model had a lower AIC than the H1*–H2* model, indicating that the residual H1* model is a better fit for the data than the H1*–H2*

Table 3.2: Model comparison between H1*–H2* and Residual H1* in distinguishing Santiago Laxopa Zapotec voice quality.

Voice Quality Contrast	Model	β	Std. Error	*t*-value	*p*-value	
Breathy vs Modal	H1*−H2*	0.04631	0.03806	1.21680	0.22372	
	Res. H1*	0.23625	0.02866	8.24177	<0.001	***
Checked vs Modal	H1*−H2*	-0.11880	0.03476	-3.41793	<0.001	***
	Res. H1*	-0.40099	0.02621	-15.30098	<0.001	***
Rearticulated vs Modal	H1*-H2	-0.09175	0.04588	-1.99968	0.04560	*
	Res. H1*	-0.44162	0.03450	-12.80027	<0.001	***

model. Although AIC comparison is usually performed on nested models, it is still a useful tool for comparing non-nested models (Burnham & Anderson 2004b, Burnham, Anderson & Huyvaert 2011, Burnham & Anderson 2004a).

Table 3.3: AIC for the H1*−H2* and residual H1* models.

Model	AIC	Δ AIC
H1*−H2* model	43443.33	11182.76
Residual H1* model	32260.57	0

3.4.4 GAMM analysis and model comparisons

The GAMM analysis shows that there are nonlinear effects present in the data. This was expected due to the dynamic nature of voice quality in SLZ. Due to the nature of GAMM analyses, it is important to visually inspect the results to see how well the model fits the data. The model that best fits the data is the one that shows the clearest distinction between the different voice qualities.

Figure 3.3 shows the predicted values for H1*—H2* from the GAMM. We see that the different voice qualities are very difficult to observe clearly. We see that modal and breathy voice occupy the same space. However, checked and rearticulated voice are more separated from modal, with both appearing lower in the graph, which is expected because of creaky voice. We also observe that the rearticulated voice begins to increase throughout the final third of the vowel until it is higher than the modal voice.

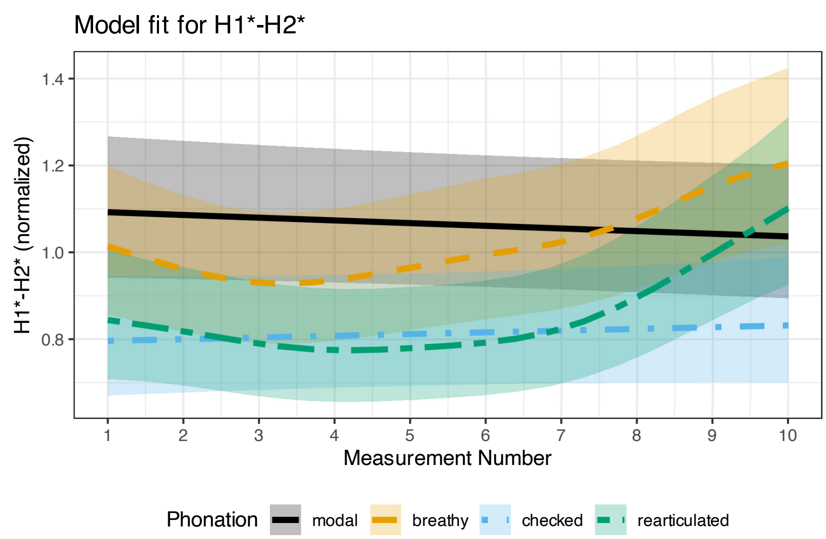


Figure 3.3: GAMM smooths for H1*—H2* across the duration of the vowel.

The difference plots in Figure 3.4 show three panels for a comparison between the modal voice and each of the other voice qualities. The leftmost panel shows that the difference between modal and breathy voice never reaches significance. This means that the breathy voice and the modal voice are essentially the same according to the model. The middle panel shows the difference between modal and checked voice. This shows that checked voice is significantly different from modal voice throughout the duration of the vowel, except for a short period of time right before measurement number 10. The rightmost panel shows the difference between modal and rearticulated voice. This panel shows that rearticulated voice is significantly different from modal voice for the first two-thirds of the vowel. However, in the final third of the vowel, rearticulated voice is not significantly different from modal voice.

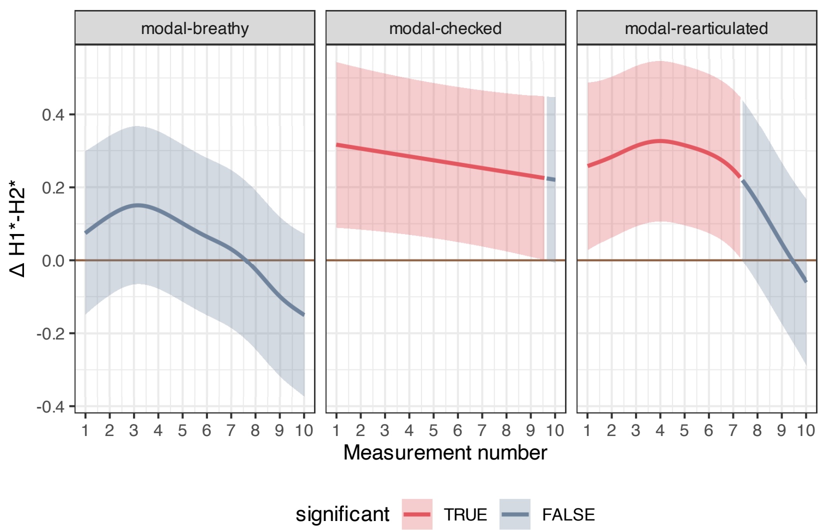


Figure 3.4: GAMM difference plots for H1*−H2* across the duration of the vowel.

The residual H1* GAMM showed a clearer distinction between the different voice qualities than H1*−H2*. This is especially clear in the smooth plot of the predicted values in Figure 3.5. In this plot, we see that beginning at measurement number 3 breathy voice is higher than modal voice. This is expected because the breathy voice is characterized by a higher spectral slope than the modal voice. We also see that the checked and rearticulated voice are lower than modal voice, which is also expected because of the presence of creaky voice, which is characterized by a lower spectral slope than modal voice. Additionally, we see that the temporal difference between checked and rearticulated voice is present in the smooth plot. The checked voice is higher than the rearticulated voice for the first two-thirds of the vowel, but then the checked voice becomes lower than the rearticulated voice in the final third of the vowel. This is expected because checked voice is characterized by a glottal occlusion at the end of the vowel, while rearticulated voice is characterized by a glottal occlusion in the middle of the vowel. Furthermore, this plot suggests that rearticulated voice is predominately more creaky than checked voice in the first two thirds of the vowel. This suggests that the glottal occulsion in the middle of the vowel predominantly affects the first half of the vowel. This difference might be due to the need to distinguish the rearticulated vowel from the checked vowel more effectively.

Figure 3.6 shows the difference plots for residual H1*. The leftmost panel shows that the difference we observed in Figure 3.5 between the modal and the breathy voice is significant for the last two thirds of the vowel. The middle panel shows that the difference observed between the modal and checked voice is significant across the entire duration of the vowel. The rightmost panel shows that the difference we observe

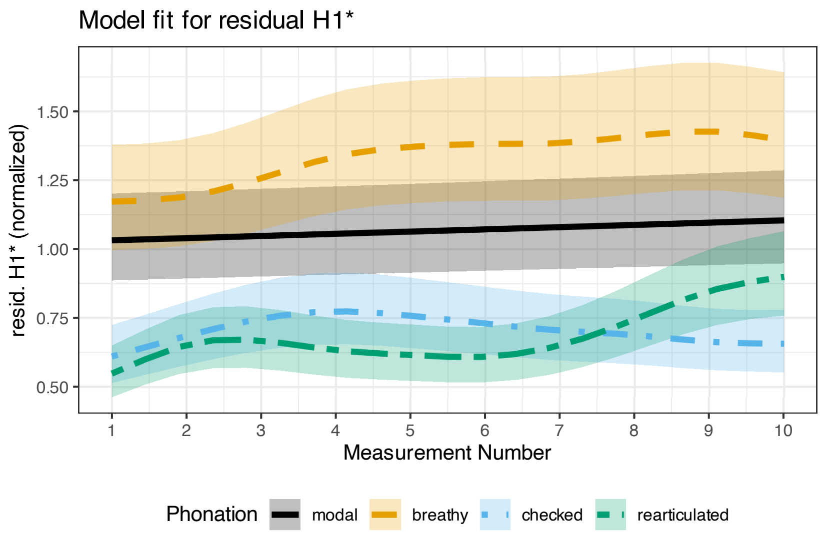


Figure 3.5: GAMM smooths and difference plots for residual H1* across the duration of the vowel.

between modal and rearticulated voice is significant for almost the entire duration of the vowel except for measurement number 10.

In comparing the two GAMM analyses, we find that the residual H1* model provides a clearer distinction between the different voice qualities than the H1*–H2* model. This is further evidence that residual H1* is a more robust measure of voice quality than H1*–H2* in Santiago Laxopa Zapotec. Additionally, the GAMM analysis shows that there appear to be some amount of phasing in the data, which the linear models were unable to capture. This is especially clear in the residual H1* model, where the difference plots show that the breathy voice is only significantly different from the modal in the final two thirds of the vowel. This suggests that the breathy

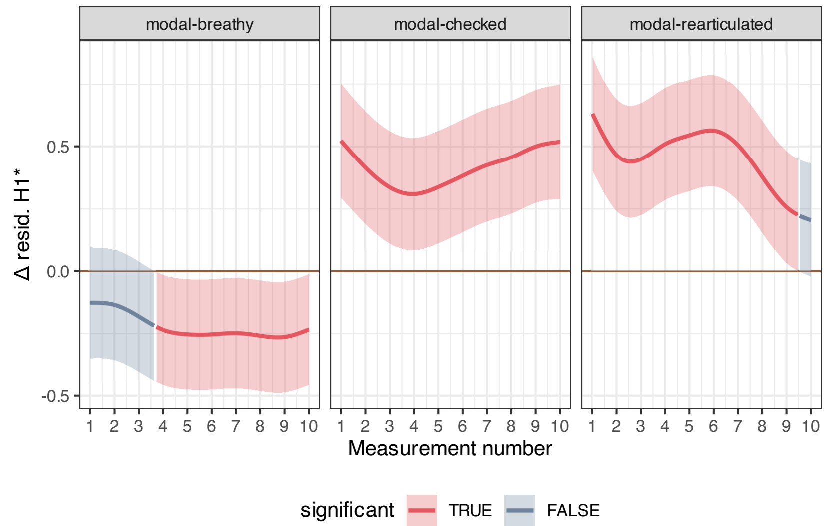


Figure 3.6: GAMM smooths and difference plots for residual H1* across the duration of the vowel.

voice aligns with the end of the vowel. This is similar to the predictions made by Silverman (1997a,b) about the alignment of nonmodal phonation in laryngeally complex languages. This will be further discussed in Chapters 6.

3.5 Conclusion

In conclusion, we find that residual H1* is a more robust measure of voice quality than H1*−H2* in Santiago Laxopa Zapotec. This is supported by the results of the linear mixed-effects models, which show that residual H1* is better at distinguishing breathy, checked, and rearticulated vowels from modal vowels. This is further supported by the AIC comparison, which shows that the residual H1* model is a better fit for the data than the H1*−H2* model. These results lend credence to the claims of Chai & Garellek (2022) and support using residual H1* instead of H1*−H2* in voice quality research, especially in laryngeally complex languages.

However, the results neither suggest nor support the notion that residual H1* should be the sole measure used to determine phonation quality. Instead, we suggest that residual H1* should be included as one of the measures that researchers should consult during their acoustic studies.

Chapter 4

The acoustic space of voice quality in Santiago Laxopa Zapotec

4.1 Introduction

This chapter studies the acoustic dimension of voice quality in Santiago Laxopa Zapotec (SLZ) using a Multidimensional Scaling (MDS) analysis of acoustic data. MDS is a statistical method that reduces the dimensionality of a dataset and visualizes the relationships between data points. This study uses MDS to visualize the acoustic space of voice quality in SLZ. This analysis provides information on the acoustic correlates of voice quality in SLZ and contributes to our understanding of the phonetic properties of this under-documented language.

This study is based on the work conducted by Keating et al. (2023) on the acoustic space of voice quality in 11 languages. However, this study focuses on a single language, SLZ, and provides a detailed analysis of the acoustic properties of voice quality in this language. The results of this study will contribute to our understanding of the phonetic properties of SLZ and how the acoustic properties of voice quality in this language compare with other languages.

4.2 Methods

4.2.1 Participants

This study uses data collected from 10 native speakers of SLZ during the summer of 2022. Participants were recruited from the community of Santiago Laxopa, Oaxaca, Mexico. All participants were native speakers of SLZ. The participants were between 18 and 60 years old and consisted of five males and five females.

4.2.2 Recordings

Participants were asked to perform a word list elicitation task consisting of 72 words. These words were selected to elicit the entire range of types of voice quality in SLZ, including modal voice, the two kinds of creaky (i.e., checked and rearticulated), and breathy voice. The words were selected based on previous research conducted as part of the Zapotec Language Project at the University of California, Santa Cruz (*Zapotec Language Project — University of California, Santa Cruz* 2022). Because participants were not literate in SLZ, the word list was prompted for them by asking them "How do you say [word in Spanish]?" by myself and another researcher in Zapotec. Participants were asked to respond with the desired word in the carrier phrase *Shnia' [WORD] chonhe lhas* "I say [WORD] three times." which was repeated three times. These utterances were recorded in a quiet environment using a Zoom H4n digital recorder. The recordings were saved as 16-bit WAV files with a sampling rate of 44.1 kHz.

4.2.3 Acoustic measuring

The resulting audio files were then processed in Praat to isolate the vowel portion of each word. The onset of the vowel was set to the second glottal pulse after the onset, and the offset of the vowel was set to the last glottal pulse before the decrease in amplitude at the end of the vowel (Garellek 2020). The vowel was then extracted and saved as a separate file for analysis.

These vowels were fed into VoiceSauce (Shue et al. 2011) to generate the acoustic measures for the studies discussed in this dissertation. Because many acoustic measures are based on the fundamental frequency, this measure was calculated using the STRAIGHT algorithm from (Kawahara, Cheveigne & Patterson 1998) to estimate the fundamental frequency in millisecond (ms) intervals. Once the fundamental frequency is calculated, VoiceSauce then uses an optimization function to locate the harmonics of the spectrum, finding their amplitudes.

VoiceSauce then uses the Snack Sound toolkit (Sjölander 2004) to find the frequencies and bandwidths of the first four formants, also at millisecond intervals. The amplitudes of the harmonics closest to these formant frequencies are located and treated as the amplitudes of the formants. These formant frequencies and bandwidths are used to correct the harmonic amplitudes for the filtering effects of the vocal tract, using Iseli, Shue & Alwan's 2007 extension of the method employed by Hanson (1997). Each vowel was measured across ten equal time intervals, resulting in 22890 data points in total. These measures were then z-scored by speaker to reduce the variation between speakers and provide a way to compare the different measures directly on similar scales.

4.2.4 Data processing

Data points with an absolute z-score value greater than three were considered outliers and excluded from the dissertation analyzes. The Mahalanobis distance was calculated in the F1-F2 panel within each vowel category. Each data point with a Mahalanobis distance greater than six was considered an outlier and excluded from the analysis. Using the Mahalanobis distance allows us to compare the data points to the mean of the F1-F2 panel for each vowel category. The larger the Mahalanobis distance is the more deviant the data point is from the mean which in turn means that the data point was improperly tracked. This is comparable to what was done in Seyfarth & Garellek (2018), Chai & Ye (2022), and Garellek & Esposito (2023).

Energy was excluded if it had a zero value and then logarithmically transformed to normalize its right-skewed distribution. Afterward, the resulting log-transformed energy was z-scored and any data point with a z-score greater than three was excluded. This outlier removal resulted in 1918 data points being removed.

All data points were then z-scored by speaker to reduce the variation between speakers and provide a way to compare the different measures directly on the same scale.

The residual H1 * was then calculated for the remaining data points following Chai & Garellek (2022). First, a linear mixed effects model was generated with the z-scored H1* as the response variable and the z-scored energy as the fixed effect. The uncorrelated interaction of the z-scored energy by speaker was treated as random. The energy factor resulting from this linear mixed-effects model was extracted. Finally, the z-scored H1* had the product of the z-scored energy and the energy factor subtracted from it to produce the residual H1* measure.

Once these steps were completed, the mean of each combination of phonation and speaker was taken for the fourth to seventh interval of the vowel. This is similar to what Keating et al. (2023) did by taking the middle of the vowel for their analysis. This choice minimizes the effect of the onset and offset of the vowel on the acoustic measures, which are more likely to be affected by the surrounding consonants and should give us the most accurate representation of the vowel quality. Because z-scores were used, this resulted in negative measures, which presents a problem for MDS analyses. To correct for this, I added the absolute value of the minimum z-score to each measure. This results in a dataset that still preserves the relative differences in the scores while providing a dataset that is all positive for the MDS analysis.

4.2.5 Statistical analysis

Multidimensional scaling analysis (MDS) is a type of dimensionality reduction to visualize the relationships between data points (Kruskal & Wish 1978). Using MDS is especially effective when many variables could contribute to the data. In the case of voice quality, this is especially warranted.

As shown in Kreiman et al. (2014, 2021) and Garellek (2020), voice quality is psychoacoustically complex and a single measure is not enough to capture the full range of voice quality. Instead, multiple measures are required that function as cues for the different types of voice quality. For example, a vowel characterized as having a breathy voice has an elevated spectral-slope and a lower harmonics-to-noise ratio than a modal voice. A creaky voice has a lowered spectral-slope and a lowered harmonics-to-noise ratio.

Because MDS analyses that contain many variables can result in rather unmeaningful results, I chose to focus on the speaker by phonation interaction. This allows us to see how speakers differ in their production of the different voice qualities. This choice to focus on speaker by phonation means that each speaker's production of each of the four phonation contrasts is represented as a single point in the MDS plot (e.g., one point for the first speaker's modal voice, one for their checked voice, one for their rearticulated voice, and one for their breathy voice). This is similar to what Keating et al. (2023) did in their study of the acoustic space of voice quality in 11 languages, except that they compared the language by voice quality interaction. Both of these interactions show us similar information. The analysis of speaker-by-voice quality shows us the acoustic space within a language, while the analysis using language-by-voice quality shows us the acoustic space between languages.

The MDS analysis was conducted using the `metaMDS` function in the `vegan` package (Oksanen et al. 2025) in the R programming language (R Core Team 2024). The Manhattan distance was used to estimate the differences between the speaker-by-phonation pairs. Because the distances are non-Euclidean, the MDS analysis was conducted using the nonmetric option.

This algorithm resulted in a solution that involves several different dimensions. The number of dimensions retained directly affects how well the original data is captured. Too many dimensions and the data are overfitted; too few, and the data are underfitted. To determine the number of dimensions to retain, I used a scree plot to plot the stress of each dimension. As shown in Figure 4.1, most of the data are captured in the first four dimensions. These four dimensions were retained for the analysis.

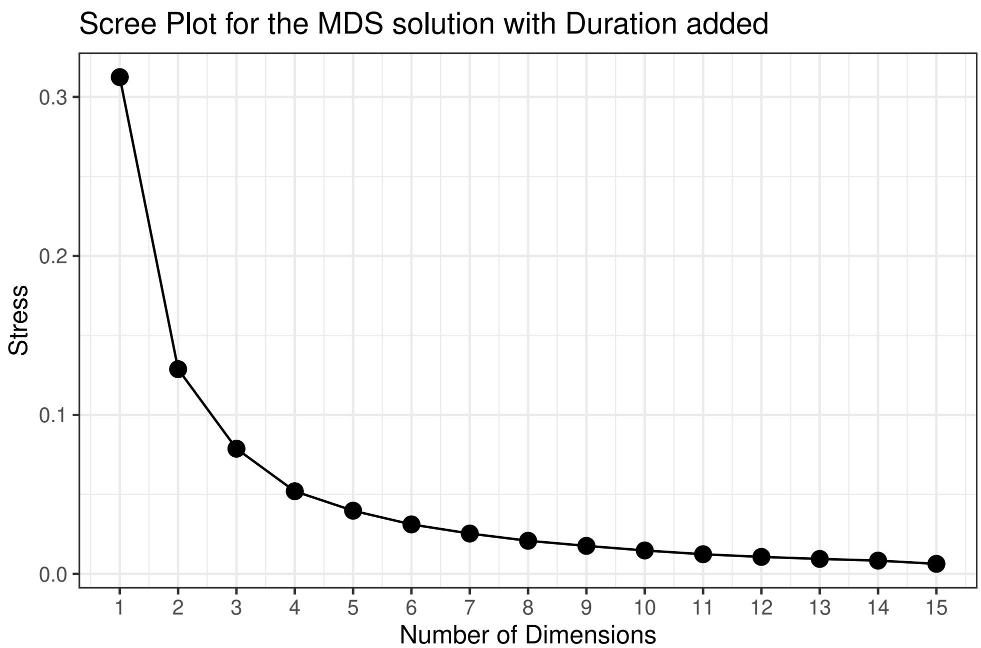


Figure 4.1: Scree plot showing the stress for each dimension for the MDS analysis.

4.3 Results

4.3.1 Acoustic space of voice quality

The results of the MDS analysis show that the acoustical space is represented primarily by a three-dimensional space.[1] In all subsequent plots, breathy voice is represented by orange, checked voice with blue, rearticulated voice with green, and modal voice with black. In each of the plots, the modal voice is generally more densely packed than the non-modal voice qualities. This is likely due to the fact that the modal voice represents approximately 60% of the data, while the non-modal voice qualities represent approximately 40% of the data.

Figure 4.2 shows the first dimension plotted against the second dimension. In this plot, we observe that breathy voice is located in the top left of the plot, modal voice is located in the bottom center of the plot, and the two types of creaky voices are located to the right of the plot, with checked voice located at the extreme right of the plot, and rearticulated voice located closer to the center. From this plot, we see that the first dimension separates breathy, modal, and creaky voices. The second dimension separates the modal voice, bottom of the plot, from the non-modal voice qualities, top of the plot.

Figure 4.3 compares the first dimension with the third dimension. In this plot, we

[1] A 3D plot showing the acoustic space can be found at https://www.mlbrinkerhoff.me/files/3d_-plot.html.

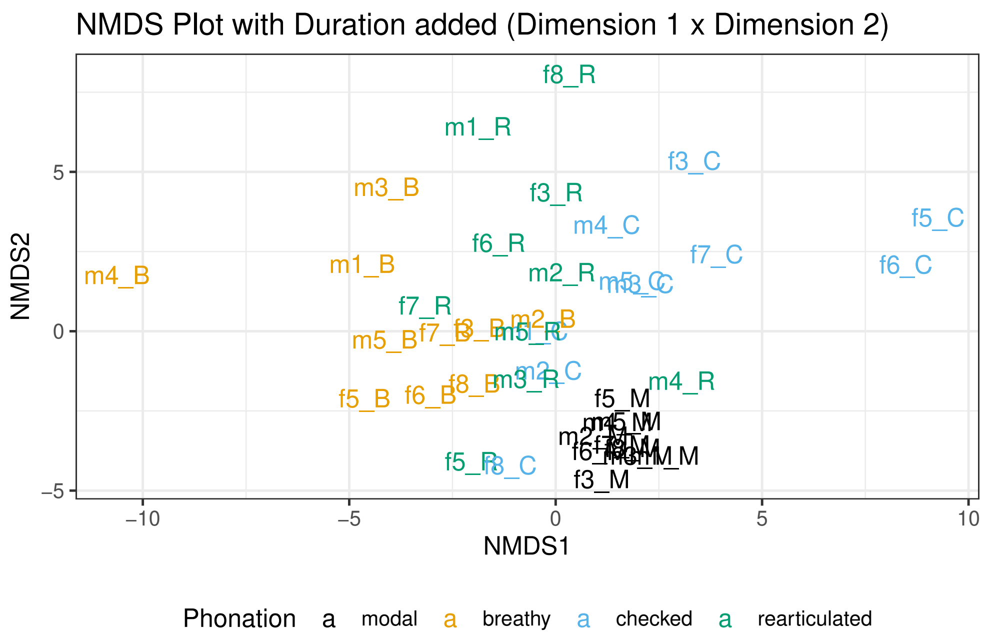


Figure 4.2: Two-dimensional MDS solution showing the first and second dimensions.

observe that the breathy voice is located in the bottom left of the plot, modal voice is located in the very center of the plot, and the two types of creaky voices are located in the top right of the plot. It should be noted that the distribution of the different voice qualities follows a line from bottom left to top right in the plot. This suggests that the first and third dimensions capture similar information about voice quality in SLZ.

Figure 4.4 shows the first dimension plotted against the fourth dimension. This plot is very similar to Figure 4.2, with the only difference being that the fourth dimension moves the modal voice from the bottom-center of the plot to almost the exact center of the plot.

Figure 4.5 shows the second dimension plotted against the third dimension. This

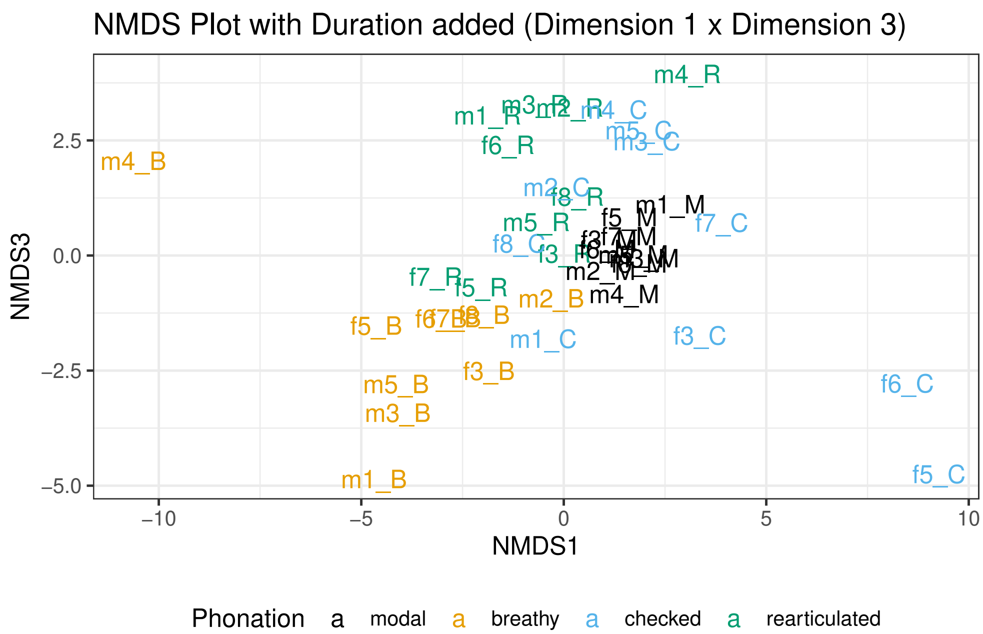


Figure 4.3: Two-dimensional MDS solution showing the first and third dimensions.

plot is essentially the same as Figure 4.3, except that the coordinates are flipped.

Figure 4.6 shows the second dimension plotted against the fourth dimension. This plot shows that the modal voice and the non-modal voice are separated into two different clusters, with modal voice located to the extreme left of the plot and the nonmodal voice qualities located to the right of the modal grouping. Again, as first seen in Figure 4.4, the fourth dimension centralizes the modal voice, but no discernible pattern is observed for the other phonations.

Figure 4.7 shows the third dimension plotted against the fourth dimension. This plot is very similar to Figure 4.4 with the exception that along the third dimension checked and rearticulated voice have swapped places. Checked voice is more central-

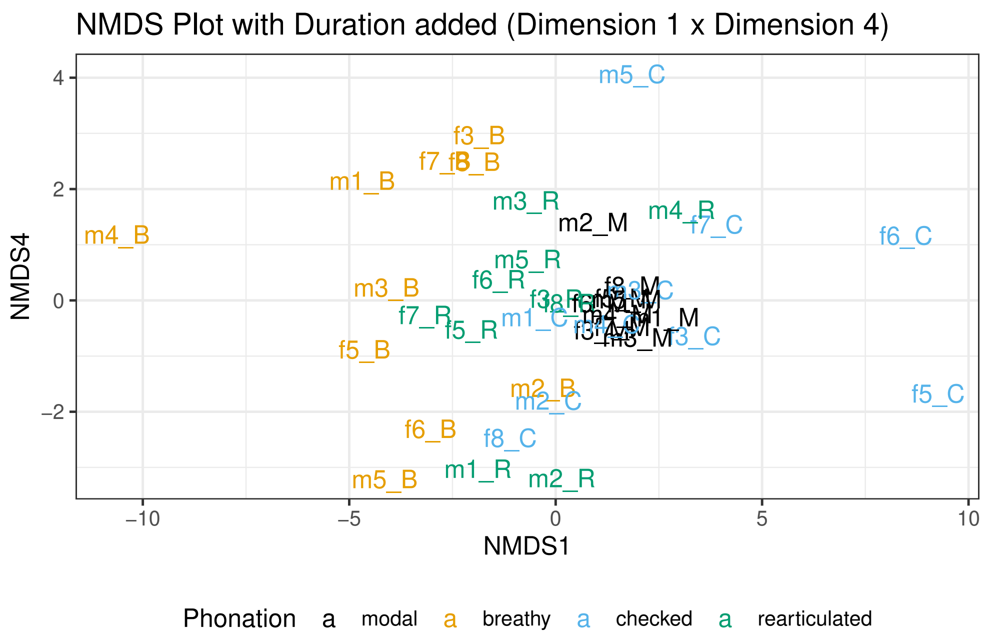


Figure 4.4: Two-dimensional MDS solution showing the first and fourth dimensions.

ized in the plot, whereas rearticulated voice is located more to the right of the plot. Again, we observe that the modal voice is located in the center of the plot.

4.3.1.1 Interim summary for dimension plots

The plots of the MDS analysis show that the acoustic space of voice quality in SLZ is primarily represented by a three-dimensional space. The first dimension and third dimensions are very similar in that they capture a continuum from breathy, to modal, and finally creaky voice. This is similar to what Keating et al. (2023) found in their study for the second dimension.

This continuum from breathy to modal to creaky is very similar to the open-

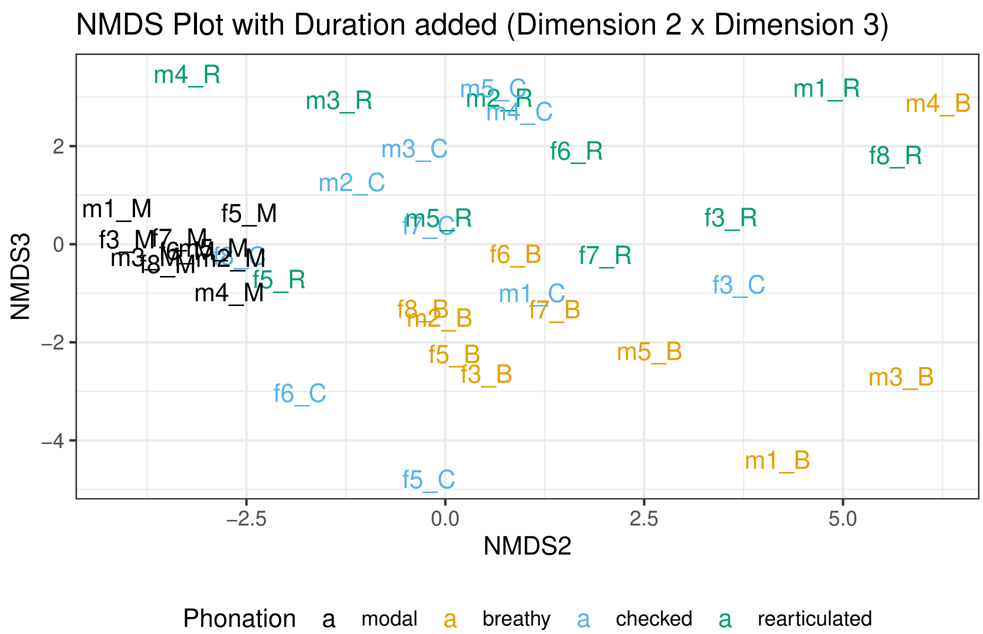


Figure 4.5: Two-dimensional MDS solution showing the second and third dimensions.

quotient model of voice quality proposed by Gordon & Ladefoged (2001), as illustrated in Figure 4.8. Because of this similarity to the open-quotient model, the first and third dimensions are likely capturing the open-quotient of the glottis.

From the plots involving the second dimension, we see that this dimension separates the modals from the non-modal ones. This is similar to what we observe with the various harmonics-to-noise ratios and the cepstral peak prominence (CPP) which are all measures of the amount of noise present in the signal across various bandwidths (de Krom 1993, Hillenbrand & Houde 1996a, Blankenship 2002, Ferrer Riesgo & Nöth 2020). In addition to the amount of noise, it also captures the strength of the vocal fold vibration, similar to what Garellek et al. (2021) found in their study of glottal

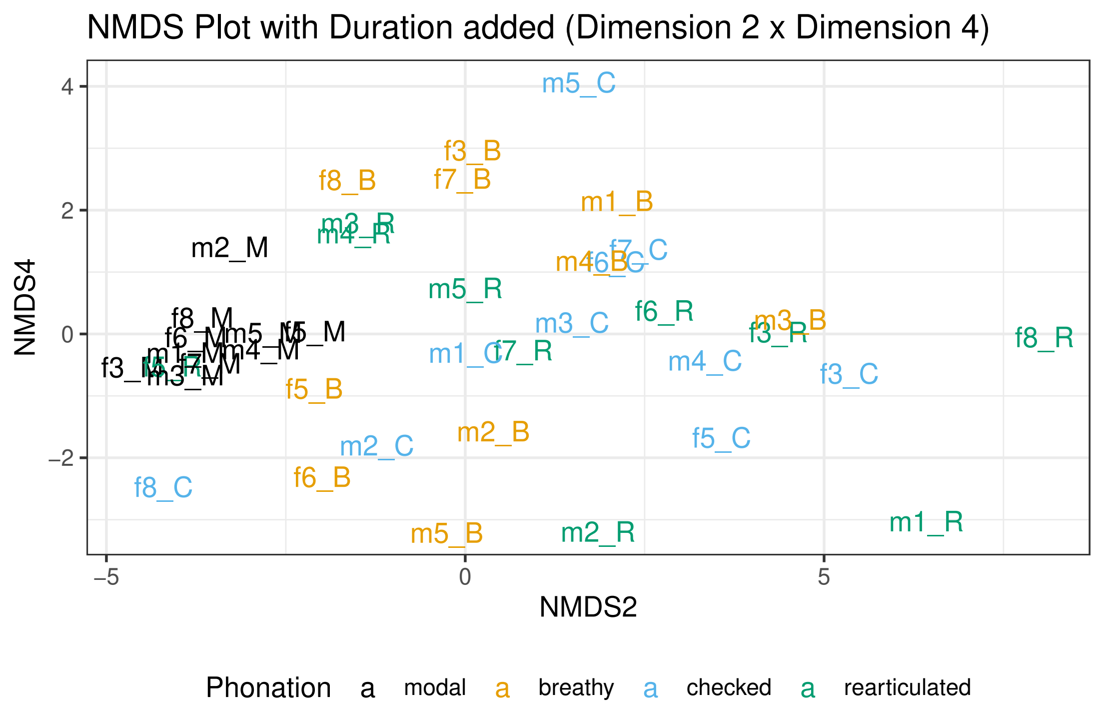


Figure 4.6: Two-dimensional MDS solution showing the second and fourth dimensions.

consonants and nonmodals. In their study, they found that the modal voice had the highest strength of vocal fold vibration (measured by the Strength of Excitation), while the non-modal voice had a lower strength of vocal fold vibration. This suggests that the second dimension is to capture the amount of aperiodic noise in the signal, the strength of the vocal fold vibration, or both.

The fourth dimension is less clear about what it is potentially capturing. In all of the plots involving the fourth dimension, we see that modal voice is always located near the center of the plot, while the nonmodal voice qualities are located around this

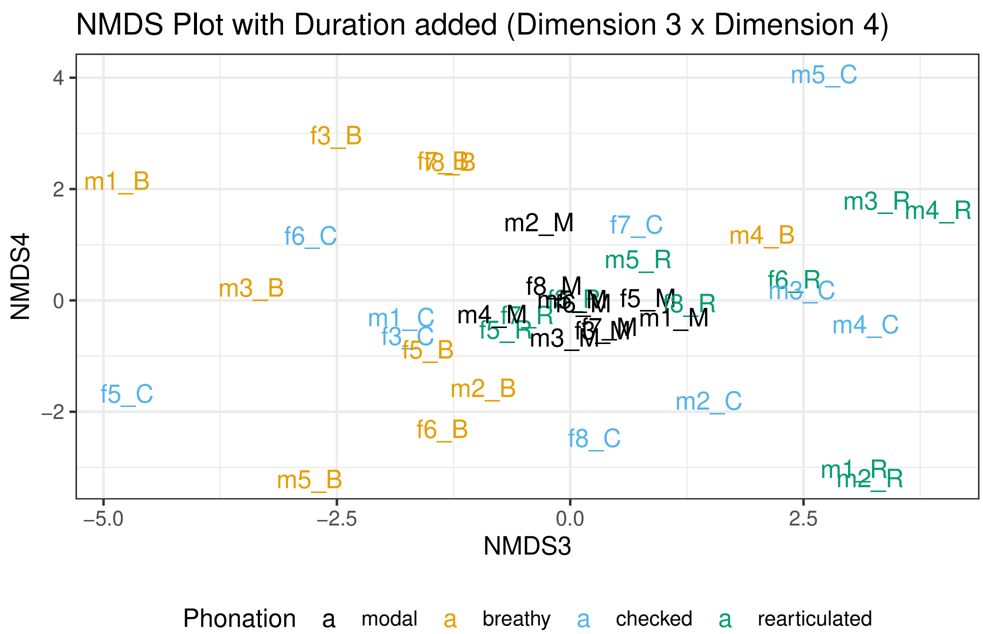


Figure 4.7: Two-dimensional MDS solution showing the third and fourth dimensions.

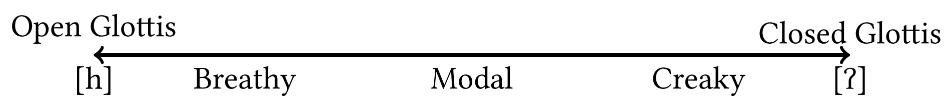


Figure 4.8: A diagram showing the relationship between breathy, modal, and creaky phonation types from Gordon & Ladefoged (2001).

point depending on the patterns from the other dimensions. This suggests that the fourth dimension possibly captures something about modal voice.

The rest of this chapter will focus on how the different acoustic measures contribute to the different dimensions of the MDS analysis. This will be followed by a general discussion of both the MDS analysis dimensions and the acoustic measures that are correlated with these dimensions. This discussion will focus on how the results of this study relate to previous work on voice quality and the implications of these results for our understanding of the acoustic space of voice quality.

4.3.2 Acoustic correlates of voice quality

Looking at the visual representation of the dimensions is only part of the full story. In order to fully understand what is occurring, we need to determine how the acoustic measures contribute to each of the different dimensions. There are two ways to do this: (i) looking at the amount of weight each acoustic measure contributes to the different dimensions, and (ii) looking at how correlated the different acoustic measures are to the different dimensions (Kruskal & Wish 1978, Hastie, Tibshirani & Friedman 2009). Both of these methods are useful in understanding the shape of the data. In the following discussion, I will take the second approach and look at the correlations between the different acoustic measures and the different dimensions. This is done to make the resulting discussion easier to follow.

Table 4.1 shows the correlation score, computed by the `cor` function in R, for each acoustic measure and dimension. The four largest correlations in each dimension are in bold. The choice to bold the four largest correlations was arbitrary and was done to make it easier to discuss. Although only the four largest correlations are bolded, there are several correlations that share similar correlation values. These cases will be discussed as needed.

Table 4.1: Correlations for each acoustic measure to the four dimensions (NMDS1, NMDS2, NMDS3, NMDS4). The four largest correlations in each dimension are bolded.

Acoustic Measure	NMDS1	NMDS2	NMDS3	NMDS4
H1*−H2*	-0.221	-0.339	0.031	0.314
H2*−H4	-0.437	0.239	**-0.689**	**-0.364**
H1*−A1*	**-0.828**	0.048	**-0.459**	0.044
H1*−A2*	**-0.855**	-0.067	-0.343	0.114
H1*−A3*	**-0.809**	-0.218	-0.297	0.126
H4*−H2k*	-0.452	-0.598	0.294	**0.366**
H2k*−H5k*	0.152	0.023	0.101	0.057
residual H1*	-0.290	-0.443	**-0.722**	0.084
H2*	-0.157	-0.555	**-0.679**	0.114
H4*	0.295	**-0.778**	0.078	**0.479**
A1*	0.756	-0.549	0.092	0.124
A2*	**0.779**	-0.476	-0.103	0.086
A3*	0.735	-0.416	-0.211	0.093
CPP	-0.590	-0.606	0.209	-0.179
HNR < 500 Hz	-0.513	**-0.792**	0.152	-0.202
HNR < 1500 Hz	-0.275	**-0.799**	0.323	-0.290
HNR < 2500 Hz	-0.327	-0.714	0.391	-0.348
HNR < 3500 Hz	-0.446	-0.644	0.393	-0.356
Strength of Excitation	-0.013	-0.741	-0.238	0.145
SHR	0.144	-0.176	0.122	**-0.597**
Energy	-0.080	**-0.793**	-0.015	0.341
Duration	-0.622	0.539	0.257	0.030

4.3.2.1 Dimension 1

From Table 4.1, we observe that the first dimension is negatively correlated with the acoustic slope measures of H1*−A1*, H1*−A2*, and H1*−A3* ($r^2 \approx -0.825$ for all three of these measures). These measures are all types of spectral slope measures which attempt to normalize the amplitude of the formant corrected H1 against the amplitudes of the formant correct harmonic closest to the first three formants. It has been noted that these measures are correlated with the open-quotient of the glottis (see Garellek 2019, 2022 for an overview of these measures). Because these measures all capture the same information and they are all highly correlated with the first dimension, we can conclude that the first dimension is primarily concerned with the spectral slope of the signal (i.e., the open-quotient of the glottis).

The first dimension is also positively correlated with the amplitude of the first three formants (that is, A1*, A2*, and A3*). These acoustic measures also share correlations of a similar value ($r^2 \approx 0.75$). It is also worth noting that these amplitude measures are all used in the normalization of the spectral slope measures (i.e., H1*−A1*, H1*−A2*, and H1*−A3*). I believe that this fact suggests that the first dimension is primarily concerned with the spectral slope of the signal, but is also factoring in the amplitude of the formants.

4.3.2.2 Dimension 2

The second dimension is strongly correlated with the harmonics-to-noise ratio (HNR) measures. The HNR measures HNR < 500 Hz, HNR < 1500 Hz, and HNR < 2500 Hz are all negatively correlated with the second dimension ($r^2 \approx -0.79$). These measures are all measures of the amount of noise in the signal and are used to indicate the amount of periodicity in the signal (e.g., whether something is modal or non-modal). This suggests that the second dimension is concerned with the amount of noise in the signal.

Furthermore, there are strong negative correlations with energy, which is the root mean squared energy of the acoustic signal, and the Strength of Excitation (SoE), which is defined as “the instant of significant excitation of the vocal-tract system during production of speech” and represents “the relative amplitude of impulse-like excitation” (Mittal, Yegnanarayana & Bhaskararao 2014: p. 1934). In other words, the SoE correlates with how strongly the vocal folds vibrate during the signal with modal voice showing the strongest amount of voicing and nonmodal phonation the least. This suggests that the second dimension is also concerned with strength of the vocal-fold vibration.

The last measure that shows a strong correlation with the second dimension is the amplitude of the fourth harmonic (H4*). This measure is negatively correlated with the second dimension ($r^2 \approx -0.78$). This measure is typically used to help normalize the amplitude of the first formant and is used in the calculation of the spectral slope measures (e.g., H2*−H4*).

Together, given the correlations of the HNR measures, energy, and SoE, we can conclude that the second dimension is primarily concerned with the periodicity of the signal (i.e., the amount of noise in the signal) and the strength of the vocal fold vibration.

4.3.2.3 Dimension 3

The third dimension is negatively correlated with the two spectral slope measures of residual H1* and H2*−H4* ($r^2 \approx -0.7$). Residual H1* is a measure that has been shown to be robust in capturing the strength of the first harmonic (Chai & Garellek 2022, Brinkerhoff & McGuire 2025), which was the original goal of using higher harmonics to normalize H1 (Fischer-Jørgensen 1968). H2*−H4* has also been shown to be another spectral slope measure that can capture the differences in amplitude between different phonation types (Garellek 2019, Garellek et al. 2016, Kreiman et al. 2014, 2021). This suggests that the third dimension is primarily concerned with the spectral slope of the signal, similar to what we observed in the first dimension.

In addition to these measures, the amplitude of the second harmonic (H2*) was also found to be negatively correlated with the third dimension ($r^2 \approx -0.69$). This measure is typically used to help normalize the amplitude of the first harmonic and is used in the calculation of the spectral slope measures (e.g., H1*−H2*, H2*−H4*, etc.). Due to its use in spectral slope measures and that H2*−H4* is also negatively correlated with the third dimension, we can conclude that H2 * also contributes to the spectral slope of the signal. This further supports the idea that the third dimension is primarily concerned with the spectral slope of the signal.

The last acoustic measure also correlated with the third dimension is the spectral slope measure of H1*−A1* ($r^2 \approx -0.46$). Even though the correlation is not as strong as the other acoustic measures, it still shows that the third dimension corresponds to the spectral slope of the signal (i.e., the open-quotient of the glottis).

4.3.2.4 Dimension 4

In the fourth dimension, we observe that the correlations are less clear than in the previous three dimensions. The strongest correlations in this dimension with the Subharmonic-to-Harmonic ratio (SHR; $r^2 \approx -0.60$) describe the relative strength of any subharmonics (interharmonics) in the signal (Sun 2002). The subharmonics in the signal correspond to alternating periods in the time domain (that is, period doubling), which typically occurs in creaky voice and with broader laryngeal constrictions (see Herbst 2021 for concerns about this acoustic measure).

The positive correlation with H4* ($r^2 \approx 0.48$) suggests that the fourth dimension may also capture some information about the amplitude of the higher harmonics. This is also true with the positive correlation with H4*−H2k* ($r^2 \approx 0.37$) and H2*−H4* ($r^2 \approx 0.36$), which are both measures that capture the spectral slope of the signal. This suggests that the fourth dimension may capture information about the amplitude of the harmonics in the signal.

4.3.2.5 Interim summary for acoustic correlates

Based on the correlations observed in Table 4.1, we can summarize the acoustic correlates of each dimension as follows: Dimension 1 captures the spectral slope of the signal, primarily through the spectral slope measures H1*−A1*, H1*−A2*, and H1*−A3*. This dimension appears to be related to the open-quotient of the glottis. Dimension 2 captures the amount of noise in the signal and the strength of vocal fold vibration, primarily through the HNR measures (HNR < 500 Hz, HNR < 1500 Hz, and HNR < 2500 Hz), Energy, and Strength of Excitation. This dimension separates modal from nonmodal voice quality, which is what periodicity is concerned with. Dimension 3 captures the spectral slope of the signal as well, just as Dimension 1 does, primarily through residual H1*, H2*−H4*, and H2*. This dimension also appears to be related to the open-quotient of the glottis. Finally, Dimension 4 captures the periodicity in the signal through SHR and also appears to capture some information about the spectral slope, as evidenced by the amplitude of the higher harmonics through H4* and H4*−H2k*. However, it is not entirely clear whether this dimension is primarily concerned with the spectral slope of the signal or the periodicity in the signal.

4.4 Discussion

The results of this study show that the voice quality in SLZ occupies an acoustic space, but also shows that this space is similar to what Keating et al. (2023) found in their study of phonation in 11 languages. However, the results of this study differ from Keating et al. (2023). Instead of a two-dimensional space like in Keating et al. (2023), SLZ's voice quality occupies a three-dimensional space. Despite these differences, the behavior of the dimensions is similar to that in Keating et al. (2023).

In the analysis presented in this chapter, we see that the first and third dimensions are primarily concerned with a spectral slope continuum from positive spectral slope to negative spectral slope which correlates to breathy voice to modal voice and finally to creaky voice. As extensive research has shown, the spectral slope of the signal is closely related to the open-quotient of the glottis, or in other words, how open or closed the glottis is during phonation (see Garellek (2022) for an overview of this history). This continuum was also found to exist in the second dimension of Keating et al.'s 2023 MDS analysis. Based on Keating et al. (2023) and the analysis presented in this chapter, at least one of the dimensions of any acoustic space related to voice quality must correspond to the spectral slope of the signal.

As mentioned in Section 4.3.1.1, the first and third dimensions of this chapter's analysis and Keating et al.'s second dimension, bear a striking resemblance to the voice quality model proposed by Gordon & Ladefoged (2001). In Gordon & Ladefoged's model, voice quality is described as being a single continuum based on how open or closed the glottis is during speech. The more open the glottis, the more breathy the phonation will be. The more closed the glottis, the more creaky the phonation will be. This model also claims that laryngeal consonants [h] and [ʔ] also exist along this continuum and represent the extreme ends of this continuum. This model can be visually represented as a line with [h] on one end and [ʔ] on the other end. The various voice qualities exist between these two extremes, as represented in Figure 4.9.

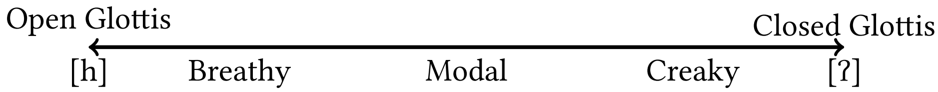


Figure 4.9: A diagram showing the relationship between breathy, modal, and creaky phonation types. Based on Gordon & Ladefoged (2001).

As mentioned above, the measures that correlate the most with the first and third dimensions are several spectral slope measures (i.e., H1*−A1*, H1*−A2*, H1*−A3*, residual H1*, and H2*-H4*). These measure correlations provide further support that the first and third dimensions of the MDS analysis are concerned with the spectral slope of the signal, which in turn is closely correlated with the state of the glottis during phonation (Holmberg et al. 1995, Kreiman, Gerratt & Antoñanzas-Barroso 2007, Garellek et al. 2016, Garellek 2019, Chai & Garellek 2022).

The second dimension of this chapter's analysis is concerned with dividing the acoustic space into modal and nonmodal voice qualities. This split between modal and nonmodal is similar to the split that exists for periodicity and strength of voicing. The more modal the voice quality is, the greater the amount of periodicity or voicing that we observe in the signal. I propose that this dimension is concerned with how harmonic the signal is, or in other words, how much noise is present in the acoustic signal. This is similar to what Keating et al. (2023) found in their study, where the first dimension of their analysis was primarily concerned with making the same split in the acoustic space. Based on these results, I propose that another dimension in the acoustic space of voice quality must correspond to the amount of noise in the signal. This proposal is further supported by the fact that the second dimension of this chapter's analysis is primarily correlated with the harmonics-to-noise ratios, energy, and SoE measures, which are all measures of the amount of noise or the amount of energy present in the acoustic signal.

The last dimension of the MDS analysis is less clear in what it is capturing. The correlation with the subharmonics-to-harmonic ratio (SHR) suggests that this dimension might be capturing whether or not their is period doubling which is a common feature in creaky voice. However, there is no easily describable pattern that emerges with this measure.

The results of this study show that the voice quality acoustic space in SLZ is primarily represented by a three-dimensional space. However, it can primarily be broken down into two aspects that the acoustic space is attempting to capture: (i) the spectral slope of the signal and (ii) the amount of noise in the signal. This can be represented as a primarily two-dimensional space with higher dimensions adding additional information related to these two primary concerns. This can be visualized as a two-dimensional space with the first dimension representing the spectral slope of the signal and the second dimension representing the amount of noise in the signal, as shown in Figure 4.10.

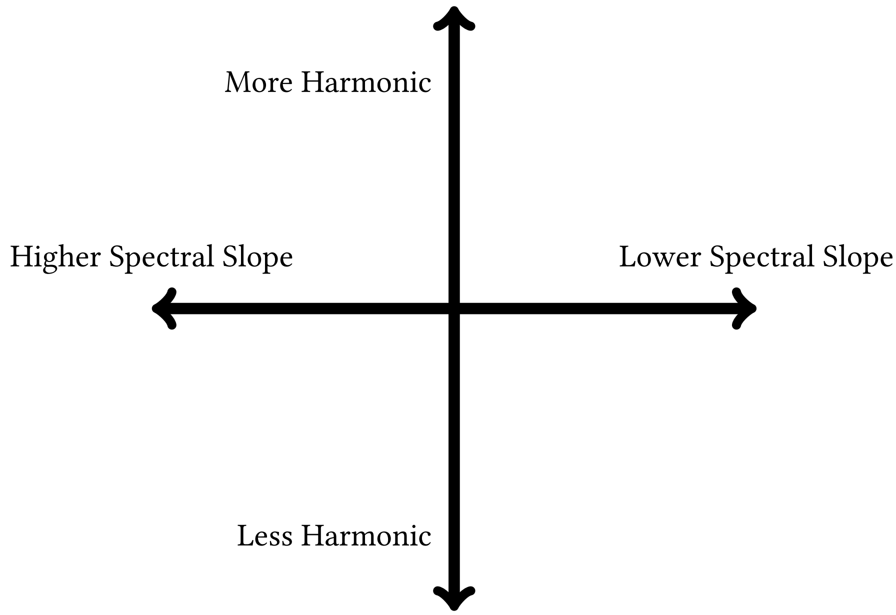


Figure 4.10: A two-dimensional representation of the acoustic space of voice quality in SLZ. The horizontal axis represents the spectral slope of the signal, while the vertical axis represents the amount of noise or energy in the signal.

4.5 Conclusion

Although the discussion has predominately been about the correlations of the measures that contribute to the different dimensions, it is important to note that the measures are not independent of each other. Instead, all of the measures contribute to the acoustic space of voice quality in SLZ to some extent or another. Just because a measure has a low correlation does not mean that it does not contribute to the acoustic space. Rather than thinking of the measures as independent of each other, it is better to think of them as a group of measures that work together to create the acoustic space of voice quality in SLZ. This is especially true given the fact that the MDS analysis is a reduction of the data to a few dimensions. This analysis offers a snapshot of the voice quality acoustic space in SLZ, but is not the full picture.

Furthermore, as will be discussed in Chapter 5, another way in which we can determine which measures are the most important is by performing a classification and regression tree analysis (Breiman et al. 1986, Breiman 1996, 2001).

Chapter 5

Trees reveal the importance of measures in SLZ

5.1 Introduction

The MDS analysis presented in Chapter 4 provides us an understanding of the acoustic shape that voice quality takes in SLZ. This shape is defined by different dimensions that correspond to the glottal configuration needed to produce each voice quality and the amount of noise in the signal. The MDS analysis also provides us with a way to visualize how the different voice qualities occupy an acoustic space. Finally, the chapter discussed how different acoustic measures are correlated with different dimensions of the acoustic space. This provides us with a potential avenue to explore which measures contribute to our understanding of the different voice qualities in SLZ. However, the analysis does not tell us which measures are the most important in making the splits between the different voice qualities. This is where decision trees are helpful as they provide a way to cut through the noise and reveal which measures play the most important role in dividing that acoustic space.

In this chapter, I present an analysis using a flavor of decision trees called Random Forests (Breiman et al. 1986, Breiman 2001). Random Forests are a type of ensemble learning method that combines multiple decision trees that are built on the same data. This allows us to take advantage of the strengths of decision trees while minimizing the weaknesses that are present in classical decision trees (see Hastie, Tibshirani & Friedman 2009, Boehmke & Greenwell 2019, James et al. 2021 for a discussion of the strengths and weaknesses of the different types of decision trees).

In this chapter, I show that only a small number of acoustic measures are needed to classify the different voice qualities and make the splits in the acoustic space.

5.2 What are Decision Trees

Decision trees are a statistical tool that helps to reveal which predictors divide the space under investigation. Essentially, this is done by stratifying or segmenting the predictor space into some number of simpler regions. The rules that divide the space into these regions are based on some aspect of the predictors (see Hastie, Tibshirani &

Friedman 2009, James et al. 2021 for explanations of the statistics and how to perform these analyzes in R).

These trees can be used for both regression and classification. In the case of regressions, it splits the predictor space into regions and calculates how the item under discussion behaves in each region. This process of splitting into regions and calculating how something responds in that region continues until some stopping rule is applied, which is usually defined to some number of terminal nodes. This resulting tree is rather large and is then pruned based on the cost-complexity pruning to a subset of itself. This subsetted tree is the tree that has minimized its cost-complexity criterion for all potential subsets. That is, it balances the trade-off between the complexity of the tree and its fit to the data.

In the case of classification, the algorithms that result in a tree are very similar to those used for regression trees. The main difference in the algorithm comes from what is used to split the nodes and how the tree is pruned. Instead of predicting a continuous outcome like with regression trees, classification trees predict a categorical outcome. The predictor space is divided into regions and within each region, the majority class is assigned as the predicted class for that region. This process continues until a stopping rule is applied, similar to regression trees. The resulting tree can also be pruned to avoid overfitting, using a cost-complexity criterion.

Decision trees are easy to interpret and visualize, making them an ideal choice for understanding the structure of data and how the different predictors interact with data (Hastie, Tibshirani & Friedman 2009, James et al. 2021).

5.3 Decision trees in linguistics

The use of decision trees in linguistics is not new. One of the first uses was done by Tagliamonte & Baayen (2012), where they illustrated the use of decision trees in investigating which sociolinguistic factors were the most important in the use of *was* versus *were* in York English.

Recently, decision trees were used to show which acoustic measures were important in making the split in the acoustic space for voice quality (Keating et al. 2023). Keating et al. performed a simple decision tree analysis to supplement their MDS analysis of voice quality in 11 languages. The results of this analysis are shown in Figure 5.1. Decision trees like the one in Figure 5.1, show the binary splits that are made in the space and what predictor, and the value of that predictor, makes that split. In the case of Keating et al.'s (2023) tree, the first split is made on the harmonics-to-noise ratio over the frequency range from 0 to 500 Hz for the middle third of each vowel. This split is made at the z-score of 0.48. If the predictor value is greater than or equal to 0.48, the dominant voice quality of that region is modal. If HNR < 500 Hz value is less than 0.48, the region needs to be further split.

The next split in the region is made on the subharmonic-to-harmonic ratio for the

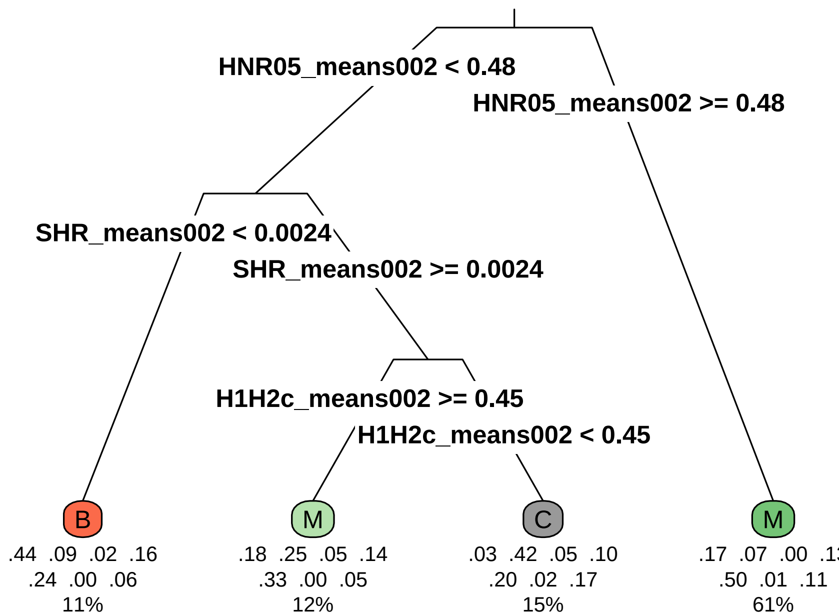


Figure 5.1: Classification tree of phonation categories from Keating et al. (2023). Abbreviations used in this figure are: HNR05_means002: harmonics-to-noise ratio over the frequency range from 0 Hz to 500 Hz for the middle third of each vowel; SHR_means002: subharmonic-to-harmonic ratio for the middle third of each vowel; H1H2c_means002: H1* – H2* for the middle third of each vowel; B: breathy, M: modal, and C: creaky phonation categories.

middle third of each vowel. If the value of this predictor is less than 0.0024, the voice quality is classified as breathy. If the value is greater than or equal to 0.0024, the region needs to be split further.

The final split in the region is made on the H1*−H2* for the middle third of each vowel. If the value of this predictor is less than 0.45, the voice quality is classified as creaky. If the value is greater than or equal to 0.45, the voice quality is classified as modal. This tree shows that using only three acoustic measures, one can classify the voice quality of the data. This is a powerful tool for understanding the importance of different acoustic measures in the acoustic space.

5.4 Growing a forest of decision trees

However, simple decision trees suffer from two main disadvantages. The first is that decision trees can suffer from high variance. In other words, the tree can be very sensitive to small changes in the data on which it was trained. The second disadvantage is that decision trees do not have the same predictive accuracy as other regression or classification models (see Hastie, Tibshirani & Friedman 2009 for a discussion). The following subsections explain two methods that have been proposed to overcome these disadvantages. The first method is called Bagging trees and the second method is called Random Forests. As will be explained in the following sections, these two methods are essentially the same except for one key difference, the number of predictors that are used. Bagging trees use all predictors in each split, whereas Random Forests use a random subset of predictors in each split.

5.4.1 Bagging trees

One way to overcome the disadvantages of simple decision trees is to make use of a technique called bootstrap aggregating, or bagging (Breiman 1996). This means that instead of growing a single tree with all of the predictors (i.e., variables in your data) in the data like in simple decision trees, we grow many trees on random samples of the data until we reach a given number of trees using the full set of predictors. Once these trees are grown, we then average across the trees to get a more stable prediction of how the regions are split and what predictors are most important in making those splits. This averaging across the trees helps to explain the variance in the data and improve the predictive accuracy of the model. However, this comes at the cost of interpretability.

In decision trees, we usually represent the splits in the data as a tree. When we use bagging, because of the large number of trees that are grown, it is impossible to represent the results in this way. Instead of using a tree, we use variable importance measures to understand which predictors are most important in dividing the data. There are two measures that are commonly used to understand the importance of variables in tree bagging: the residual sum of squares (RSS) for regression trees and the Gini index, which is a measure of *purity*, for classification. In regression trees, the amount of RSS that decreases due to the splits over a given predictor is recorded and averaged across all the trees. In the classification trees, the total amount that the Gini index decreases for each predictor is recorded and averaged across all trees. The higher the value of the RSS, or Gini index, the more important that predictor is in making the splits in the data. These are then graphed to show the importance of each predictor in the data with the most important predictors at the top of the graph.

The exact number of trees to be grown is not known *a priori*. Instead, the number of trees is determined by the user and is usually determined by the number of trees needed to stabilize the prediction. This is done by comparing multiple models that were built with different numbers of trees and determining which number of trees produces the most stable prediction. This is done by comparing the predictions of the different models and calculating the variance of the predictions across the different models. The model that produces the most stable prediction is the one chosen.

5.4.2 Random Forests

However, subsequent research by Breiman (2001), showed that bagging sometimes would not stabilize and that they would overfit the data. To remedy this problem, Breiman proposed that instead of growing multiple trees that consider all predictors every time, we should only consider a random subset of predictors in each split. This means that bagging and Random Forests are essentially the exact same thing except for the number of predictors that are considered at each split. This is a very important distinction, as it allows us to grow trees that are more stable and less likely to overfit the data.

Research has shown that generally speaking the number of predictors that must be considered in each split (that is, m_{try}) is usually the square root of the total number of predictors in the data for classification trees and one-third of the total number of predictors for regression trees (Breiman 2001, Sandri & and Zuccolotto 2008, Hastie, Tibshirani & Friedman 2009, Janitza & Hornung 2018, Boehmke & Greenwell 2019, James et al. 2021). This means that if we have 81 predictors in our data, we would only consider ≈ 9 predictors at each split for classification trees ($\sqrt{81} = 9$) and ≈ 27 predictors for regression trees ($\frac{81}{3} = 27$). However, this is not a hard and fast rule, and the number of predictors that are considered at each split can be tuned to improve the performance of the model. This is done by comparing the predictions of the different models and calculating the variance of the predictions across the different models. The model that produces the most stable prediction is the one chosen.

5.4.3 How to interpret the results

The benefit of using Bagging and Random Forests is that they are able to produce more stable predictions than simple decision trees. This is because they are able to average across the predictions of multiple trees instead of growing a single tree. This allows us to take advantage of the strengths of decision trees while minimizing the weaknesses that are present in classical decision trees.

However, the benefits of using these methods come at the cost of interpretability. In classical decision trees, the results of the analysis are presented as a tree plot, similar to the tree in Figure 5.1. This allows us to see how the different predictors are used to split the data. This is not possible in bagging and random forests due to the large number of trees that are grown and then used to average the predictions. Instead, we used variable importance plots to show how important a variable is across all trees in making the data splits. This is done by calculating a measure of importance for each variable and then plotting the results. In the case of random forests, the most common measure of importance is called the Gini Index, which is a measure of how pure the split is generally. The Gini index is calculated for each variable and then averaged across all trees. The higher the Gini index, the more important that variable is in making the splits in the data. This is similar to how we interpret the results of a classical decision tree, where the most important predictors are at the top of the tree and the least important predictors are at the bottom. In bagging and random forests, we are not concerned with the exact value of the Gini index, but rather with the mean decrease in the Gini index. This is because the Gini index is a measure of how pure the split is and not a measure of how important the variable is in making the splits in the data. The mean decrease in the Gini index is a measure of how much the variable contributes to the overall purity of the split. For bagging, this is the only measure that is considered.

There is some evidence that the Gini index is not always the best measure for random forests. For example, Strobl et al. (2007) showed that the Gini index can be biased in some cases with random forests. To get around this fact, interpretation of random forests also frequently consult the permutation importance in addition to the Gini Index. The permutation importance measures the change in the model's prediction accuracy when the values of a variable are randomly permuted. This means that the variable is shuffled and the model is re-evaluated. The difference in accuracy between the original model and the permuted model is then used to measure the importance of that variable. The higher the difference, the more important that variable is in splitting the data.

Tagliamonte & Baayen (2012) made use of both the Gini index and the importance of permutation to evaluate the importance of the different sociolinguistic predictors in their analysis. The results of their analysis are shown in Figure 5.2. The Gini index is shown on the left, and the permutation importance is shown on the right. The y-axis shows the different predictors for both plots. We observe in this graph that for the impurity importance (i.e., Gini index), the most important predictors are those that are at the top of the graph. The most important predictor is the one that has the largest

Gini index, in this case the Individual variable. The second most important predictor is the Age variable, etc. The permutation importance is shown on the right and is interpreted in the same way. However, they kept the ordering of the predictors the same across the two plots. The most important predictor according to the permutation importance is the one that has the largest value in this case, the Individual variable. The second most important predictor is the Proximate1 variable, etc.

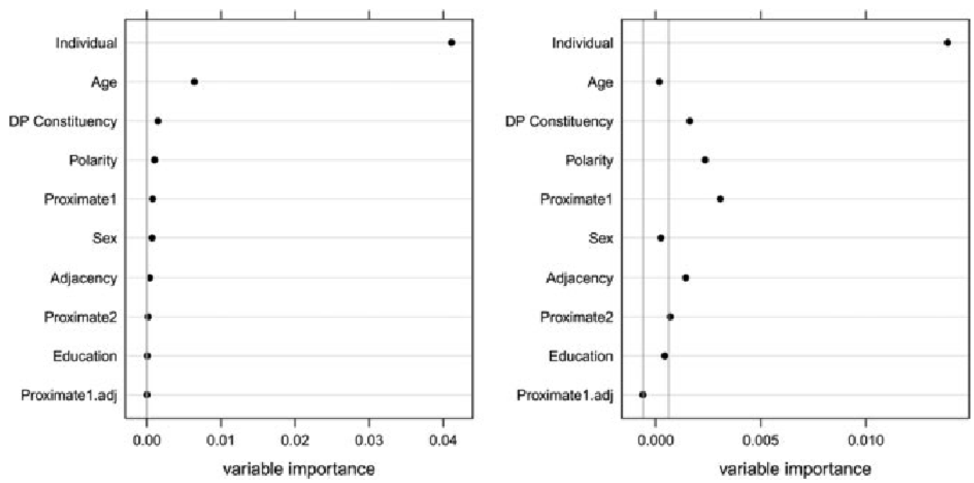


Figure 5.2: Variable importance plot from Tagliamonte & Baayen (2012). The plot on the left shows the impurity importance (i.e., Gini index) and the plot on the right shows the permutation importance. The y-axis shows the different predictors for both plots.

Looking at both the impurity and permutation importance, we can get a better understanding of which predictors are the most important in making the splits in the space. For Tagliamonte & Baayen (2012), this meant that the Individual variable was the most important predictor in making the splits in the data. The other predictors were also important, but we now see differences between the two plots in which are important. In general, the researcher should look at both the impurity importance and the permutation importance to gain an understanding of which predictors are the most important in making the splits in the data. If there is a overlap between the two plots, then we can be more confident that the predictor is important in making the splits in the data. If there is a large difference between the importance of a predictor between the two plots, then we should be cautious about that predictor's importance.

5.5 Random Forests in SLZ

In this section, I will present the results of a Random Forest analysis that was performed on the SLZ data. The goal of this analysis is to understand which acoustic measures are the most important in achieving the separations in the acoustic space of the SLZ. The data collection and analysis methods are similar to those used in the MDS analysis presented in Chapter 4. The main difference is that I will be using a Random Forests analysis instead of an MDS analysis. The results of this analysis will be presented in the following sections.

5.5.1 Methods

5.5.1.1 Participants

This study uses data collected from 10 native speakers of SLZ during the summer of 2022. Participants were recruited from the community of Santiago Laxopa, Oaxaca, Mexico. All participants were native speakers of SLZ. The participants were between 18 and 60 years old and consisted of five males and five females.

5.5.1.2 Recordings

Participants were asked to perform a word list elicitation task consisting of 72 words. These words were selected to elicit the entire range of types of voice quality in SLZ, including modal voice, the two kinds of creaky (i.e., checked and rearticulated), and breathy voice. The words were selected based on previous research conducted as part of the Zapotec Language Project at the University of California, Santa Cruz (*Zapotec Language Project — University of California, Santa Cruz* 2022). Because participants were not literate in SLZ, the word list was prompted for them by asking them "How do you say [word in Spanish]?" by myself and another researcher in Zapotec. Participants were asked to respond with the desired word in the carrier phrase *Shnia' X chonhe lhas* [ʃnːiaˀ X tʃone ɾas] "I say X three times." which was repeated three times. These utterances were recorded in a quiet environment using a Zoom H4n handheld digital recorder. The recordings were saved as 16-bit WAV files with a sampling rate of 44.1 kHz.

5.5.1.3 Acoustic measuring

The resulting audio files were then processed in Praat to isolate the vowel portion of each word. The onset of the vowel was set to the second glottal pulse after the onset, and the offset of the vowel was set to the last glottal pulse before the decrease in amplitude at the end of the vowel (Garellek 2020). The vowel was then extracted and saved as a separate file for analysis.

These vowels were fed into VoiceSauce (Shue et al. 2011) to generate the acoustic measures for the studies discussed in this dissertation. Because many acoustic measures are based on the fundamental frequency, this measure was calculated using the STRAIGHT algorithm from (Kawahara, Cheveigne & Patterson 1998) to estimate the fundamental frequency in millisecond (ms) intervals. Once the fundamental frequency is calculated, VoiceSauce then uses an optimization function to locate the harmonics of the spectrum, finding their amplitudes.

VoiceSauce then uses the Snack Sound toolkit (Sjölander 2004) to find the frequencies and bandwidths of the first four formants, also at millisecond intervals. The amplitudes of the harmonics closest to these formant frequencies are located and treated as the amplitudes of the formants. These formant frequencies and bandwidths are used to correct the harmonic amplitudes for the filtering effects of the vocal tract, using Iseli, Shue & Alwan's 2007 extension of the method employed by Hanson (1997). Each vowel was measured across ten equal time intervals, resulting in 22890 data points in total. These measures were then z-scored by speaker to reduce the variation between speakers and provide a way to compare the different measures directly on similar scales.

5.5.1.4 Data processing

Data points with an absolute z-score value greater than three were considered outliers and excluded from the analysis. The Mahalanobis distance was calculated in the F1-F2 panel within each vowel category. Each data point with a Mahalanobis distance greater than six was considered an outlier and excluded from the analysis. Using the Mahalanobis distance allows us to compare the data points to the mean of the F1-F2 panel for each vowel category. The larger the Mahalanobis distance is the more deviant the data point is from the mean which in turn means that the data point was improperly tracked. This is comparable to what was done in Seyfarth & Garellek (2018), Chai & Ye (2022), and Garellek & Esposito (2023).

Energy was excluded if it had a zero value and then log-transformed to normalize its right-skewed distribution. Afterward, the resulting log-transformed Energy was z-scored, and any data point with a z-score greater than three was excluded. This outlier removal resulted in 1918 data points being removed.

All data points were then z-scored by speaker to reduce the variation between speakers and provide a way to compare the different measures directly on the same scale.

Residual H1* for the remaining data points following Chai & Garellek (2022). First, a linear mixed effects model was generated with the z-scored H1* as the response variable and the z-scored energy as the fixed effect. The uncorrelated interaction of the z-scored energy by speaker was treated as random. The energy factor resulting from this linear mixed-effects model was extracted. Finally, the z-scored H1* had the product of the z-scored energy and energy factor subtracted from it to produce the residual H1* measure.

Once these steps were completed, the mean of each combination of phonation and speaker was taken for the fourth to seventh interval of the vowel. This is similar to what Keating et al. (2023) did by taking the middle of the vowel for their analysis. This choice minimizes the effect of the onset and offset of the vowel on the acoustic measures, which are more likely to be affected by the surrounding consonants and should give us the most accurate representation of the vowel quality. Because z-scores were used, this resulted in negative measures, which presents a problem for MDS analyses. To correct for this, I added the absolute value of the minimum z-score to each measure. This results in a dataset that still preserves the relative differences in the scores.

5.5.2 Parameter selection

In order to determine the correct number of trees and the number of predictors to use at each split, I followed the methods outlined in Boehmke & Greenwell (2019) and James et al. (2021) for parameter selection. The data was split into a training set and a test set. The training set was used to train the model and the test set was used to evaluate the performance of the model. The training set consisted of 70% of the data and the test set consisted of 30% of the data. The split was done so that the distributions of the different voice qualities was the same between the training and test sets. The training set was then used to tune the parameters of the model. The parameters that were tuned were the number of trees, the number of predictors to use at each split, the amount to sample from the training set, and whether to sample with replacement. The values for the parameters were chosen based on the results of a grid search.

Using a grid search allowed me to systematically search through the different combinations of parameters to find the best combination for the model. Another reason was that this allowed me to determine if Bagging (i.e., using all of the predictors at each split) or Random Forests (i.e., using a random subset of the predictors at each split) was the best model for the data. The model whose parameters resulted in the most accurate model was chosen as the final model.

Figure 5.3, shows the results of the grid search in relation to the number of trees and the number of predictors to use at each split. The x-axis shows the number of trees that were used in the model and the y-axis shows the percentage of incorrectly classified out-of-bag tokens. The lower the out-of-bag error percentage, the better the model is at predicting unseen data. The colored lines show the different values of m_{try} that were used in the model. When $m_{try} = 22$, the model uses all the predictors in each split and is therefore a bagging model. The other potential values for m_{try} are the values corresponding to 5%, 15%, 25%, and 40% of the total number of predictors. Furthermore, the default values of $m_{try} = \sqrt{22}$ and $m_{try} = \frac{22}{3}$ for classification and regression trees were considered, these values were recommended by Boehmke & Greenwell (2019).

Figure 5.3 shows that the model with $m_{try} = 22$ (i.e., bagging) is not the best model for the data. This is because it has a very high percentage of out-of-bag error compared to the other models. The model that performed the best was the model that trained on 300 trees and whose $m_{try} = 5$ had the lowest percentages of out-of-bag errors. Furthermore, it was found that sampling only 80% of the replacement data produced the best model.[1]

5.6 Results

The results of the Random Forest analysis are shown in Figure 5.4. The plot on the left shows the impurity importance which uses the mean decrease in the Gini index

[1] The full grid search results can be found online at my website: mlbrinkerhoff.me.

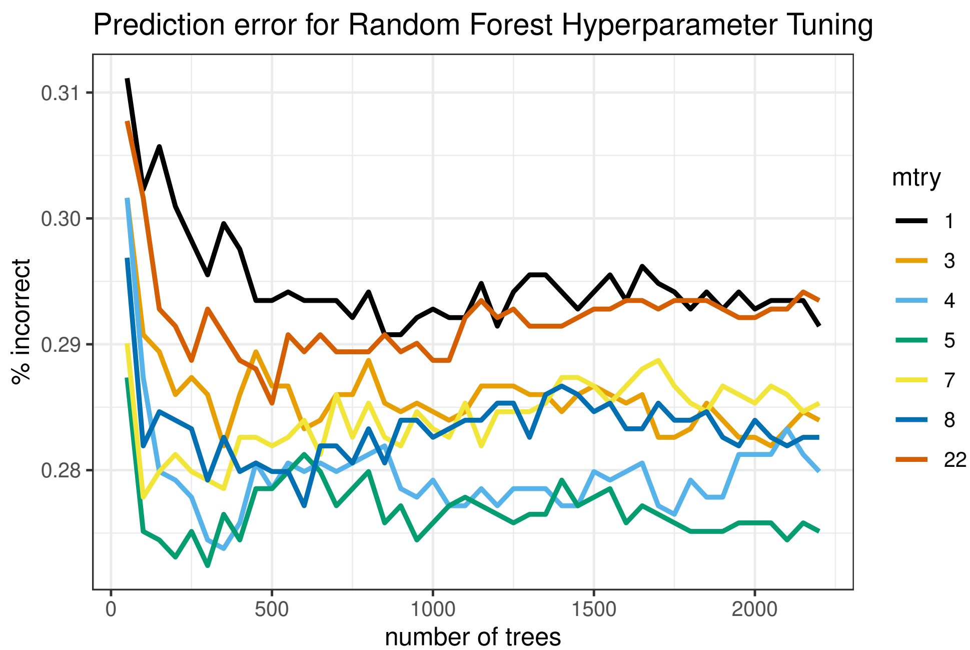


Figure 5.3: Plot showing the percent of inaccurately classified phonation types as a function of the number of trees ran. The different colored lines indicate the different m_{try} values.

to measure the importance of each predictor. The Gini index is a measure of how pure the split is and is calculated for each variable and then averaged across all trees. The higher the Gini index, the more important that variable is in making the splits in the data. The plot on the right shows the permutation importance, which measures the change in the model's prediction accuracy when the values of a variable are randomly permuted (i.e., the variable is shuffled and the model is reevaluated). The difference in accuracy between the original model and the permuted model is then used to measure the importance of that variable. The higher the difference, the more important that variable is in splitting the data. The y-axis shows all 21 different predictors for both plots. However, the predictors in each plot are ranked in descending order of importance.

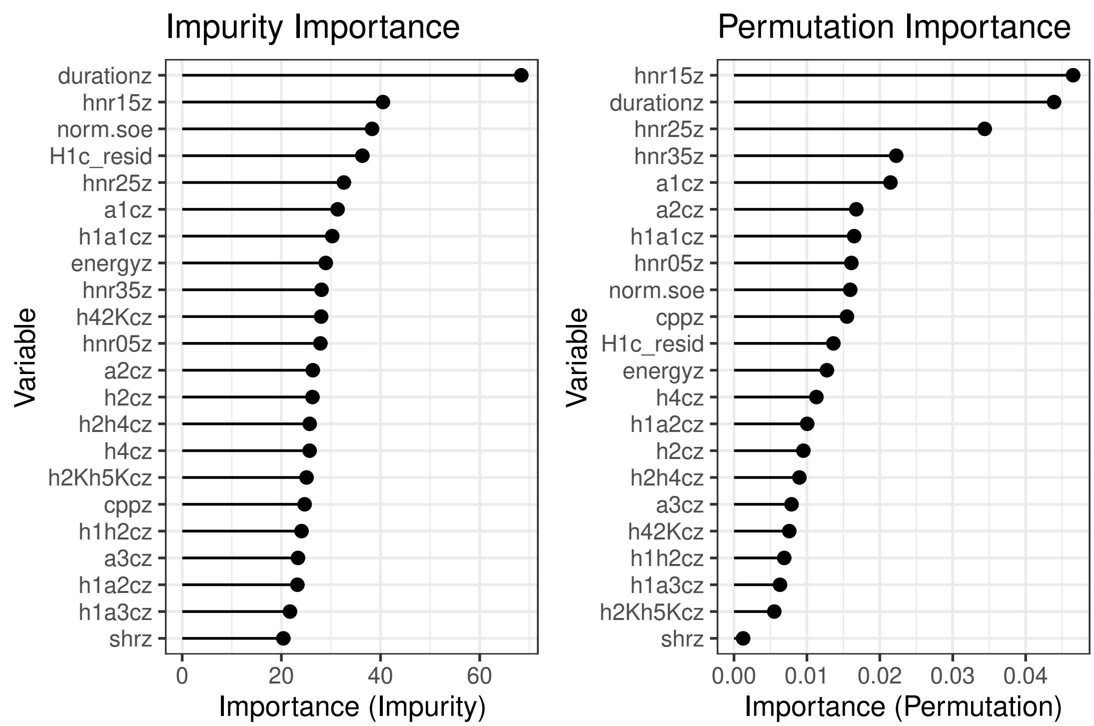


Figure 5.4: Variable importance plots showing the impurity importance and permutation importance of each acoustic measure.

From Figure 5.4, we can see that the two most important predictors for classifying the different phonations are: (i) duration and (ii) HNR < 1500 Hz. The reason for this is that both of these measures appear as the two highest predictors in terms of impurity and permutation. Each of the predictors suggested by the model will be discussed in more detail below.

The rest of the variables require some discussion to determine their importance. This is because the next several predictors in both plots are not consistent. In the impurity importance plot, we see that Strength of Excitation and residual H1* are the next most important measures. However, their importance decreases when we consider permutation. Instead of being the third and fourth most important impurity predictors, they are now the ninth and eleventh most important permutation predictors. Because both measures are sifted only slightly and are in the upper half of both plots, we can assume that these acoustic measures still play an important role to some extent.

Furthermore, we see that A1* is relatively consistent between the two plots. In the impurity plot on the left it is the sixth most important predictor and in the permutation plot on the right it is the fifth most important predictor. This suggests that A1* plays an important role in classifying the different phonations. The same reasoning holds for H1*−A1*, which is the seventh most important predictor in both plots.

5.7 Discussion of the results

This section will discuss the results of the Random Forest analysis in two parts. The first part will discuss the results of the Random Forest analysis in relation to the MDS analysis presented in Chapter 4. The second part will discuss certain acoustic measures that were shared by the analyses and those that were uniquely chosen by the Random Forest analysis.

5.7.1 Comparing the MDS and bagging trees

There was a large amount of overlap between which measures were correlated with the dimensions of the MDS analysis and the important predictors found by the Random Forest analysis. The correlations for the MDS analysis were found in Chapter 4 and are shown in Table 4.1. They are repeated here as Table 5.1 for convenience.

The rest of this section will discuss each of the measures where there was overlap between which measures showed a correlation to a MDS dimension and the Random Forest results. Furthermore, I will discuss duration and A1* due to how important the Random Forest model showed these measures. These measures are: (i) duration, (ii) A1*, (iii) H1*−A1*, (iv) residual H1*, (v) HNR < 1500 Hz, and (vi) Strength of Excitation. Each of these measures will be discussed in the following subsections in the order described above.

5.7.2 Importance of duration

In the Random Forest analysis, duration was found to be one of the most important predictors in classifying the different phonations. This importance is reflected in the fact that it is the most important predictor in the impurity importance plot and the second most important predictor in the permutation importance plot.

Table 5.1: Correlations for each acoustic measure to the four dimensions (NMDS1, NMDS2, NMDS3, NMDS4). The four largest correlations in each dimension are bolded.

Acoustic Measure	NMDS1	NMDS2	NMDS3	NMDS4
H1*−H2*	-0.221	-0.339	0.031	0.314
H2*−H4	-0.437	0.239	**-0.689**	**-0.364**
H1*−A1*	**-0.828**	0.048	**-0.459**	0.044
H1*−A2*	**-0.855**	-0.067	-0.343	0.114
H1*−A3*	**-0.809**	-0.218	-0.297	0.126
H4*−H2k*	-0.452	-0.598	0.294	**0.366**
H2k*−H5k*	0.152	0.023	0.101	0.057
residual H1*	-0.290	-0.443	**-0.722**	0.084
H2*	-0.157	-0.555	**-0.679**	0.114
H4*	0.295	**-0.778**	0.078	**0.479**
A1*	0.756	-0.549	0.092	0.124
A2*	**0.779**	-0.476	-0.103	0.086
A3*	0.735	-0.416	-0.211	0.093
CPP	-0.590	-0.606	0.209	-0.179
HNR < 500 Hz	-0.513	**-0.792**	0.152	-0.202
HNR < 1500 Hz	-0.275	**-0.799**	0.323	-0.290
HNR < 2500 Hz	-0.327	-0.714	0.391	-0.348
HNR < 3500 Hz	-0.446	-0.644	0.393	-0.356
Strength of Excitation	-0.013	-0.741	-0.238	0.145
SHR	0.144	-0.176	0.122	**-0.597**
Energy	-0.080	**-0.793**	-0.015	0.341
Duration	-0.622	0.539	0.257	0.030

Duration frequently plays an important role in Zapotec phonation contrasts. Many descriptions of Zapotec languages have shown that duration is often an important correlate for checked vowels (Ariza-García 2018). This has also been reported by Chai (2025) for Yateé Zapotec, a variety of Zapotec spoken in the Sierra Norte region of Oaxaca, Mexico, and a close relative of SLZ. Chai (2025) found that duration was the most important percept for distinguishing between Yateé Zapotec's checked and rearticulated vowels. Chai also found that the longer the duration of the vowel, the more likely the speakers of Yateé Zapotec were to classify the vowel as rearticulated. Similar results for checked vowels have been reported for other Zapotec varieties (Arellanes Arellanes 2009, 2010, Chávez-Peón 2010, López Nicolás 2016, Merrill 2008), other members of the Oto-Manguean family (e.g., Campbell 2014), and crosslinguistically (Gao & Kuang 2022, Chai & Ye 2022).

In their study on contextual enhancement effects on phonation contrasts in San Pablo Macuiltianguis Zapotec, Barzilai & Riestenberg (2021) found that duration played a role in distinguishing between the different types of phonation, regardless of the phrasal context. They found that modal vowels were the longest and checked vowels the shortest. In regards to rearticulated vowels, they claim that they do not function as a separate phonation type but rather are a sequence of a checked vowel followed by a modal vowel. Despite this reanalysis, we would then expect the duration of the rearticulated vowel to be longer, given that they are a sequence of two vowels.

In SLZ, the duration distribution between the different phonation types is shown in Figure 5.5. The y-axis shows the duration of each of the different phonation types in z-scores, again, z-scores were used to minimize individual speaker variation. The x-axis shows each of the four different phonations. The plot shows that the breathy voice has the longest duration and the rearticulated voice has the longest duration, followed by the checked voice. The shortest duration is found in the modal voice.

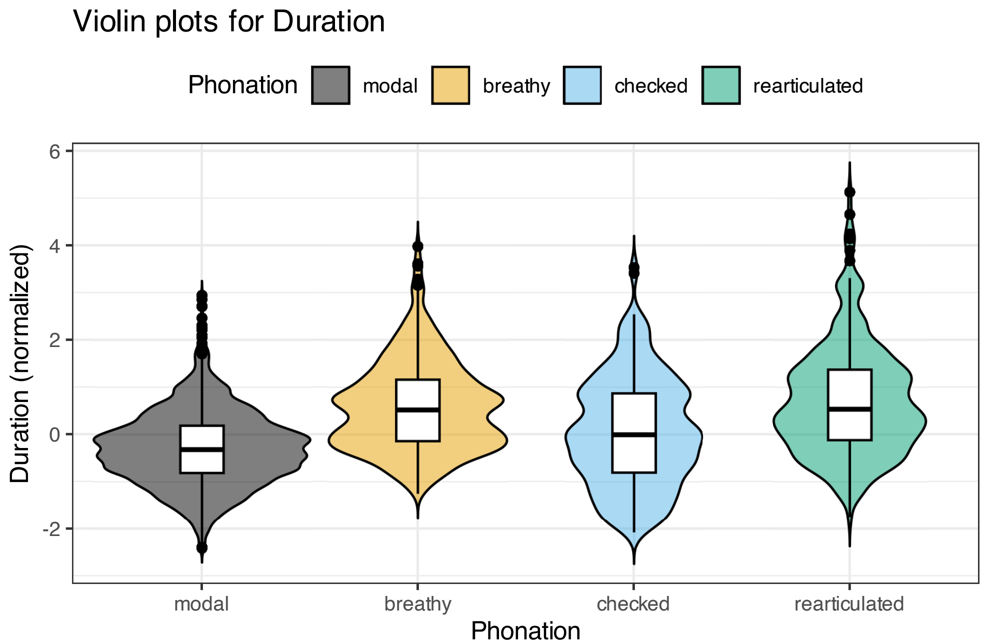


Figure 5.5: Plot showing the distribution of duration across the different voice qualities in SLZ.

These results are somewhat surprising because of the generalization that checked vowels are vowels that are abruptly stopped at the end of the vowel. This is not the case in SLZ. The checked vowels in SLZ are indeed the shortest of the three non-modal phonations. However, they are not the shortest when modal is included. This is likely due to the fact that checked vowels are often limited to final open word syllables in SLZ, and there is a general trend in Zapotec phonology for these syllables to be longer or minimally bimoraic (e.g., Chávez-Peón 2010, Nellis & Hollenbach 1980, Uchihara & Pérez Báez 2016).

When we consider the behavior of nonmodal phonation cross-linguistically, we see that the behavior of duration is consistent with what is generally found. For example, Gordon & Ladefoged (2001) and Esposito & Khan (2020) both report that nonmodal voice qualities are typically longer than modal voice. One of the reasons for this is that by lengthening the vowel, the speaker is able to increase the amount of time that the glottal source is able to produce the desired voice quality, which increases the likelihood that the listener will be able to perceive the desired voice quality.

The behavior of duration warrants further investigation in SLZ. If checked vowels are indeed longer than modal vowels, then this would suggest that the checked vowels in SLZ are not the same as the checked vowels in other Zapotec languages. Furthermore, vowels have been claimed to undergo a process of vowel lengthening in Zapotec languages depending on the type of syllable it is in and whether the coda is a fortis or lenis consonant (Nellis & Hollenbach 1980, Uchihara & Pérez Báez 2016).

If the vowel is in an open syllable or a closed syllable with a lenis coda, then the vowel is lengthened. If the vowel is in a closed syllable with a fortis coda, then the vowel is not lengthened. If this behavior is true for SLZ, this means that the duration of the vowel depends not only on the type of phonation but also on the type of syllable in which it is placed. This requires further investigation to determine if the duration is affected by the phonation and type of syllable.

5.7.3 Importance of A1*

The measure that was found to be the most important for classifying the SLZ phonation contrasts was A1*. This measure captures the amplitude of the harmonic closest to the first formant. This measure is typically not found in voice quality research except as a way to normalize the amplitude of the first harmonic (i.e., H1*). This goes back to Fischer-Jørgensen (1968) who used this as one of the ways to correct for the high-pass filtering, in addition to the widely used H1*−H2* measure. However, A1* as a standalone measure is not typically used in voice quality research. Therefore, what we are to make of this measure's importance in SLZ is not clear as is its behavior in regards to the different phonations.

When we look at how A1* is distributed across the different voice qualities in SLZ, as seen in Figure 5.6, we see that the modal vowels are found at the top of the chart with the other phonation contrasts located lower on the chart. The phonation that is found at the bottom of the chart is a breathy voice.

The pattern of behavior for A1* is very similar to what is found in descriptions of the behavior of the frequency of F1 in breathy voice contexts. For example, differences in the frequency of F1 are typically used to distinguish register differences in Southeast Asian languages (Brunelle 2005, 2012, Brunelle & Kirby 2016).

As described in Brunelle & Kirby (2016), many of the languages of Southeast Asia have what is called register. The linguistic term register is defined as "the redundant

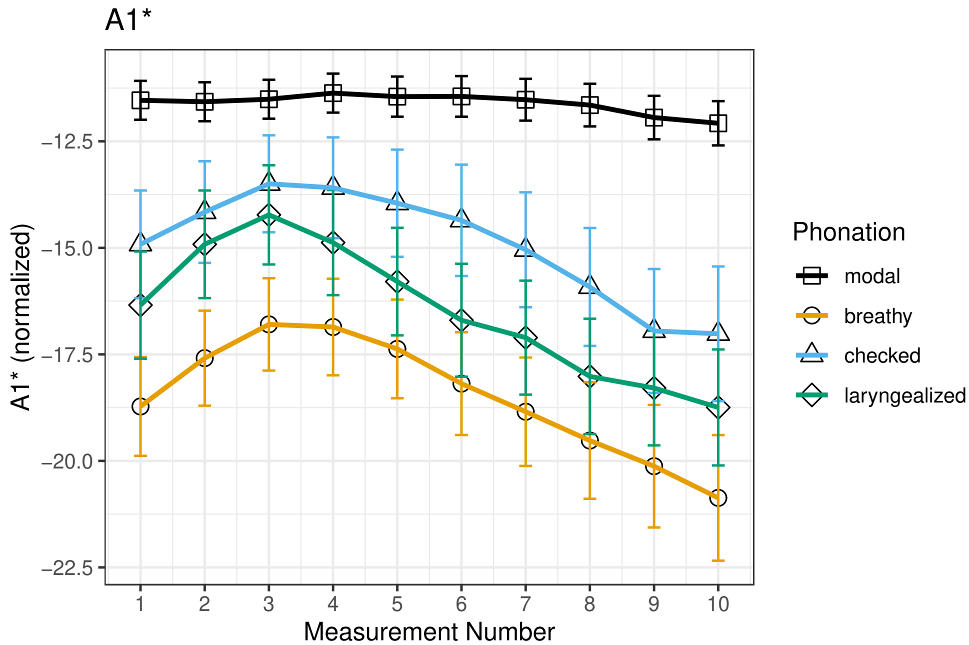


Figure 5.6: Plot showing the distribution of A1* across the different voice qualities in SLZ.

use of pitch, voice quality, vowel quality, and durational differences to distinguish (typically two) contrastive categories" (Brunelle & Kirby 2016: p. 193), and was first used by Henderson (1952) to describe the categorical contrasts found in Khmer. The characteristics that define the higher and lower registers are found in Table 5.2.

From this table, the lower register is what interests us here. The lower register is associated with breathy voice and a lowered first formant. There is evidence that breathy voice frequently has a lower first formant than modal voice in paralinguistic settings for English (Lotto, Holt & Kluender 1997). However, these studies do not discuss the amplitude of the first formant but rather the frequency of the first formant.

Table 5.2: Possible phonetic correlates of register. From Brunelle & Kirby (2016).

Higher Register	**Lower Register**
Higher pitch	Lower pitch
Tense/Modal voice	Lax/Breathy voice
Monophthongs/shorter vowels	Diphthongs/longer vowels
Raised F1/lower vowels/[+ATR]	Lowered F1/higher vowels/[-ATR]
Plain stops/shorter VOT	Aspirated stops/longer VOT

Another comparison can be found in the research on nasality, where this measure is discussed extensively either alone or in association with the nasal pole (e.g., Chen 1997, Delvaux 2009, Macmillan et al. 1999, Pruthi & Espy-Wilson 2004, Schwartz 1968, Stevens 2000, Styler 2015, 2017). These studies discuss how in nasalized contexts the amplitude of the first formant is typically found to be lower than in oral contexts, again similar to what we see in Figure 5.6.

There is a large body of research that discusses that nasality is closely associated with breathy voice or glottal consonants, in a phenomenon called *rhinoglottophilia* (Matisoff 1975, Ohala 1975, Ohala & Ohala 1993, Bennett 2016). In Blevins & Garrett (1993) and Matisoff (1975), this association is attributed to the acoustic and perceptual similarities between nasalization and breathy voice. In one study, Garellek, Ritchart & Kuang (2016) showed that nasalization in three different Yi languages was associated with a slender voice. The authors suggested that breathy voice during nasalization can arise from misperception or as a type of phonetic enhancement.

In terms of what is going on in SLZ, it is not clear if the lowering of A1* for breathy voice can be attributed to the same observations as discussed above. I suggest that there are three different possibilities to explain what is going on in SLZ. First; because SLZ does not have phonemic nasalization, it is possible that speakers of SLZ are using nasalization as a way to phonetically enhance the contrast of breathy voice. Essentially, the reverse of what was reported by Garellek, Ritchart & Kuang (2016). This possibility could be tested acoustically by performing an experiment to detect nasal airflow during breathy vowels. The second possibility is that the same measures that work for detecting nasality can also be used to detect breathy voice. This second possibility could be easily tested by examining whether A1* and the other measures for nasality work in other breathy voice contexts cross-linguistically. The third possibility is that the lowering of A1* is a result of subglottal resonances, which is much harder to test than the other two possibilities.

5.7.4 Importance of H1*−A1*

The Random Forest analysis found H1*−A1* to be one of the most important predictors in classifying the different phonations. This is because of its high ranking in both the impurity and permutation importance plots. In the impurity plot, it is the seventh most important predictor, and in the permutation plot, it is the sixth most important predictor.

The acoustic measure H1*−A1* is a measure of the difference between the amplitude of the first harmonic and the amplitude of the harmonic closest to the first formant. This acoustic measure falls into the category of spectral slope measures. As discussed in Chapter 3, these measures capture the differences in the amplitude of the harmonics of the spectrum. The higher the values of the spectral slopes, the more breathy the phonation is (Fischer-Jørgensen 1968). The lower the spectral slope values, the more creaky the phonation is.

The distribution of H1*−A1* across the different phonation types is shown in Figure 5.7. The y-axis is the z-scored H1*−A1* values and the x-axis shows the ten equally spaced intervals throughout the duration of the vowel.

The plot shows that breathy voice has the highest H1*−A1* score, which is consistent with what we would expect from a spectral slope measure. Things are more complicated when we look at the other non-modal phonations. It is expected that checked voice and rearticulated voice have a lower spectral slope than modal voice because of their association with creaky voice. However, this is not the case in Figure 5.7.

The plot shows that the checked voice has a spectral slope that is identical to that of the modal voice. Additionally, we see that the rearticulated voice has a spectral slope that is higher than the modal voice except at measurement intervals three and four, where it is identical to the modal voice. This behavior for a rearticulated voice is not too surprising, especially at the end of the vowel.

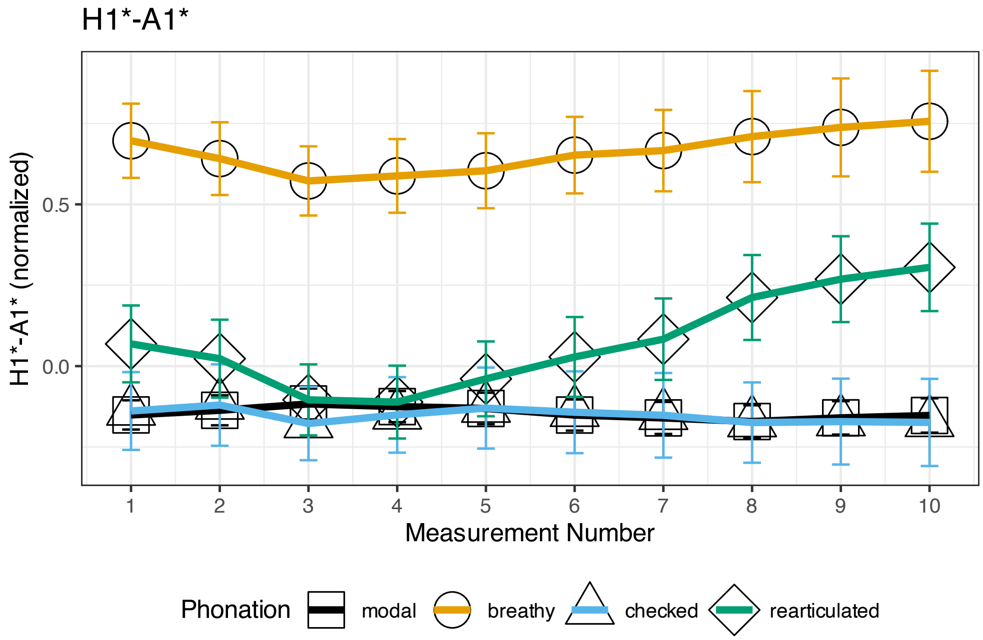


Figure 5.7: Plot showing the distribution of H1*−A1* across the different voice qualities in SLZ. Each point represents the mean of the ten equally spaced intervals across the duration of the vowel and the error bars represent a 95% confidence interval.

One reason for this is that when listening to the rearticulated vowels, the portion of the vowel after the glottal occlusion sounds somewhat breathy. One of the possible explanations for this is that in order to produce the glottal occlusion, the speaker is tightly constricting the vocal folds, which causes either a glottal stop or creaky voice. The speaker then relaxes the vocal folds in order to produce the modal voice which stereotypically occurs after this glottal occlusion. However, one way for the speaker to quickly relax the vocal folds is to open them as much as possible to reach the ideal position for the modal voice. This general behavior of opening the vocal folds to produce a modal voice is similar to what is found in a breathy voice. This could be a possible explanation for why the rearticulated voice has a higher H1*−A1* score than modal voice.

5.7.5 Importance of Residual H1*

Another spectral slope measure that was found to be important in the Random Forest and MDS analyses was residual H1*. This measure is similar to H1*−A1* in that it captures the differences in the amplitude of the harmonic of the spectrum. However, residual H1* is a measure that is calculated after removing the effect of energy on H1*.

As residual H1* is a type of spectral slope measure, it is expected to show similar behavior to other spectral slope measures. That is, a breathy voice should be associated with a higher score than a modal voice. Since the two types of laryngealization are closely associated with creaky voice, they should be associated with a lower score than modal voice. Additionally, there is a temporal difference between the two types of laryngealization, it is expected that checked voice will have a lower score toward the end of the vowel and rearticulated voice will have a lower score in the middle of the vowel.

In Figure 5.8, we see that the predictions are correct. Breathy voice is associated with a higher score than the modal voice. The two types of laryngealization are associated with a lower score than the modal voice.

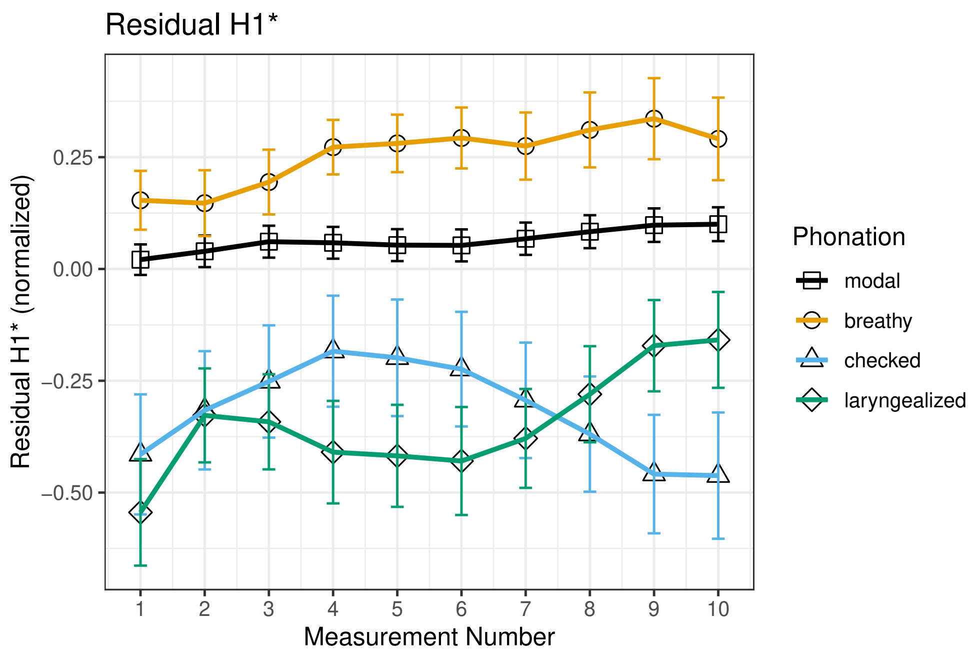


Figure 5.8: Plot showing the distribution of residual H1* across the different voice qualities in SLZ.

A further analysis of the results of this measure using generalized additive (mixed) models (GAM(M)s; Hastie & Tibshirani 1986, Wood 2017, Sóskuthy 2017, Wieling 2018) shows that the temporal behavior of the two types of laryngealization is also correct. Checked voice is associated with a lower score toward the end of the vowel, and rearticulated voice is associated with a lower score in the middle of the vowel. Additionally, the results of the GAMM show that for the first three time points there is no significant difference between the breathy and modal voice. The suggestion is that the breathy voice is more closely associated with the latter part of the vowel than with the beginning. The full results of the GAMM for residual H1* are discussed in

Chapter 3.

5.7.6 Importance of the HNR <1500 Hz

The acoustic measure HNR $<$ 1500 Hz is a harmonics-to-noise ratio measure that is calculated over the frequency range from 0 Hz to 1500 Hz and is a measure of the amount of noise in the signal, or in other words, it is a measure of periodicity. In these measures, modal voice is associated with a higher score than nonmodal phonations. This is because modal voice is associated with a higher degree of periodicity than nonmodal phonations because a-periodicity is a defining feature of nonmodal phonations (e.g., Hillenbrand & Houde 1996a, Blankenship 1997, Kent & Ball 1999).

As seen in Figure 5.9, this is exactly what we see in SLZ. Modal voice is associated with a higher score than non-modal phonations. Among nonmodal phonations, breathy and rearticulated voice have a higher HNR $<$ 1500 Hz than checked voice.

The fact that breathy and rearticulated voice has a higher HNR $<$ 1500 Hz than checked voice is not surprising. As discussed in Chapter 2, rearticulated and breathy voices are associated with more modal-like qualities than checked voice. In most instances of rearticulated voice, the vowel is produced with modal voice except for a small portion of the middle of the vowel where the vowel is produced with creaky voice or a glottal occlusion. Breathy voice typically appears very periodic but with a high degree of noise in the signal. Checked voice on the other hand, is associated

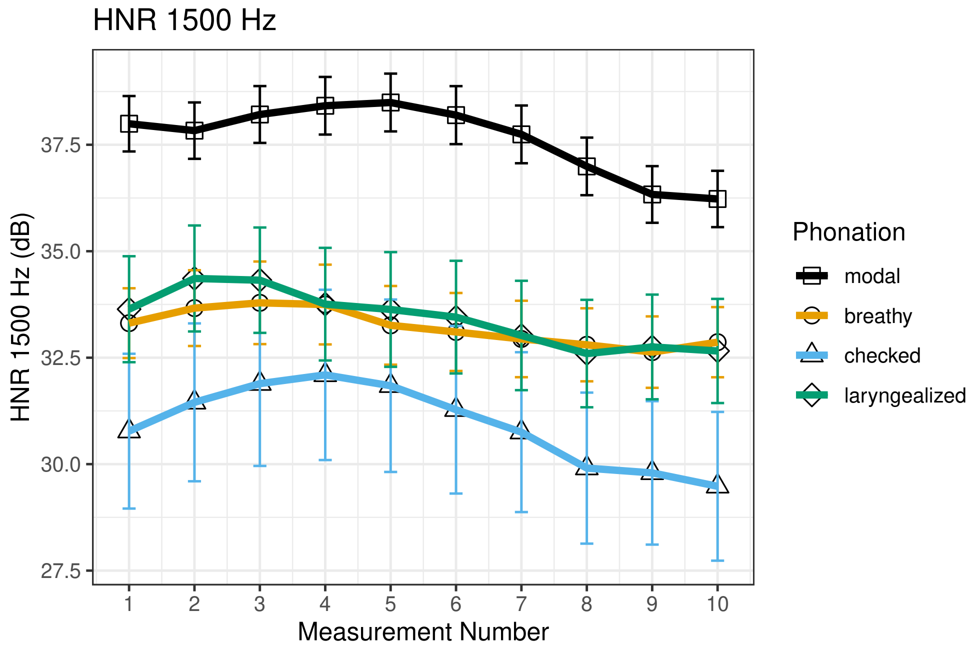


Figure 5.9: Plot showing the distribution of HNR < 1500 Hz across the different voice qualities in SLZ.

with a high degree of aperiodicity in the signal from the creaky voice that is produced at the end of the vowel. This is why checked voice has the lowest HNR < 1500 Hz of nonmodal phonations.

This measure will be very important in Chapter 6, where I will use it to model laryngeal complexity.

5.7.7 Importance of Strength of Excitation

Strength of Excitation (SoE) is a measure that is defined as "the instant of significant excitation of the vocal-tract system during production of speech [and represents] the relative amplitude of impulse-like excitation" (Mittal, Yegnanarayana & Bhaskararao 2014: p. 1934). This measure is typically associated with the amplitude of the voicing, or in other words, the degree of voicing during production (Murty & Yegnanarayana 2008, Mittal, Yegnanarayana & Bhaskararao 2014). This measure has been shown to be an effective measure for showing the effects of laryngealization on the amplitude of voicing (Garellek et al. 2021) and has been shown to be effective in distinguishing the different phonations in the Oto-Manguean languages (Chai, Fernández & Mendez 2023, Weller et al. 2023a,b, 2024).

Following Garellek et al. (2021), it is expected that the modal voice will have the highest SoE score. It is also expected that, because the breathy, rearticulated, and checked voice will have a lower SoE score. The reason for this expectation is that according to Garellek et al. (2021), there is a strong tendency for all types of laryngealization to have a dampening effect on voicing. This is evidenced by the lower SoE scores for these phonations. The score for SoE ranges from 0 to 1, where 0 is no voicing and 1 is full voicing.

In Figure 5.10, we observe that the modal voice has the highest SoE score. The three non-modal phonations are all lower. We can ignore the SoE values at measurement numbers one and ten because they represent the beginning and end of the vowel and are more likely to be affected by the coarticulatory effects of the surrounding consonants. However, we can see that breathy voice has a higher SoE score than checked voice and rearticulated voice. Additionally, we see that between checked and rearticulated voice, we see that checked voice has a higher SoE score in the first half of the vowel and rearticulated voice has a higher SoE score in the second half of the vowel. This is consistent with what we would expect from the behavior of the phasing difference between checked and rearticulated vowels.

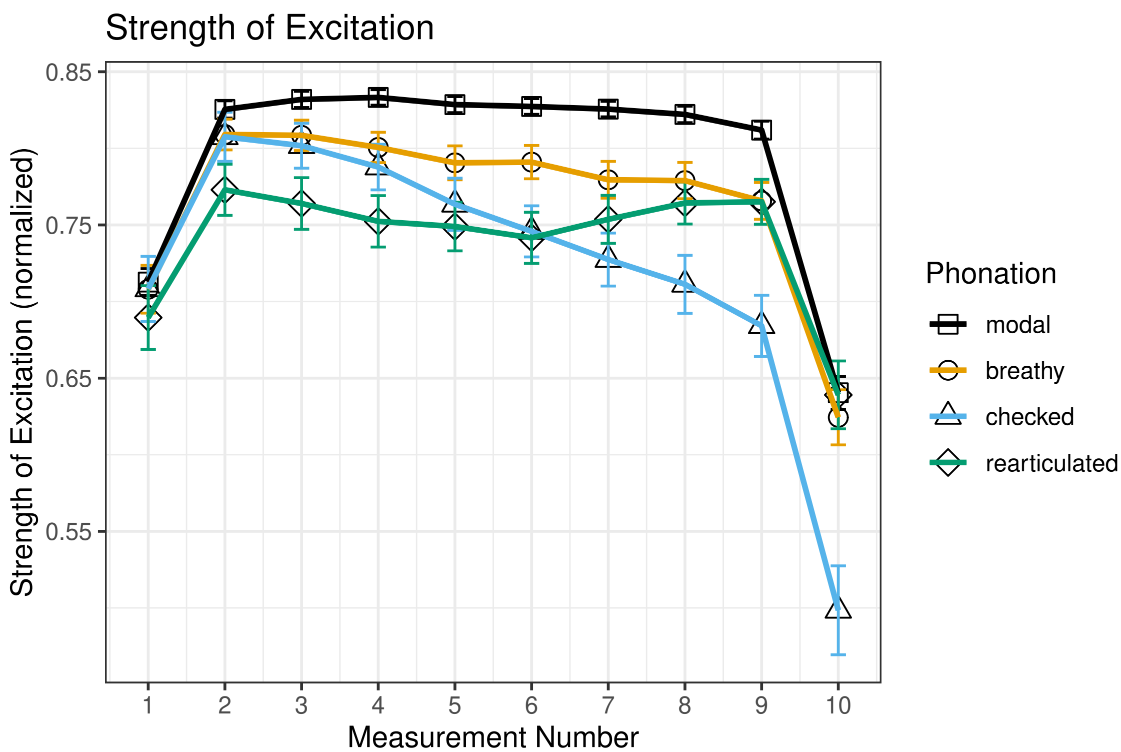


Figure 5.10: Plot showing the distribution of Strength of Excitation across the different voice qualities in SLZ.

One thing to note about the SoE measures is that even though nonmodal phonations are associated with a lower SoE score than modal voice, the values for these measures are still extremely high. This likely indicates that even though laryngealization is present in these phonations, the speakers are only weakly laryngealizing the vowel. This is consistent with the claims made by Silverman (1997a,b) that laryngeally complex languages that do not show a phasing relationship between modal and non-modal phonation must weakly produce non-modal phonation. The claims about laryngeal complexity will be discussed in more detail in Chapter 6.

5.8 Conclusion

In summary, we find that the results of the Random Forest and the MDS analyses showed a lot of overlap between which measures were correlated and/or important for classifying SLZ's four phonation types. In both analyses, A1*, H1*−A1*, residual H1*, HNR $<$ 1500 Hz, and SoE were found to be important acoustic measures. We saw that each of these measures was able to capture some aspect of the phonation types. It also shows that the spectral slope measures and harmonics-to-noise ratio measures are generally the most important measures for classifying the different phonation types. This is consistent with previous work on the acoustics of phonation (e.g., Garellek 2019). We also saw that SoE plays an important role in classifying the different phonation types. This contributes to the growing body of literature that shows that SoE is an important measure for classifying the different phonation types (Chai, Fernández & Mendez 2023, Garellek et al. 2021, Weller et al. 2023a,b, 2024).

Interestingly, the results of the Random Forest analysis showed that duration was the most important predictor in classifying the different phonations. This is consistent with what has been found in other Zapotec languages (e.g., Barzilai & Riestenberg 2021, Chai 2025, Chávez-Peón 2010). However, when we plotted the duration, it showed that modal vowels had the shortest duration instead of checked vowels having the shortest duration. The effect of duration needs to be further investigated in SLZ, especially with respect to differences in syllable type and the claimed effects of vowel lengthening in the various Zapotec languages (e.g., Chávez-Peón 2010, Merrill 2008, Nellis & Hollenbach 1980, Pickett, Villalobos & Marlett 2010, Uchihara & Pérez Báez 2016).

Chapter 6

Laryngeal complexity's realization in Santiago Laxopa Zapotec

6.1 Introduction

Laryngeal complexity is a term used to describe languages that have contrastive tone and phonation, such as the Oto-Manguean languages (Blankenship 1997, 2002, Silverman 1997a,b). Laryngeally complex languages have two ways that tone and phonation can be realized: (i) in a temporally ordered fashion, or (ii) simultaneously (Silverman 1997a). In the first case, tone and phonation are phased concerning one another, with one portion of the vowel realized with modal phonation with the acoustic correlates of tone and the other with nonmodal phonation with the acoustic correlates of phonation. In the second case, tone and phonation are not phased with one another. Instead, the acoustic correlates of phonation or tone are weakly applied, so both tone and phonation are realized simultaneously.

Laryngeal complexity in Oto-Manguean languages generally follows the first pattern, where tone and phonation are phased with each other, such as Jalapa Mazatec (Blankenship 1997, 2002, Silverman 1997a,b). However, Zapotec languages are described as lacking this phasing. Instead, tone and phonation are described as "reach[ing] a degree of equilibrium, weakening enough to be present simultaneously (Herrera Zendejas 2000: p. 558)." However, this description is based on the subjective impression of what was observed in the spectrograms for Isthmus Zapotec and not any quantitative analysis of the acoustic properties of the phonation.

As discussed in Chapter 2, Santiago Laxopa Zapotec (SLZ) is also a laryngeally complex language and, based on Herrera Zendejas's (2000) claim, should lack phasing and instead have tone and phonation realized simultaneously. This chapter presents an acoustic analysis of tone and phonation in SLZ in a similar vein as DiCanio (2012a), Garellek & Keating (2011), Kelterer & Schuppler (2020), and Weller et al. (2023a,b, 2024). This is done by looking at the acoustic properties of SLZ phonation using a combination of Strength of Excitation (SoE), Harmonic-to-Noise Ratio (HNR) < 1500 Hz and $f0$ perturbations. This analysis demonstrates that SLZ weakly produces phonation and that there is phasing between portions of the vowel realized with modal phonation and portions realized with nonmodal phonation.

These measures were selected based on the results of the MDS analysis in Chapter 4 and the Random Forest analysis in Chapter 5. MDS and Random Forest analyses showed that SoE and HNR < 1500 Hz were the most important measures to distinguish between the different phonation types. SoE has previously been used to show that there is a clear phasing between modal and rearticulated vowels in San Sebastián del Monte Mixtec (Weller et al. 2023a,b, 2024). Similarly, $f0$ perturbation has been used previously to show that there is phasing between modal and nonmodal phonation in several Oto-Manguean languages (Garellek & Keating 2011, DiCanio 2012a, Kelterer & Schuppler 2020). To date, HNR measures have not been widely used to demonstrate phasing, except for cepstral peak prominence for Chichimec (Kelterer & Schuppler 2020).

The remainder of this chapter will be organized as follows. First, in Section 6.2, I will briefly overview laryngeal complexity, phasing, and recoverability of tone and phonation in Section 6.2. In Section 6.3, I will discuss previous analyses of laryngeal complexity. In Section 6.4, I will discuss the methods used to analyze laryngeal complexity in SLZ. I will present the analysis results in Section 6.5. Finally, in Section 6.6, I will discuss the implications of these results for our understanding of laryngeal complexity in SLZ.

6.2 What is Laryngeal Complexity?

Laryngeal complexity is defined as the contrastive use of tone and phonation within the same syllabic nucleus (Blankenship 1997, 2002, Silverman 1997a,b). This use of contrastive tone and phonation is one of the defining characteristics of the Oto-Manguean languages (Silverman 1997a). However, it is not limited to just these languages. It has also been used to describe the behavior of tone and phonation interactions in languages outside of the Oto-Manguean languages; such as the Tibeto-Burman languages of Mpi and Tamang (Silverman 1997a,b), the Mayan language Yucatec Mayan (Frazier 2013), and to describe the behavior of coarticulatory pitch and phonation in the Germanic language Danish (Frazier 2013, Peña 2022, 2024).

6.2.1 Phasing and recoverability

According to Silverman (1997a,b), one of the defining aspects of laryngeal complexity is the concept of phasing and recoverability; the phonation and tone are phased with respect to each other in a way that lends itself to a listener's ability to recover the underlying phonation and tone. In practical terms, laryngeally complex vowels have two components: a modal phonation portion where tone is realized, and a nonmodal phonation portion of the vowel. For the researcher, two portions of the vowel can be analyzed temporally and spectrally (Silverman 1997a: p. 237).

For example, in the Oto-Manguean language Jalapa Mazatec, breathiness or creakiness is realized only on the first portion of the vowel, either as a full laryngeal consonant or as a laryngeal feature on the vowel (Silverman 1997a: p. 238). The second portion of the vowel is realized as a modal voice vowel with one of the three tones belonging to the tonemes of the language.

Silverman argues that three principles help explain why laryngeal complexity should be temporally ordered or phased: (i) sufficient acoustic distance, (ii) sufficient articulatory compatibility, and (iii) optimal auditory salience.

6.2.1.1 Sufficient acoustic distance

Silverman (1997a) argues that sufficient acoustic distance is necessary to recover phonation and tone. Silverman explained that listeners do not rely on the fundamental frequency alone to perceive pitch. Instead, listeners use the signal's harmonic spacing and pulse period to perceive the pitch (Ritsma 1967, Remez & Rubin 1993). The harmonic spacing and pulse periods are present for modal phonation and encode a salient pitch value. However, the harmonic spacing and pulse periods are often obscured or absent during nonmodal phonation.

For breathy voice, there is a general weakening of the harmonic structure that makes it difficult for the listener to recover the pitch (Silverman 1997b). Creaky voice, conversely, had obscured the pulse periods due to its aperiodic and unstable glottal vibration (Ladefoged & Maddieson 1996). This behavior was observed in Mazatec, where the harmonic structure is gone and the pulses are indiscernible (Kirk, Ladefoged &

Ladefoged 1993). Furthermore, pitch perception is rendered indiscernible when pulse periods are varied by 10% or more (Rosenberg 1966).

These observations lead Silverman (1997a) to conclude that if a period glottal wave is either obscured (as with breathy voice) or not present (as with creaky voice), the acoustic signal cannot encode a salient pitch value. For laryngeally complex languages, it is better if phonation and tone are phased to allow the listener to recover the underlying phonation and tone.

6.2.1.2 Sufficient articulatory compatibility

Another essential point for Silverman's theory about laryngeal complexity has to do with the articulatory compatibility of the phonation and tone. One of the guiding ideas behind this principle is that there is a principle of least effort in biological motor systems such as speech production (Lindblom 1983). According to Lindblom (1983), speech gestures can be considered distinct motor goals in our speech production system. The speaker achieves these goals by coordinating the articulators with the least effort, manifesting through either sequencing or coarticulating the gestures.

An example of this articulatory compatibility comes from nasalization. In nasal contexts, the velum is lowered to couple the nasal cavity with the oral cavity, introducing nasal resonances to the signal. This velum-lowering gesture is compatible with the gestures needed in the oral cavity to produce different vowel qualities. This leads to the production of nasal vowels in languages like French or Portuguese. This lowering also occurs in languages that do not have contrastive nasal vowels, such as English, where the velum is lowered in anticipation of a nasal consonant (e.g., Ohala 1975).

For Silverman's 1997 theory of laryngeal complexity, the articulatory mechanisms for tone and phonation are precisely the same, which leads to a need to phase the two to make use of the same articulatory gestures optimally. However, there is a growing body of literature that shows that tone and phonation are much more complex and rely on the entire larynx, not just the vocal folds (see Esling et al. 2019 for a detailed discussion of this). This matter will be discussed again in Section 6.6.

6.2.1.3 Optimal auditory salience

The last important point for Silverman's theory of laryngeal complexity is the idea of optimal auditory salience. Research into the behavior of nerve responses to acoustic signals has shown that the auditory system is sometimes more and other times less recoverable depending on the signal, which led Bladon (1986) to propose two major principles for auditory phonetics: (i) the on/off response asymmetry and (ii) short-term adaptation.

The on/off response asymmetry principle originates in the work of Tyler et al. (1982). This principle states that "spectral changes whose response in the auditory nerve is predominantly an onset of firing are much more perceptually salient than those producing an offset. (Silverman 1997a: p. 249);" Changes in the signal that cause an increase in the firing rate in the auditory nerve are easier to perceive when this firing starts compared to when it ends.

The second principle, short-term adaptation, states that "after a rapid onset of auditory nerve discharge at a particular frequency, there is a decay to a moderate level of discharge, even though the same speech sound is continuing to be produced (Silverman 1997a: p. 250)" and is based on the work of Delgutte (1982). This means that the auditory nerve is less sensitive to changes in the signal after a rapid onset of firing, which in turn means that the auditory nerve is less able to recover the underlying phonation and tone when the signal is not changing.

For Silverman, these two principles are important because they help provide an acoustic explanation for why laryngeal complexity should be phased. Phasing tone and phonation enable the listener to recover the signal more easily. This is because when there is a change in the signal, the auditory nerve has the greatest chance of recovering the signal. Regarding tone and phonation, when there is a change from tone's modal phonation to nonmodal phonation or from nonmodal phonation to tone's modal phonation, the listener can perceive the difference better than if the two were produced simultaneously. Furthermore, these two principles suggest that some phasing relationships are more perceptually salient than others. Based on these two principles, things are easier to perceive if they are produced first or if there is a significant change in the signal.

To illustrate this point graphically, Silverman provides a series of figures that show the relationship between the articulatory, acoustic, and perceptual components of laryngeal complexity. These figures are shown in Figures 6.1, 6.2, and 6.3.

Figure 6.1 shows the sequence characteristics of [ha˥], where breathiness is first produced, followed by a modal vowel with a high tone. The figure shows that when the breathiness is first produced, the auditory nerve response is high, followed by a sharp decline throughout the rest of the breathiness. As the energy increases when the modal vowel with high tone is produced, there is a sharp increase in the auditory nerve response, which is followed by a gradual decline, making it easier for the listener to perceive the transition and recover the underlying phonation and tone.

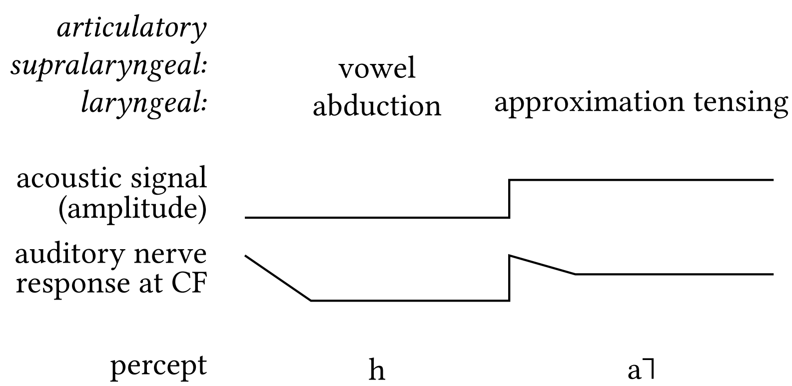


Figure 6.1: Schematic representation of the characteristics of [ha˥] sequences. Adaptation of figure from Silverman (1997a).

For Figure 6.2, the sequence is reversed. In this case, the high-tone modal vowel is produced first, followed by breathiness. The figure shows that when the high tone modal vowel is produced first, the auditory nerve response is at its highest, followed by a sharp decline through the rest of the modal vowel portion. Because no energy increase is associated with breathiness, there is a continual decrease in the auditory nerve response. It is difficult for the listener to perceive the transitions and recover the underlying phonation and tone.

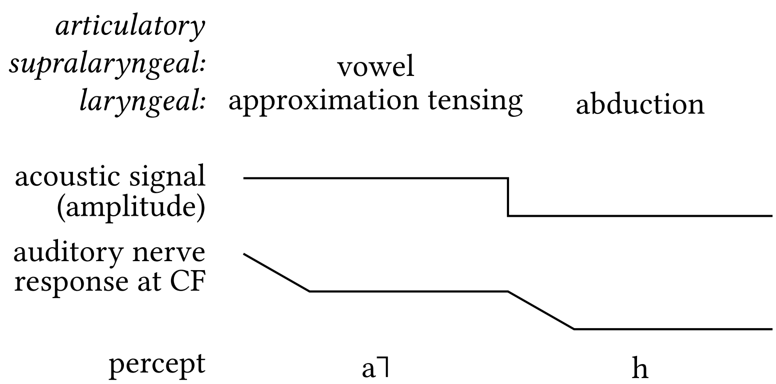


Figure 6.2: Schematic representation of the characteristics of [a˥h] sequences. Adaptation of figure from Silverman (1997a).

About the third type of sequence, [aha˥], the figure shows that the auditory nerve response is at its highest when the modal vowel is produced first. This is followed by a decline similar to what we see in Figure 6.2. However, because there is a high-tone modal vowel, we see a sharp increase in auditory nerve response. Again, this is not as optimal as having the breathiness produced first.

In summary, Silverman claims that due to this auditory response asymmetry, we see a difference in the auditory nerve response depending on the sequencing of the phonation and tone. For Silverman, this implies an implicational hierarchy between

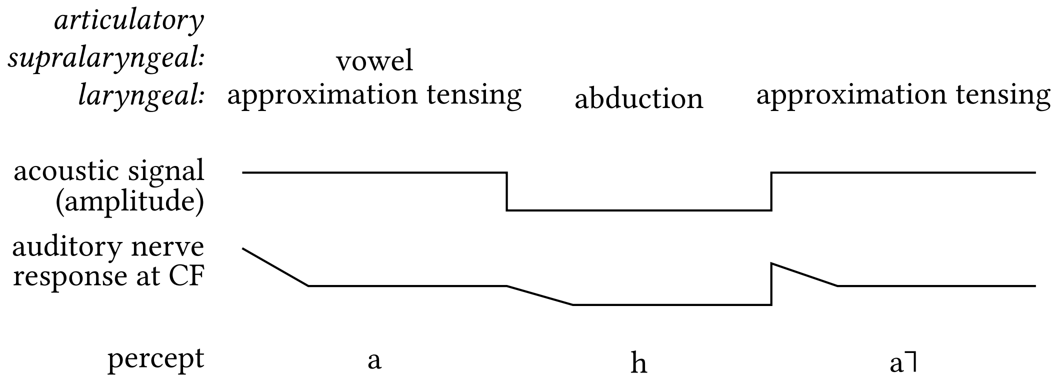


Figure 6.3: Schematic representation of the characteristics of [aha˥] sequences. Adaptation of figure from Silverman (1997a).

the different types of sequences found in laryngeal complexity. This implicational hierarchy will be discussed in more detail in Section 6.2.2.

6.2.2 Implicational hierarchy of laryngealization

Another aspect of Silverman's 1997 laryngeal complexity theory is that there is an implicational hierarchy in the phasing and ordering of phonation and tone. This hierarchy is based on the way laryngealization appears in three Oto-Manguean languages. In this implicational hierarchy, laryngealization can only appear in three ways: prevocalic, postvocalic, or interrupted. In the prevocalic laryngealization appears before the vowel. In the postvocalic case, the laryngealization appears after the vowel. Laryngealization interrupts a vowel and appears in the middle in the interrupted case.

According to Silverman (1997a), if a language has interrupted laryngealization, it must also have postvocalic laryngealization. If a language has postvocalic laryngealization, it must also have prevocal laryngealization. In support of his claims, Silverman (1997a) provides data from three Oto-Manguean languages: Jalapa Mazatec, Comaltepec Chinantec, and Copala Trique. These languages are shown in Table 6.1.

Table 6.1: Implicational hierarchy of laryngeal complexity. The symbols h and ʔ represent laryngealization. The symbol V represents where the modal vowel is located in relation to the laryngealization. Modified from Silverman (1997a).

Language	**Prevocalic**	**Postvocalic**	**Interrupted**
Jalapa Mazatec	hV˥, ʔV˥	–	–
Comaltepec Chinantec	hV˥, ʔV˥	Vh˥, Vʔ˥	–
Copala Trique	hV˥, ʔV˥	Vh˥, Vʔ˥	VhV˥, VʔV˥

The implicational hierarchy seems to hold for many descriptions of languages with laryngeal complexity, which is the case in the other Trique languages (DiCanio 2008, 2010, 2012a,b, 2014, DiCanio et al. 2020, Elliott, Edmondson & Cruz 2016, Hollenbach 1984). However, as mentioned in Frazier (2013), it is not clear how accurate or robust this implicational hierarchy is actually. It is not always clear whether something is a laryngeal consonant or a laryngeal feature in the vowel. In fact, Silverman (1997a,b) treats laryngeal consonants and laryngeal features in vowels as the same thing. For example, in many Trique languages, the laryngealization is realized as a laryngeal consonant (DiCanio 2008, 2010, 2012a,b, 2014, DiCanio et al. 2020, Elliott, Edmondson & Cruz 2016, Hollenbach 1984). Still, in Jalapa Mazatec, the laryngealization is realized as a laryngeal feature on the vowel (Kirk, Ladefoged & Ladefoged 1993, Garellek & Keating 2011).

Another issue with the implicational hierarchy is that it is based on only three languages. The relatively small sample size does not capture the full range of variation in the Oto-Manguean languages. For example, in many Mixtec languages, laryngealization is understood to be a feature of the vowel rather than a consonant (e.g., Cortés, Mantenuto & Steffman 2023, Eischens 2022, Gerfen 1999, Gerfen & Baker 2005). Additionally, in many of these languages, the larngealization can appear either in the middle of the vowel, which Silverman (1997a) calls interrupted, or at the end of the vowel (e.g., Cortés, Mantenuto & Steffman 2023, Eischens 2022). This is directly in contrast to the implicational hierarchy, which states that if a language has interrupted laryngealization, it must also have postvocalic laryngealization and prevocalic laryngealization.

Not only is the violation of the implicational hierarchy the case in Mixtecan languages, but it is also the case in other branches of the Oto-Manguean languages. It is often the case that Zapotec languages have only interrupted and postvocalic vowels or just interrupted vowels (see Ariza-García 2018 for a typology of phonation in Zapotec languages). For example, Avelino (2004, 2010) argues that Yalálag Zapotec only has interrupted laryngealization as a vowel feature and postvocalic laryngealization with a laryngeal consonant. However, there is no prevocal laryngealization in the language.

This is in direct contrast to the implicational hierarchy of laryngeal complexity. This is also true for laryngealization in Santa Ana del Valle Zapotec (Esposito 2004, 2012).

6.3 Previous analyses of laryngeal complexity

Previous analyses of laryngeal complexity fall into two categories: (i) descriptive studies and (ii) instrumental studies. Most descriptive studies focus on describing the tone and voice quality patterns and how they interact. For example, Frazier (2013) describes the phonetic properties of tone and voice quality in Yucatec Mayan. In this study, Frazier describes how Yucatec Mayan is one of the few Mayan languages that has developed tonal contrasts. Additionally, it has a series of vowels that have a high tone with glottalization that variably surface as rearticulated vowels[1] or as a vowel with creaky voice. Frazier notes that these vowels show clear evidence of phasing between the tone and the voice quality for most speakers. In these vowels, the first portion of the vowel is always modal and is produced with the high tone. The second portion of the vowel is produced with creaky voice, which significantly obscures the pitch.

Not only is this the case in Yucatec Mayan, but it is also the case in languages that primarily have a phonation contrast with tone/pitch as a secondary cue like Danish

[1]This is very similar to rearticulated vowels in SLZ, where a vowel has a period of glottalization in the middle of the vowel that is often realized as a glottal stop. This is different from how Baird (2011) describes "broken" or "rearticulated" vowels in K'ichee, another Mayan language.

(Fischer-Jørgensen 1989, Grønnum, Vazquez-Larruscaín & Basbøll 2013, Peña 2022, 2024). In Danish, there is a phonation contrast between modal voice and a type of creaky voice that is called *stød*. Research has shown that *stød* is also associated with a secondary cue of a heightened $f0$ (Fischer-Jørgensen 1989, Grønnum, Vazquez-Larruscaín & Basbøll 2013). Peña (2022, 2024) showed that even though Danish is not traditionally classified as a laryngeally complex language it still shows evidence of phasing between the primary and secondary cues to stød, with the primary cue of phonation being produced in the second half of the syllablic rhyme and the secondary cue of pitch being made in the first half of the syllablic rhyme.

In contrast to descriptive studies, instrumental studies have focused on the acoustic properties of laryngeal complexity, mainly how laryngealization affects $f0$ and the harmonic structure of the vowel. For example, Garellek & Keating (2011) found that in Jalapa Mazatec, the laryngealization causes $f0$ perturbation in the first part of the vowel. Garellek & Keating conclude that this is evidence for the phasing between laryngealization and tone, as predicted by Silverman's 1997 theory of laryngeal complexity. Additionally, this same phenomenon of $f0$ perturbation is the case with laryngealization in some varieties of Trique, again showing that there is phasing between modal and nonmodal phonation (DiCanio 2012a). For these previous studies, the primary evidence for laryngeal complexity has been the perturbation of the $f0$ signal in the vowel portion affected by nonmodal phonation.

In recent studies, researchers have looked at other measures besides $f0$ to determine if phasing exists. For example, Weller et al. (2023a,b, 2024) have investigated laryngeal complexity in San Sebastián del Monte Mixtec using a combination of $f0$ and Strength of Excitation (SoE). This measure correlates with the strength of voicing. They found a clear phasing between modal and nonmodal phonation in the rearticulated vowels of the language.

However, these accounts offer only a limited window into the question of laryngeal complexity. Because there has been a focus on determining whether or not there are $f0$ perturbations in the signal, these previous studies have missed the opportunity to look at the full range of acoustic properties. Weller et al. (2023a,b, 2024) have done an excellent job in showing that looking at SoE, in addition to $f0$, we can better understand the phasing between tone and voice quality. However, looking at other acoustic properties of the signal is still necessary to understand laryngeal complexity fully.

The harmonic-to-noise ratio measure is one of those classes that promises to provide a better understanding. Harmonic-to-noise ratio measures are a class of measures that look at the ratio of the harmonic energy to the noise energy in the signal. This class of measures has been invaluable in determining whether there is aperiodicity in the signal (de Krom 1993, Ferrer Riesgo & Nöth 2020, Garellek 2019). It is well understood that aperiodicity is one of the defining characteristics of nonmodal phonation (Ladefoged & Maddieson 1996) and implies that harmonic-to-noise ratio measures can be used to determine whether there is aperiodicity in the signal and if there is a clear phasing between modal and nonmodal phonation.

6.4 Analysis of laryngeal complexity

6.4.1 Methods

6.4.1.1 Participants

This study uses data collected from 10 native speakers of SLZ during the summer of 2022. Participants were recruited from the community of Santiago Laxopa, Oaxaca, Mexico. All participants were native speakers of SLZ. The participants were between 18 and 60 years old and consisted of five males and five females.

6.4.1.2 Recordings

Participants were asked to perform a word list elicitation task consisting of 72 words. These words were selected to elicit the entire range of types of voice quality in SLZ, including modal voice, the two kinds of creaky (i.e., checked and rearticulated), and breathy voice. The words were selected based on previous research conducted as part of the Zapotec Language Project at the University of California, Santa Cruz (*Zapotec Language Project — University of California, Santa Cruz* 2022). Because participants were not literate in SLZ, the word list was prompted for them by asking them

"How do you say [word in Spanish]?" by myself and another researcher in Zapotec. Participants were asked to respond with the desired word in the carrier phrase *Shnia' X chonhe lhas* [ʃnːiaˀ X tʃone ɾas] "I say X three times." which was repeated three times. These utterances were recorded in a quiet environment using a Zoom H4n handheld digital recorder. The recordings were saved as 16-bit WAV files with a sampling rate of 44.1 kHz.

6.4.1.3 Acoustic measuring

The resulting audio files were then processed in Praat to isolate the vowel portion of each word. The onset of the vowel was set to the second glottal pulse after the onset, and the offset of the vowel was set to the last glottal pulse before the decrease in amplitude at the end of the vowel (Garellek 2020). The vowel was then extracted and saved as a separate file for analysis.

These vowels were fed into VoiceSauce (Shue et al. 2011) to generate the acoustic measures for the studies discussed in this dissertation. Because many acoustic measures are based on the fundamental frequency, this measure was calculated using the STRAIGHT algorithm from (Kawahara, Cheveigne & Patterson 1998) to estimate the fundamental frequency in milliseconds (ms) intervals. Once the fundamental frequency is calculated, VoiceSauce then uses an optimization function to locate the harmonics of the spectrum, finding their amplitudes.

VoiceSauce then uses the Snack Sound toolkit (Sjölander 2004) to find the frequencies and bandwidths of the first four formants, also at millisecond intervals. The amplitudes of the harmonics closest to these formant frequencies are located and treated as the amplitudes of the formants. These formant frequencies and bandwidths are used to correct the harmonic amplitudes for the filtering effects of the vocal tract, using Iseli, Shue & Alwan's 2007 extension of the method employed by Hanson (1997). Each vowel was measured across ten equal time intervals, resulting in 22890 data points in total. These measures were then z-scored by speaker to reduce the variation between speakers and provide a way to compare the different measures directly on similar scales.

6.4.1.4 Data processing

Data points with an absolute z-score value greater than three were considered outliers and excluded from the analysis. The Mahalanobis distance was calculated in the F1-F2 panel within each vowel category. Each data point with a Mahalanobis distance greater than six was considered an outlier and excluded from the analysis. Using the Mahalanobis distance allows us to compare the data points to the mean of the F1-F2 panel for each vowel category. The larger the Mahalanobis distance, the more deviant the data point is from the mean, which implies that the data point was improperly tracked. This methodology is comparable to what was done in Seyfarth & Garellek (2018), Chai & Ye (2022), and Garellek & Esposito (2023).

Energy was excluded if it had a zero value and then logarithmically transformed to normalize its right-skewed distribution. Afterward, the resulting log-transformed energy was z-scored, and any data point with a z-score greater than three was excluded. This outlier removal resulted in 1918 data points being removed.

All data points were then z-scored by speaker to reduce the variation between speakers and provide a way to compare the different measures directly on the same scale.

Residual H1* for the remaining data points following Chai & Garellek (2022). First, a linear mixed effects model was generated with the z-scored H1* as the response variable and the z-scored energy as the fixed effect. The uncorrelated interaction of the z-scored energy by speaker was treated as random. The energy factor resulting from this linear mixed-effects model was extracted. Finally, the z-scored H1* had the product of the z-scored energy and energy factor subtracted from it to produce the residual H1* measure.

As mentioned in Chapter 3, the distribution of tone and phonation was not equal in all combinations of the two; certain combinations of tone and phonation were over-represented in the data set. As seen in Table 3.1, which is repeated here as Table 6.2, the tone and phonation distribution is unequal across all combinations.

Modal voice was the only phonation that occurred across all possible tones, and low tone was the only tone that occurred across all possible phonations. This imbalance in the data is a problem because it can lead to biased results when analyzing

Table 6.2: Distribution of the number of syllables containing the combination of tone and voice quality in the wordlist.

	High	Mid	Low	Rising	Falling
Modal	14	9	15	2	10
Breathy	—	—	11	—	2
Checked	1	—	9	—	—
Rearticulated	1	—	4	—	4

laryngeal complexity. Only low tone phonations were used to analyze laryngeal complexity to correct for this potential confound. By limiting the analysis to only low tone, we ensure that the dataset accurately captures the effects of the interaction of phonation on a specific tone. This will allow us to ignore the potential confounding factor of how creaky voice is not uniformly produced when it appears with different tones (Keating, Garellek & Kreiman 2015).

6.4.1.5 Statistical analysis

Generalized additive mixed models (GAMM; Hastie & Tibshirani 1986, Wood 2017) are a type of statistical model similar to generalized linear mixed models, but allow for smoothing functions of continuous predictors. This allows for modeling nonlinear relationships between the predictors and the response variable, especially over time. Using GAMMs will enable us to capture the nonlinear relationships between the predictors and the response variable, which is essential for understanding how phonation affects the acoustic measures over time.

The GAMMs were fitted to the data using the `mgcv` package in R (Wood 2017) and plotted using the `tidygam` package (Coretta 2024). Three GAMMs were fitted for each acoustic measure: $f0$, HNR < 1500 Hz, and SoE as the response variable. These three measures were chosen based on previous research, with respect to $f0$, and the results of the MDS and Random Forest analyses. Phonation was treated as a fixed effect to model the main effect of phonation on the acoustic measures. A smooth term for time (i.e., measurement number) was included to model the nonlinear relationship between time. A second smooth term was included to allow time to vary by phonation type. Random effects for speaker and the interaction between speaker and phonation were included to account for variability in the data due to individual differences between speakers and individual differences in how they produce the phonation types. A tensor product interaction between time and repetition was also included as a random effect to account for the variability in the data due to the different repetitions the speakers were asked to complete.

By modeling the GAMMs in this way, each model can capture four things: (i) the main effect of phonation on the acoustic measures, (ii) the nonlinear relationships in time both overall and specific for each phonation type, (iii) speaker variability and its interaction with phonation, and (iv) the variability in the data due to the different repetitions the speakers were asked to complete. This allows for a more accurate modeling of the data and a better understanding of how phonation affects the acoustic measures under consideration.

6.5 Results

This section presents the results of the three GAMM analyses. I present the GAMM model fit and the model difference plot for each measure. The model fit plot shows the GAMM's predicted values for each phonation over time. The model difference plot shows the difference between the modal phonation and each of the other phonations. This allows us to see how each nonmodal voice quality differs from modal phonation over time.

6.5.1 Fundamental frequency

Figure 6.4 shows the model fit for $f0$. Each line represents the predicted values of the GAMM for each of the phonations over time. The shaded area represents the 95% confidence interval. The model's fit shows that the modal voice (the black solid line) has the highest predicted values for $f0$. The nonmodal phonations have lower predicted values, with breathy voice (the orange dashed line) having the lowest predicted values. Checked and rearticulated voice (the blue and green dashed lines, respectively) have predicted values similar to and lie between modal and breathy voices. Beginning at measurement number 8, we see that checked and rearticulated voices begin to separate. For checked voice, the predicted values increase, while for rearticulated voice, the predicted values decrease until measurement number 10.

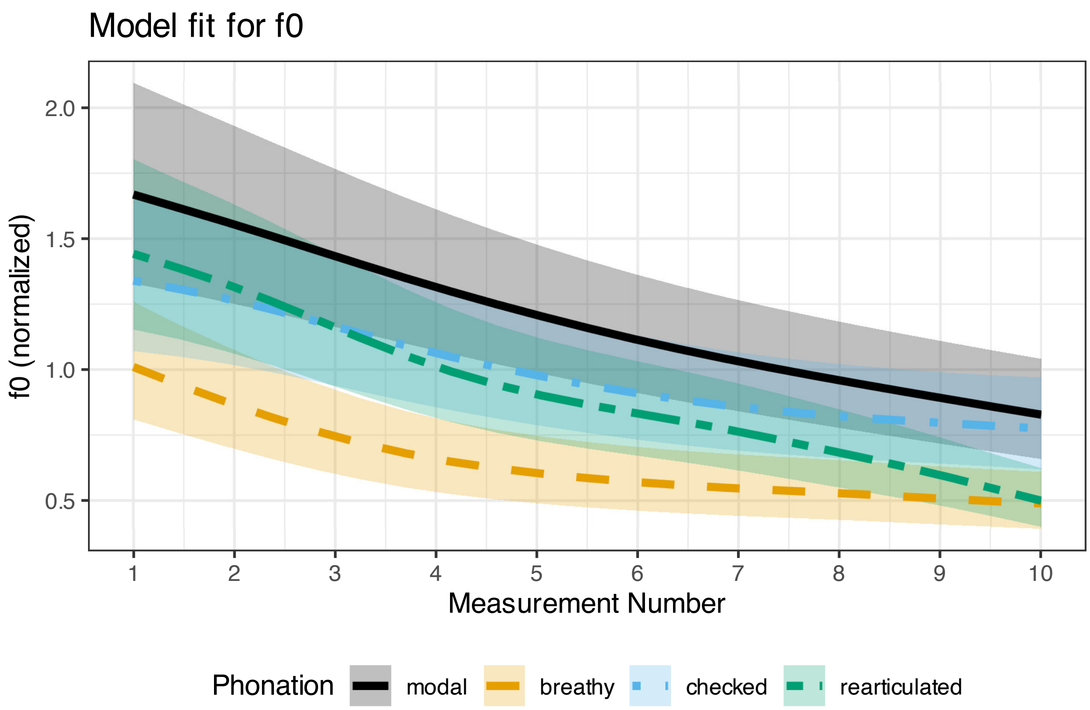


Figure 6.4: Model fit for $f0$. Each line represents the predicted values from the GAMM for each phonation over time. The shaded area represents the 95% confidence interval.

The observations from the model fit are only part of the story when evaluating GAMMs. The other evidence that we need is whether the differences we observe in the model's fit are significant. This is done through difference plots that compare the values for a given comparison (in this case modal versus one of the nonmodal phonations) and return whether or not the difference is significant. In these plots, the red line indicates that the difference between modal and the other phonation types is significant. The blue line indicates that the difference between modal and nonmodal phonation is insignificant. This evaluation method is the most common method for evaluating GAMMs (see Sóskuthy 2021 for a discussion about assessing GAMMs).

Figure 6.5 shows the corresponding difference plot for $f0$. The leftmost panel shows the difference plot between modal and breathy voice. The red line indicates that the differences observed between modal and breathy voice in Figure 6.4 are significant throughout the vowel length. The center panel shows that the differences between modal and checked voice are insignificant at any point during the vowel, and in terms of $f0$, there is no difference between these two phonations. The rightmost panel shows that the difference between modal and rearticulated voice is also not significant for the first part of the vowel. However, beginning at measurement number 7, we see that the differences are significant.

This suggests that in terms of $f0$, breathy voice significantly differs from modal voice throughout the vowel. This is consistent with the descriptions for breathy voice cross-linguistically, where one of its acoustic cues is a lower $f0$ than modal vowels (e.g., Hillenbrand & Houde 1996a). However, rearticulated vowels are only significantly different from modal vowels for the final third of the vowel. The results suggest that in terms of $f0$ perturbation, only rearticulated vowels show phasing.

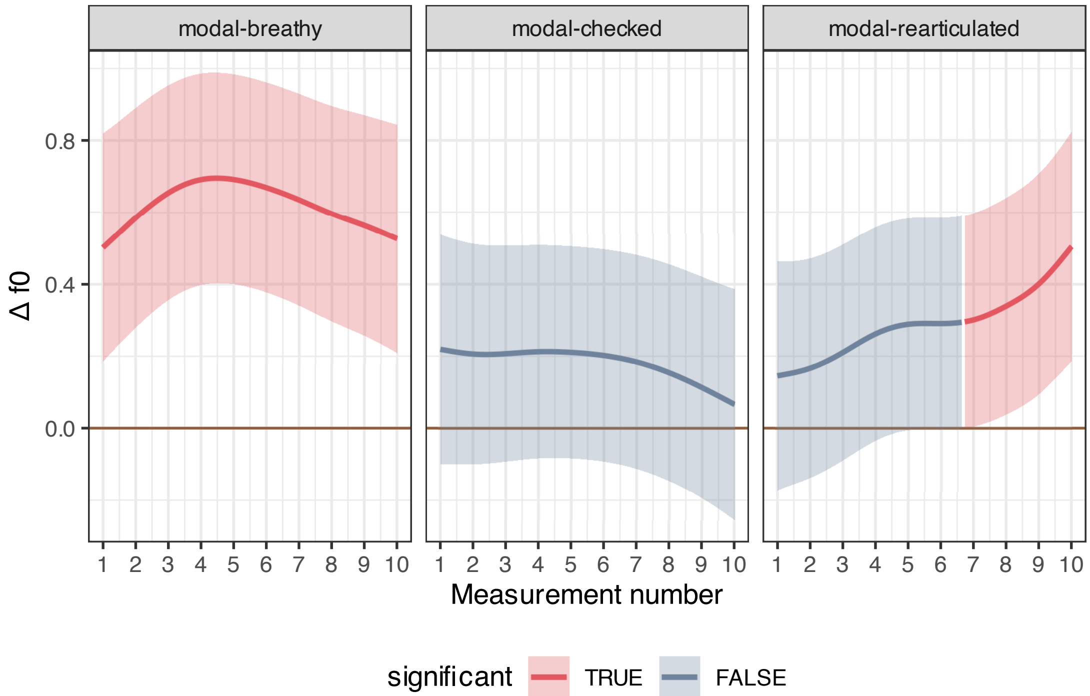


Figure 6.5: Plot of the difference between modal and each of the nonmodal phonation types.

6.5.2 Harmonic-to-noise ratio

The model fit for HNR < 1500 Hz is shown in Figure 6.6. In the fit of the model, we see that modal voice has the highest predicted values for HNR < 1500 Hz, which is consistent with previous research that has shown that modal voice has higher HNR than nonmodal phonation types (e.g., Blankenship 1997, 2002, de Krom 1993, Garellek 2012, 2019, Gerratt & Kreiman 2001). Additionally, we see that the three nonmodal phonations are all very similar to each other. However, checked vowels show a dip at the end of the vowel, indicating more noise in the signal.

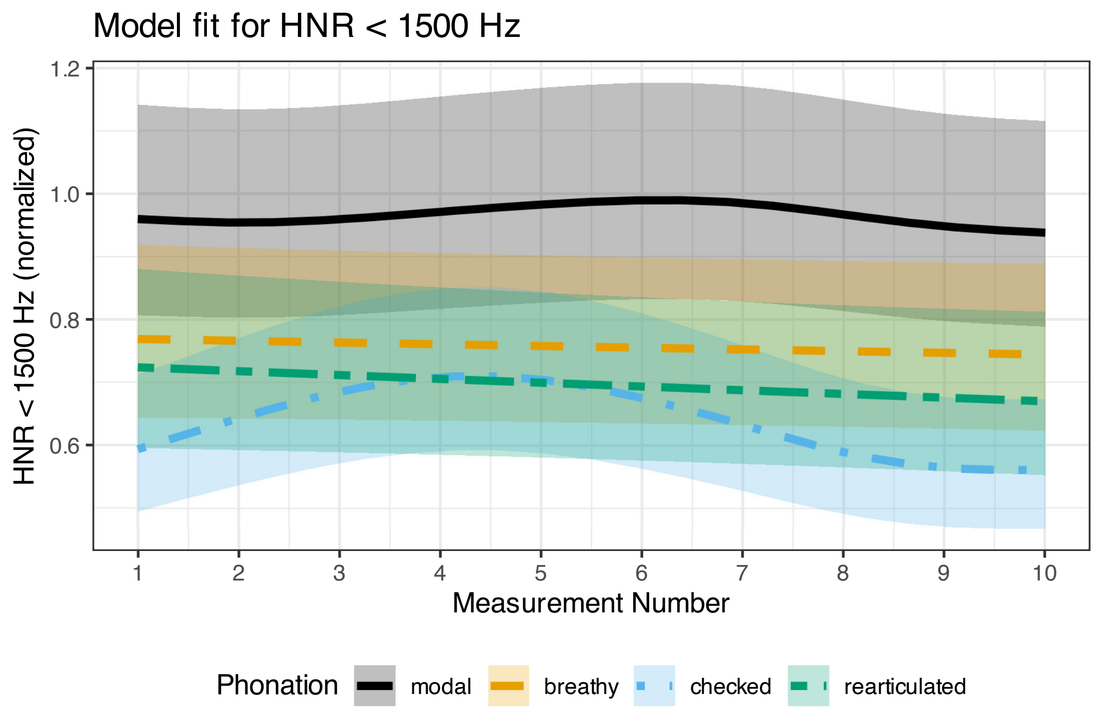


Figure 6.6: Model fit for HNR. Each line represents the predicted values from the GAMM for each phonation over time. The shaded area represents the 95% confidence interval.

Figure 6.7 shows the difference plots for each nonmodal phonation. Similarly to Figure 6.5, the leftmost panel shows the difference between modal and breathy voice. This difference is significant from measurement number 4 to halfway between measurements 8 and 9.

The center and leftmost panels show that the difference between modal and non-modal phonations is significant across the entire length of the vowel. This indicates that checked and rearticulated vowels are associated with greater aperiodicity throughout the vowel length.

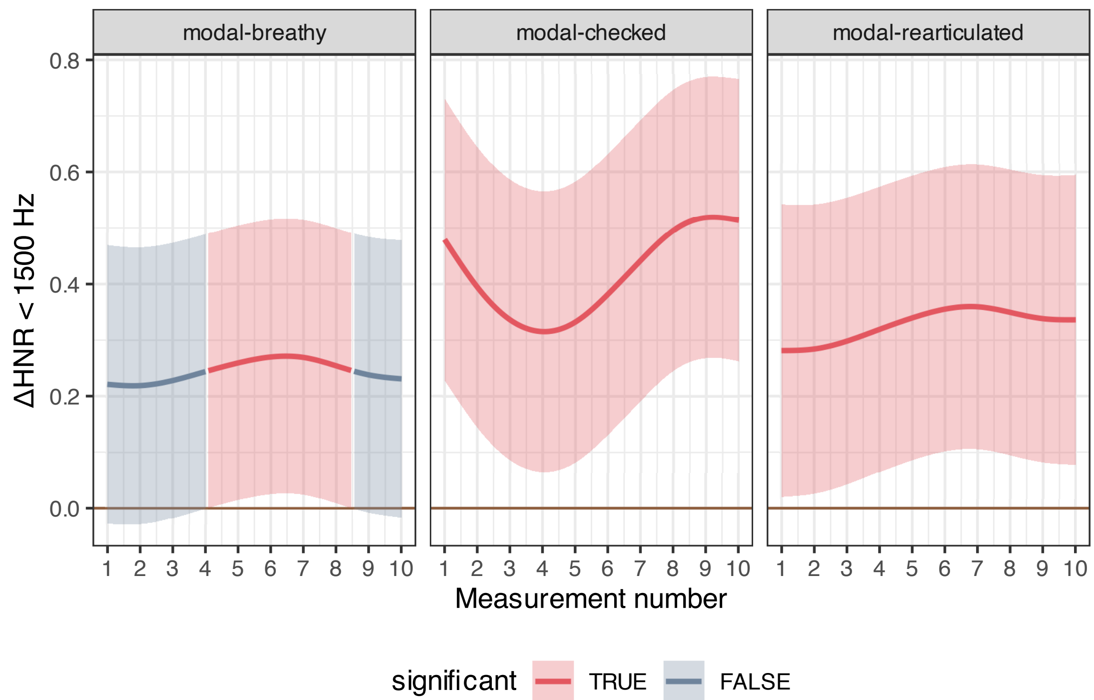


Figure 6.7: Plot of the difference between modal and each of the nonmodal phonation types.

From HNR < 1500 Hz, we see that all three nonmodal phonations have more noise in their signal than modal vowels. As I discussed in Section 5.7.6, HNR measures are typically associated with aperiodicity in the signal. If there is phasing between modal phonation and nonmodal phonation as predicted by Silverman (1997a), we would expect to see that these phonations should have periods where they exhibit more modal-like properties or are not significantly different from modal phonation, which is what we observe in the breathy voice. Breathy vowels show a period of greater aperiodicity than modal vowels in the second half of the vowel, especially if we ignore the interval between 9 and 10 due to coarticulation with the surrounding consonants. Furthermore, this indicates that there is a phasing relationship in breathy vowels where the first half of the vowel is modal and the second half is breathy.

6.5.3 Strength of excitation

As discussed in Section 5.7.7, SoE is an acoustic measure that correlates with the strength of voice. Figure 6.8 shows the model's fit for SoE. In the model fit, we see that modal voice has the highest predicted values for SoE, which is consistent with Garellek et al. (2021) where modal vowels are associated with a higher SoE than nonmodal phonation because of the greater strength in voicing.

Breathy voice is identical to the modal voice until just after measurement number 2, and then begins to decrease but becomes identical to the modal vowels shortly after measurement number 9. Checked vowels are similar to breathy vowels in that they are identical to modal vowels until just after measurement number 2. However, the checked vowels sharply decreased SoE throughout the vowels until measurement number 10 was reached. However, rearticulated vowels begin lower than the other three phonations but increase in SoE until they are identical to the modal voice between measurements 8 and 9.

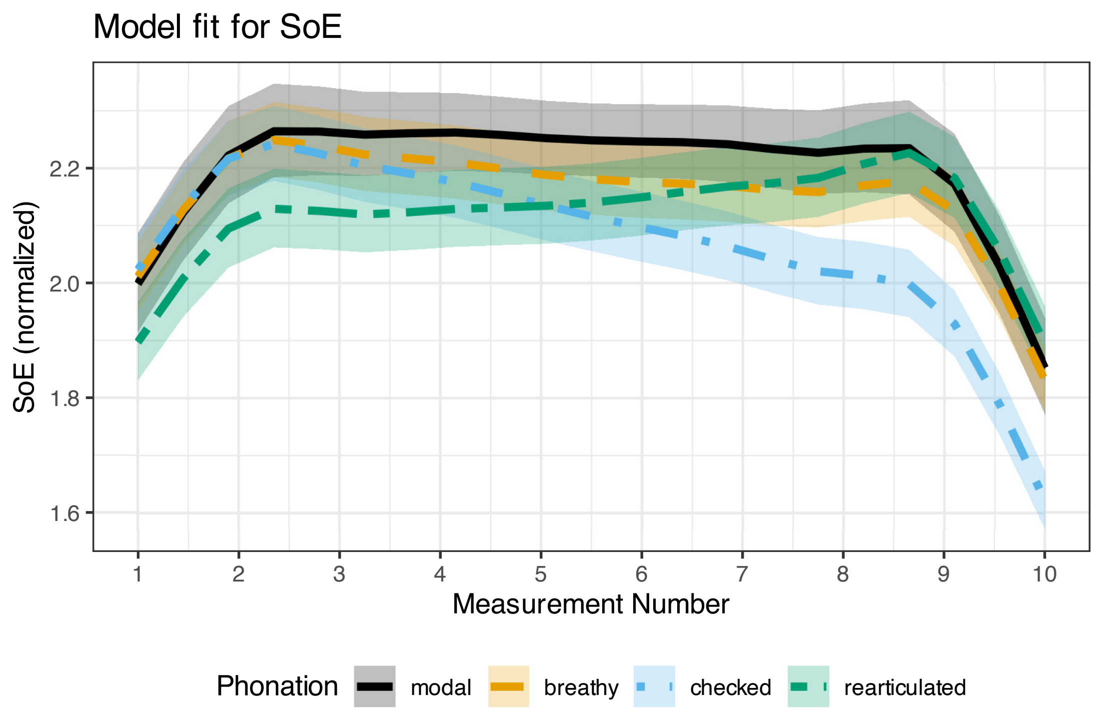


Figure 6.8: Model fit for SoE. Each line represents the predicted values from the GAMM for each phonation over time. The shaded area represents the 95% confidence interval.

In evaluating whether or not the observed differences from the model fit are significant, we can again turn to difference plots to help us determine whether or not the differences were significant. Figure 6.9 shows the difference plots for each nonmodal phonation.

The leftmost panel shows the difference between modal and breathy voice. The differences between modal and breathy vowels were only significant between measurement numbers 4 and 8. This is similar to what we observed in Figure 6.7 for HNR < 1500 Hz. We can ignore the interval between measurements 9 and 10 for breathy vowels because of the coarticulation with the surrounding consonants.

The center panel shows the difference between modal and checked voice. We observe that checked vowels are significantly different from modal vowels from measurement number 3 to the end of the vowel. However, we can consider the interval between measurements 9 and 10 because checked vowels have a phontactic restriction in the nominal domain that prevents them from occurring in closed syllables. The only time they can accure in a closed syllable is when a morpheme or clitic boundary intervenes, such as in the word *beku'=nh* [bekuˀ=ŋ] 'the dog', where the checked vowel occurs at a morpheme boundary.

The rightmost panel shows the difference between modal and rearticulated voice. The differences between modal and rearticulated vowels were only significant between measurement numbers 1 and 7.

The SoE measure shows that the strength of voicing is not drastically different between modal and nonmodal phonation. This suggests that nonmodal phonations are not as strongly articulated in SLZ when compared to nonmodal phonations in other languages (e.g., Garellek et al. 2021, Garellek, López-Francisco & Amith 2025, Weller et al. 2023a,b, 2024). This weak articulation was predicted by Silverman and is consistent with the idea that laryngealization is not as strongly articulated in Zapotec languages as it is in other Oto-Manguean languages (Herrera Zendejas 2000).

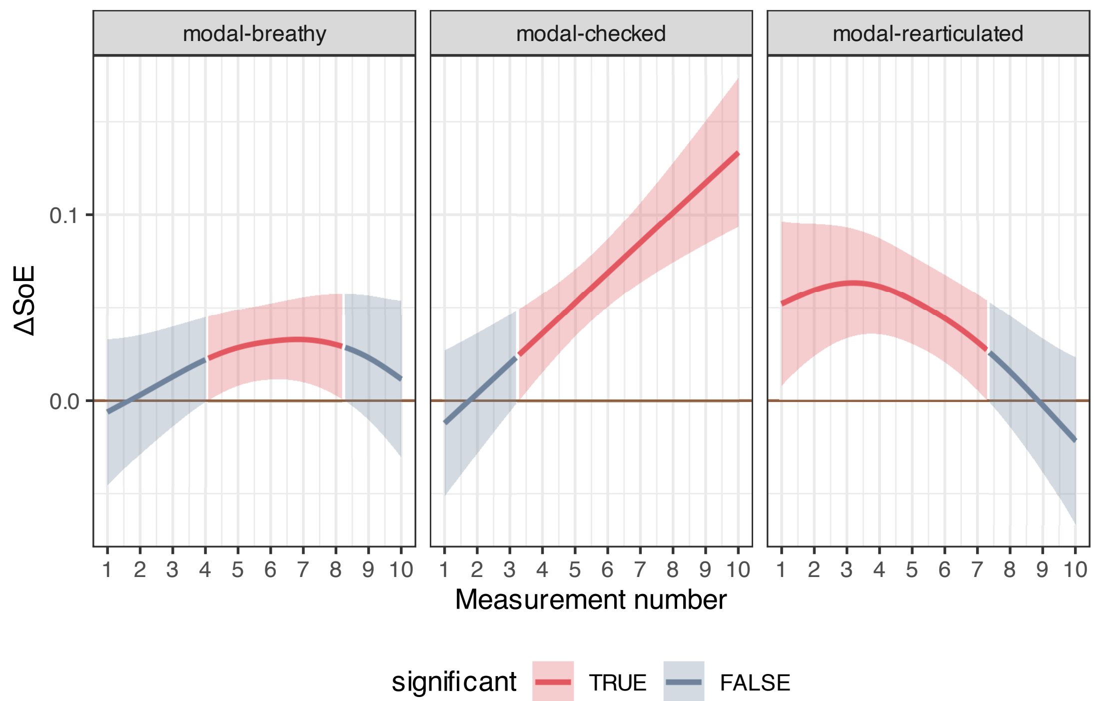


Figure 6.9: Plot of the difference between modal and each of the nonmodal phonation types.

6.6 Discussion

As an Oto-Manguean language, it is not surprising that SLZ exhibits laryngeal complexity. As a Zapotec language, SLZ is expected to offer counterarguments to some of the claims made by Silverman (1997a,b) regarding the implicational hierarchy. The reason for this is that Zapotec languages are known to typically only have interrupted and postvocalic laryngealization (Arellanes Arellanes 2009, 2010, Ariza-García 2018, Avelino 2004, 2010, Esposito 2004, 2010b). This means that Zapotec languages offer a strong counterargument against the implicational hierarchy because they lack any prevocal laryngealization. Furthermore, according to Herrera Zendejas (2000), Zapotec languages do not have a phasing relationship between modal and nonmodal phonation. These languages have achieved "equilibrium" between tone and laryngealization.

The results of this study show that these expectations for Zapotec languages do not hold in SLZ. Instead, the results, when taken in aggregate, show that SLZ exhibits a phasing relationship between modal and nonmodal phonation in support of Silverman's 1997, 1997 claims about laryngeal complexity. This is supported by the way breathy and checked voice shows a phasing pattern with modal phonation preceding a period of laryngealization, or using Silverman's 1997 vocabulary apostvocalicc laryngealization. Rearticulated vowels, while on the surface appear to belong to the interrupted, do not show this. Instead, these vowels begin with nonmodal phonation and then end with modal phonation. This classifies them as belonging to the Silverman's 1997 prevocal laryngealization.

The pattern of nonmodal phonation in prevocalic laryngealization agrees with the implicational hierarchy for laryngeal complexity. As discussed above in Section 6.2.2, the implicational hierarchy states that if a language has laryngealization, it should follow that if it has interrupted laryngealization, it should also have postvocalic laryngealization. If it has postvocalic laryngealization, it should also have prevocalic laryngealization. This can be seen in the table presented above in Table 6.1, which is repeated here as Table 6.3.

Table 6.3: Implicational hierarchy of laryngeal complexity. The symbols h and ʔ represent laryngealization. The symbol V represents where the modal vowel is located in relation to the laryngealization. Modified from Silverman (1997a).

Language	**Prevocalic**	**Postvocalic**	**Interrupted**
Jalapa Mazatec	hV˥, ʔV˥	–	–
Comaltepec Chinantec	hV˥, ʔV˥	Vh˥, Vʔ˥	–
Copala Trique	hV˥, ʔV˥	Vh˥, Vʔ˥	VhV˥, VʔV˥

As stated previously, the results of this study support this implicational hierarchy for SLZ. SLZ has both prevocal laryngealization with its rearticulated vowels and postvocalic laryngealization with its checked vowels and breathy vowels. However, this does not mean that this implicational hierarchy does not need further investigation. As discussed in Section 6.2.2, the implicational hierarchy is based on a few languages and does not consider the full range of laryngeal complexity. While the implicational hierarchy is a good starting point, it must be tested against more languages to see if it holds.

Additionally, the implicational hierarchy needs to be refined to clarify whether it concerns just laryngeal consonants, laryngealization on vowels, or both. It is also unclear whether the implicational hierarchy is a universal property of laryngeal complexity or if it is just a property of the languages studied so far. Additionally, the implicational hierarchy needs to be further tested to see if it holds even against languages that are not laryngeally complex, like Oto-Manguean languages. This is important because it will help us better understand the nature of cross-linguistic interactions between tone and phonation.

The results of the study also suggest that, in keeping with Silverman's 1997 claims, laryngealization in SLZ appears to be more weakly implemented as evidenced by the relatively narrow difference in SoE between modal and nonmodal phonation. This is only possible if the amplitude of voicing is not entirely different between modal and nonmodal phonation.

This study also shows that to get a complete picture of laryngeal complexity in a language, we need to look at more than just $f0$ perturbation. We can better understand the phasing relationship between modal and nonmodal phonation by looking at harmonics-to-noise ratio measures and measures of voicing in addition to $f0$. This is important because it allows us to see how these different measures interact and how they can be used to understand laryngeal complexity in a language better. This is also necessary because voice quality is a multidimensional phenomenon (e.g., Kreiman et al. 2014, 2021). If we only focus on one acoustic phenomenon, such as $f0$ perturbation, we will miss out on the full range of acoustic properties in the signal. This is especially true for laryngeal complexity, with multiple dimensions to consider. By looking at a broader range of acoustic properties, we can better understand how these different measures interact and how they can be used to better understand laryngeal complexity in a language.

6.7 Conclusion

In conclusion, the results of this study show that Silverman's 1997 claims about laryngeal complexity continue to hold after almost 30 years. SLZ exhibits a phasing relationship between modal and nonmodal phonation in its laryngeally complex vowels. Evidence for this phasing was only possible by looking for $f0$ perturbations and using acoustic measures for harmonics-to-noise and Strength of Excitation. This, further, shows that laryngeal complexity is also a multidimensional phenomenon and needs to be carefully considered in light of multiple acoustic measures. If only one acoustic measure is used, such as $f0$ perturbation, we will miss out on the full range of acoustic properties in the signal.

This study shows that the implicational hierarchy for laryngeal complexity holds for SLZ. With its three phonation types, it exhibits both prevocalic and postvocalic laryngealization. This also indicates that interrupted laryngealization is not present in this language. This is important because it shows that the implicational hierarchy is a valid way to think about laryngeal complexity and how it interacts with tone and phonation. However, further work is needed to establish whether this hierarchy holds for all tone and phonation interactions or only for laryngeally complex languages.

Chapter 7

Conclusion

This dissertation investigated the acoustics of voice quality in Santiago Laxopa Zapotec and its interactions with tone. There were four main questions that this dissertation addressed: (i) does the recently proposed residual H1* acoustic measure more effective captures the phonation contrasts between Santiago Laxopa Zapotec's four phonations than the traditional H1*−H2* acoustic measure; (ii) how is the acoustic landscape of voice quality in Santiago Laxopa Zapotec structured; (iii) which acoustic measures most effectively capture and classify the voice quality contrasts; (iv) and how do those acoustic measures help explain Santiago Laxopa Zapotec's laryngeal complexity?

7.1 Summary of findings

This dissertation showed in Chapter 3, that Chai & Garellek's (2022) residual H1* is a more effective acoustic measure for capturing the phonation contrasts in Santiago Laxopa Zapotec than the traditional H1*−H2* measure. Furthermore, this measure was also shown to play an important role in defining the phonation's acoustic landscape in Santiago Laxopa Zapotec, as demonstrated in Chapter 4. Additionally, Chapter 5 showed that residual H1* also was one of the important acoustic measures for classifying the phonation contrasts in Santiago Laxopa Zapotec according to a Random Forest model. These results contribute to validating Chai & Garellek's (2022) residual H1* measure as a reliable acoustic measure for phonation contrasts as does recent work by Chai, Fernández & Mendez (2023) and Garellek, López-Francisco & Amith (2025).

Chapter 4 demonstrated that SLZ's phonation largely occupies a three-dimensional acoustic landscape defined by spectral slope and harmonic structure/noise. These findings are consistent with the findings from Keating et al. (2023). In both studies, these first two dimensions are correlated with spectral slope and harmonic structure. The study presented in Chapter 4 showed that SLZ additionally needed a third dimension, also correlated with spectral slope, to fully capture phonation's dimensionality. The results of these studies suggest that phonation is primarily characterized by just spectral slope and harmonic structure. Furthermore, the results of this study suggest that when higher dimensionality is needed, spectral slope or harmonic structure/noise are also correlated with these dimensions.

Chapter 5 demonstrated that SLZ's phonation relies on a small set of acoustic measures to effectively classify the phonation contrasts. The Random Forest model presented in this chapter showed that many of the same measures that were correlated with the dimensions of the acoustic landscape were also important for classification. The results of the Random Forest analysis showed that the most important acoustic measures for classification were: (i) duration, (ii) A1*, (iii) H1*−A1*, (iv) residual H1*, (v) HNR $<$ 1500 Hz, and (vi) Strength of Excitation.These findings suggest that the phonation contrasts in SLZ are not only characterized by spectral slope and harmonic structure/noise, but also by the duration of the vowel and the strength of excitation. These results contribute to our understanding of how phonation contrasts can be effectively classified using a small set of acoustic measures.

The previous chapters were important in establishing which acoustic measures needed to be consulted for Chapter 6's analysis of how phonation interacts with tone in SLZ. This chapter investigated these interaction through generalized additive mixed model analysis of $f0$, HNR $<$ 1500 Hz, and Strength of Excitation; revealing two facts about how laryngeal complexity manifests itself in SLZ. First, the results of this chapter showed that nonmodal phonation is not strongly articulated in SLZ. This was demonstrated by the very small, though statistically significant, differences in

Strength of Excitation between the modal and nonmodal phonations. If nonmodal phonation were strongly articulated in SLZ, we would expect to see larger differences in Strength of Excitation between the modal and nonmodal phonations. This first fact confirms Herrera Zendejas's (2000) claim that Zapotec languages do not articulate their nonmodal phonation strongly. Second, the results of this chapter also showed that contrary to Herrera Zendejas's (2000) claims that Zapotec languages lack phasing of modal and nonmodal phonation, SLZ does exhibit phasing of modal and nonmodal phonation. These results demonstrate that Silverman (1997a,b) was correct in their assertion that laryngeal complexity primarily exhibits itself through phasing of modal and nonmodal phonation. The results additionally showed that despite lacking strong articulation of nonmodal phonation, SLZ does exhibit phasing of modal and nonmodal phonation which runs counter to Silverman's claims that either you exhibit phasing with strong articulation or you lack phasing and weakly articulate nonmodal phonation.

These finding contribute to our understanding of phonation and laryngeal complexity; however, these findings also raise several questions important questions that require further investigation. These questions will be discussed in the following sections.

7.2 Future directions

Despite the growing number of perceptual studies on phonation or tone, perception research into laryngeal complexity is lacking. This dissertation has shown shown that phonation is primarily analyzed through spectral slope and the harmonic structure/noise of the signal, which is correlated to the open quotient of the glottis and the periodicity of the signal. However, it is still unclear how these tones and phonation contrasts are perceived by speakers of SLZ.

Research has shown that even though these acoustic measures are effective for classifying phonation contrasts, they do not always correlate with how speakers perceive these contrasts. For example, Hillenbrand & Houde (1996b) and Gerfen & Baker (2005) found that speakers of American English and Mixtec, respectively, primarily relied on changes in $f0$ and amplitude to perceive intervocalic laryngealization. Furthermore, checked vowels have as their percepts a short duration and falling $f0$ contour are more important cues than late-phased glottalization (Brunelle & Finkeldey 2011, Chai 2022, 2025, Garellek et al. 2013). Klatt & Klatt (1990) found that aspiration noise was the important cue for distinguishing breathy and modal phonation apart in English; while Hillenbrand & Houde (1996a) found that listeners primarily relied on the cepstral peak prominence to distinguish breathy and modal phonation. Because of phonation's multidimensionality, it is likely that listeners may rely on different acoustic cues to perceive phonation contrasts (Esposito 2010a, Kreiman et al. 2014, 2021).

Besides Gerfen & Baker (2005), these studies have primarily investigated languages that lack laryngeal complexity. Therefore, it is unclear how speakers of SLZ perceive the phonation contrasts in their language. This is an important question to address because it will help us understand how phonation contrasts are perceived in languages with laryngeal complexity.

Bibliography

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Adler, Jeff, Steven Foley, Jed Pizarro-Guevara, Kelsey Sasaki & Maziar Toosarvandani. 2018. The derivation of verb initiality in Santiago Laxopa Zapotec. In Jason Merchant, Line Mikkelsen, Deniz Rudin & Kelsey Sasaki (eds.), *A reasonable way to proceed: Essays in honor of Jim McCloskey*, 31–49. Santa Cruz, Berkeley, Chicago: University of California.

Adler, Jeff & Maho Morimoto. 2016. Acoustics of phonation types and tones in Santiago Laxopa Zapotec. *The Journal of the Acoustical Society of America* 140(4). 3109–3109. https://doi.org/10.1121/1.4969713.

Arellanes Arellanes, Francisco. 2009. *El sistema fonológico y las propiedades fonéticas del zapoteco de San Pablo Güilá: descripción y análisis formal*. Mexico City, Mexico: El Colegio de México Doctoral thesis.

Arellanes Arellanes, Francisco. 2010. Dos ‘Grados’ De Laringización Con Pertinencia Fonológica En El Zapoteco De San Pablo Güilá. In Esther Herrera Zendejas (ed.), *Entre cuerdas y velo*, 1st edn., vol. 9 (Estudios Fonológicos de Lenguas Otomangues), 85–122. Mexico City, Mexico: El Colegio de México. https://doi.org/10.2307/j.ctv6jmx5b.7.

Ariza-García, Andrea. 2018. Phonation types and tones in Zapotec languages: A synchronic comparison. *Acta Linguistica Petropolitana* XIV(2). 485–516. https://doi.org/10.30842/alp2306573714220.

Arras, Kai Oliver. 1998. *An Introduction to Error Propagation: Derivation, Meaning and Examples of Equation CY= FX CX FXT*. Technical Report EPFL-ASL-TR-98-01 R3. Lausanne, Switzerland: Swiss Federal Institute of Technology Lausanne (EPFL).

Avelino, Heriberto. 2004. *Topics in Yalálag Zapotec, with particular reference to its phonetic structures*. Los Angeles, CA: University of California, Los Angeles dissertation. 315 pp.

Avelino, Heriberto. 2010. Acoustic and Electroglottographic Analyses of Nonpathological, Nonmodal Phonation. *Journal of Voice* 24(3). 270–280. https://doi.org/10.1016/j.jvoice.2008.10.002.

Baird, Brandon. 2011. Phonetic and phonological realizations of ’broken glottal’ vowels in K’ichee’. In Kirill Shklovsky, Pedro Mateo Pedro & Jessica Coon (eds.), *Pro-*

ceedings of FAMLi 1: Formal Approaches to Mayan Linguistics (MIT Working Papers in Linguistics 63). Cambridge, MA: MIT Press.

Barzilai, Maya L. & Katherine J. Riestenberg. 2021. Context-dependent phonetic enhancement of a phonation contrast in San Pablo Macuiltianguis Zapotec. *Glossa: a journal of general linguistics* 6(1). 1–36. https://doi.org/10.5334/gjgl.959.

Beam de Azcona, Rosemary G. 2007. Problems in Zapotec Tone Reconstruction. *Annual Meeting of the Berkeley Linguistics Society* 33(2). 3. https://doi.org/10.3765/bls.v33i2.3497.

Bennett, Ryan. 2016. Mayan phonology. *Language and Linguistics Compass* 10(10). 469–514. https://doi.org/10.1111/lnc3.12148.

Bhattacharyya, A. 1943. On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society* 35. 99–110.

Bladon, Anthony. 1986. Phonetics for hearers. In Graham McGregor (ed.), *Language for Hearers*, 1. ed (Language and Communication Library 8), 1–24. Oxford: Pergamon Press.

Blankenship, Barbara. 1997. *The time course of breathiness and laryngealization in vowels*. Los Angeles, CA: University of California, Los Angeles dissertation.

Blankenship, Barbara. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30(2). 163–191. https://doi.org/10.1006/jpho.2001.0155.

Blevins, Juliette & Andrew Garrett. 1993. The Evolution of Ponapeic Nasal Substitution. *Oceanic Linguistics* 32(2). 199–236. https://doi.org/10.2307/3623193.

Boehmke, Brad & Brandon M. Greenwell. 2019. *Hands-On Machine Learning with R*. New York: Chapman and Hall/CRC. 484 pp. https://doi.org/10.1201/9780367816377.

Breiman, Leo. 1996. Bagging predictors. *Machine Learning* 24(2). 123–140. https://doi.org/10.1007/BF00058655.

Breiman, Leo. 2001. Random Forests. *Machine Learning* 45(1). 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen & Charles J. Stone. 1986. *Classification and regression trees*. Boca Raton, Fla.: Taylor and Francis. 358 pp.

Brinkerhoff, Mykel Loren, John Duff & Maya Wax Cavallaro. 2021. Downstep in Santiago Laxopa Zapotec and the prosodic typology of VSO languages. In *Manchester Phonology Meeting*. Manchester, England.

Brinkerhoff, Mykel Loren, John Duff & Maya Wax Cavallaro. 2022. Tonal patterns and their restrictions in Santago Laxopa Zapotec.

Brinkerhoff, Mykel Loren & Grant McGuire. 2025. Using residual H1* for voice quality research. *JASA Express Letters* 5(2). 025501. https://doi.org/10.1121/10.0035881.

Brunelle, Marc. 2005. *Register in Eastern Cham: Phonological, phonetic and sociolinguistic approaches*. Ithaca, NY: Cornell University dissertation. 355 pp.

Brunelle, Marc. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37(1). 79–96. https://doi.org/10.1016/j.wocn.2008.09.003.

Brunelle, Marc. 2012. Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham. *The Journal of the Acoustical Society of America* 131(4). 3088–3102. https://doi.org/10.1121/1.3693651.

Brunelle, Marc & Joshua Finkeldey. 2011. Tone perception in Sgaw Karen. In *ICPhS*, 372–375.

Brunelle, Marc & James Kirby. 2016. Tone and Phonation in Southeast Asian Languages. *Language and Linguistics Compass* 10(4). 191–207. https://doi.org/10.1111/lnc3.12182.

Burnham, Kenneth P. & David R. Anderson (eds.). 2004a. *Model Selection and Multimodel Inference*. New York, NY: Springer. https://doi.org/10.1007/b97636.

Burnham, Kenneth P. & David R. Anderson. 2004b. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research* 33(2). 261–304. https://doi.org/10.1177/0049124104268644.

Burnham, Kenneth P., David R. Anderson & Kathryn P. Huyvaert. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65(1). 23–35. https://doi.org/10.1007/s00265-010-1029-6.

Butler H., Inez M. 1997. *Diccionario zapoteco de Yatzachi: Yatzachi el Bajo, Yatzachi el Alto, Oaxaca.* 1. ed (Serie de vocabularios y diccionarios indígenas ”Mariano Silva y Aceves” no. 37). Tucson, AZ: Instituto Lingüístico de Verano. 528 pp.

Campbell, Eric W. 2011. Zenzontepec Chatino aspect morphology and Zapotecan verb classes. *International Journal of American Linguistics* 77(2). 219–246.

Campbell, Eric W. 2014. *Aspects of the Phonology and Morphology of Zenzontepec Chatino, a Zapotecan Language of Oaxaca, Mexico.* Austin, TX: University of Texas at Austin dissertation.

Campbell, Eric W. 2017a. Otomanguean historical linguistics: Exploring the subgroups. *Language and Linguistics Compass* 11(7). e12244. https://doi.org/10.1111/lnc3.12244.

Campbell, Eric W. 2017b. Otomanguean historical linguistics: Past, present, and prospects for the future. *Language and Linguistics Compass* 11(4). e12240. https://doi.org/10.1111/lnc3.12240.

Chai, Yuan. 2022. *Phonetics and phonology of checked phonation, syllables, and tones*. San Diego, CA: University of California San Diego dissertation.

Chai, Yuan. 2025. Perception of checked and rearticulated phonations: Effect of duration and glottalization phasing.

Chai, Yuan, Adrián Fernández & Briseida Mendez. 2023. Phonetics of glottalized phonations in Yateé Zapotec. In Radek Skarnitzl & Jan Volín (eds.), *Proceedings of the 20th International Congress of Phonetic Sciences – ICPhS 2023*, 1751–1755. International Phonetic Association.

Chai, Yuan & Marc Garellek. 2022. On H1–H2 as an acoustic measure of linguistic phonation type. *The Journal of the Acoustical Society of America* 152(3). 1856–1870. https://doi.org/10.1121/10.0014175.

Chai, Yuan & Shihong Ye. 2022. Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min. *Languages* 7(1). 1–27. https://doi.org/10.3390/languages7010047.

Chávez-Peón, Mario E. 2008. Phonetic cues to stress in a tonal language: Prosodic prominence in San Lucas Quiiaviní Zapotec. In Susie Jones (ed.), *Proceedings of* *the 2008 annual conference of the Canadian Linguistic Association*. 13. Vancouver, Canada: University of British Columbia.

Chávez-Peón, Mario E. 2010. *The interaction of metrical structure, tone, and phonation types in Quiaviní Zapotec*. University of British Columbia dissertation. https://doi.org/10.14288/1.0071253.

Chen, Marilyn Y. 1997. Acoustic correlates of English and French nasalized vowels. *The Journal of the Acoustical Society of America* 102(4). 2360–2370. https://doi.org/10.1121/1.419620.

Coretta, Stefano. 2024. *Tidygam: Tidy Prediction and Plotting of Generalised Additive Models*. manual.

Cortés, Félix, Iara Mantenuto & Jeremy Steffman. 2023. San Sebastián del Monte Mixtec. *Journal of the International Phonetic Association*. 1–22. https://doi.org/10.1017/S0025100322000226.

Crowhurst, Megan J., Niamh E. Kelly & Amador Teodocio. 2016. The influence of vowel laryngealisation and duration on the rhythmic grouping preferences of Zapotec speakers. *Journal of Phonetics* 58. 48–70. https://doi.org/10.1016/j.wocn.2016.06.001.

Delgutte, Bertrand. 1982. Some correlates of phonetic distinctions at the level of the auditory nerve. In Rolf Carlson & Björn Granström (eds.), *The representation of speech in the peripheral auditory system: proceedings of the Symposium on the Rep-*

resentation of Speech in the Peripheral Auditory System held in Stockholm, Sweden on May 17-19, 1982, 131–150. Amsterdam: Elsevier Biomedical Press.

Delvaux, Véronique. 2009. Perception du contraste de nasalité vocalique en français. *Journal of French Language Studies* 19(1). 25–59. https://doi.org/10.1017/S0959269508003566.

DiCanio, Christian. 2008. *The Phonetics and Phonology of San Martín Itunyoso Trique*. Berkeley, CA: University of California, Berkeley dissertation.

DiCanio, Christian. 2009. The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association* 39(2). 162–188. https://doi.org/10.1017/S0025100309003879.

DiCanio, Christian. 2010. Itunyoso Trique. *Journal of the International Phonetic Association* 40(2). 227–238. https://doi.org/10.1017/S0025100310000034.

DiCanio, Christian. 2012a. Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics* 40(1). 162–176. https://doi.org/10.1016/j.wocn.2011.10.006.

DiCanio, Christian. 2012b. The phonetics of fortis and lenis consonants in Itunyoso Trique. *International Journal of American Linguistics* 78(2). 239–272. https://doi.org/10.1086/664481.

DiCanio, Christian. 2014. Cue weight in the perception of Trique glottal consonants. *The Journal of the Acoustical Society of America* 135(2). 884–895. https://doi.org/10.1121/1.4861921.

DiCanio, Christian, Basileo Martínez Cruz, Benigno Cruz Martínez & Wilberto Martínez Cruz. 2020. Glottal toggling in Itunyoso Triqui. *Phonological Data and Analysis* 2(4). 1–28. https://doi.org/10.3765/pda.v2art4.3.

Eischens, Benjamin. 2022. *Tone, Phonation, and the Phonology-Phonetics Interface in San Martín Peras Mixtec*. UC Santa Cruz.

Elliott, A. Raymond, Jerold A. Edmondson & Fausto Sandoval Cruz. 2016. Chicahuaxtla Triqui. *Journal of the International Phonetic Association* 46(3). 351–365. https://doi.org/10.1017/S0025100315000389.

Enfield, Nick J. 2005. Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.

Esling, John H., Scott R. Moisik, Allison Benner & Lise Crevier-Buchman. 2019. *Voice Quality: The Laryngeal Articulator Model.* 1st edn. (Cambridge Studies in Linguistics 162). Cambridge University Press. https://doi.org/10.1017/9781108696555.

Esposito, Christina M. 2004. Santa Ana del Valle Zapotec phonation. *UCLA Working Papers in Phonetics* (103). 71–105.

Esposito, Christina M. 2010a. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 38(2). 306–316. https://doi.org/10.1016/j.wocn.2010.02.002.

Esposito, Christina M. 2010b. Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association* 40(2). 181–198. https://doi.org/10.1017/S0025100310000046.

Esposito, Christina M. 2012. An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics* 40(3). 466–476. https://doi.org/10.1016/j.wocn.2012.02.007.

Esposito, Christina M. & Sameer ud Dowla Khan. 2020. The cross-linguistic patterns of phonation types. *Language and Linguistics Compass* 14(12). https://doi.org/10.1111/lnc3.12392.

Ferrer Riesgo, Carlos A. & Elmar Nöth. 2020. What Makes the Cepstral Peak Prominence Different to Other Acoustic Correlates of Vocal Quality? *Journal of Voice* 34(5). 806.e1–806.e6. https://doi.org/10.1016/j.jvoice.2019.01.004.

Fischer-Jørgensen, Eli. 1968. Phonetic Analysis of Breathy (Murmured) Vowels in Gujarati. *Annual Report of the Institute of Phonetics University of Copenhagen* 2. 35–85. https://doi.org/10.7146/aripuc.v2i.130674.

Fischer-Jørgensen, Eli. 1989. Phonetic Analysis of the Stød in Standard Danish. *Phonetica* 46(1–3). 1–59. https://doi.org/10.1159/000261828.

Foley, Steven, Nick Kalivoda & Maziar Toosarvandani. 2018. Forbidden clitic clusters in Zapotec. In Daniel Edmiston, Marina Ermolaeva, Emre Hakgüder, Jackie Lai, Kathryn Montemurro, Brandon Rhodes, Amara Sankhagowit & Michael Tabatowski (eds.), *Proceedings of the Fifty-third Annual Meeting of the Chicago Linguistic Society*, 87–102.

Foley, Steven & Maziar Toosarvandani. 2022. Extending the Person-Case Constraint to Gender: Agreement, Locality, and the Syntax of Pronouns. *Linguistic Inquiry* 53(1). 1–40. https://doi.org/10.1162/ling_a_00395.

Frazier, Melissa. 2013. The phonetics of Yucatec Maya and the typology of laryngeal complexity. *STUF - Language Typology and Universals* 66(1). 7–21. https://doi.org/10.1524/stuf.2013.0002.

Gao, Xin & Jianjing Kuang. 2022. Phonation Variation as a Function of Checked Syllables and Prosodic Boundaries. *Languages* 7(3). 171. https://doi.org/10.3390/languages7030171.

Garellek, Marc. 2012. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics* 40(1). 152–161. https://doi.org/10.1016/j.wocn.2011.10.003.

Garellek, Marc. 2019. The phonetics of voice. In William F. Katz & Peter F. Assmann (eds.), *The Routledge handbook of phonetics* (Routledge Handbooks in Linguistics), 75–106. Abingdon, Oxon ; New York, NY: Routledge.

Garellek, Marc. 2020. Acoustic Discriminability of the Complex Phonation System in !Xóõ. *Phonetica* 77(2). 131–160. https://doi.org/10.1159/000494301.

Garellek, Marc. 2022. Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality. *Journal of Phonetics* 94. 101155. https://doi.org/10.1016/j.wocn.2022.101155.

Garellek, Marc, Yuan Chai, Yaqian Huang & Maxine Van Doren. 2021. Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association*. 1–28. https://doi.org/10.1017/S0025100321000116.

Garellek, Marc & Christina M. Esposito. 2023. Phonetics of White Hmong vowel and tonal contrasts. *Journal of the International Phonetic Association* 53(1). 213–232. https://doi.org/10.1017/S0025100321000104.

Garellek, Marc & Patricia Keating. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41(2). 185–205. https://doi.org/10.1017/S0025100311000193.

Garellek, Marc, Patricia Keating, Christina M. Esposito & Jody Kreiman. 2013. Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America* 133(2). 1078–1089. https://doi.org/10.1121/1.4773259.

Garellek, Marc, Osbel López-Francisco & Jonathan D. Amith. 2025. At the margins: phonology and phonetics of Zongozotla Totonac glottalization. *Laboratory Phonology* 16(1). https://doi.org/10.16995/labphon.10925.

Garellek, Marc, Amanda Ritchart & Jianjing Kuang. 2016. Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics* 59. 110–121. https://doi.org/10.1016/j.wocn.2016.09.001.

Garellek, Marc, Robin Samlan, Bruce R. Gerratt & Jody Kreiman. 2016. Modeling the voice source in terms of spectral slopes. *The Journal of the Acoustical Society of America* 139(3). 1404–1410. https://doi.org/10.1121/1.4944474.

Gerfen, Chip. 1999. *Phonology and phonetics in Coatzospan Mixtec* (Studies in Natural Language and Linguistic Theory v. 48). Dordrecht ; Boston: Kluwer Academic. 304 pp.

Gerfen, Chip & Kirk Baker. 2005. The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics* 33(3). 311–334. https://doi.org/10.1016/j.wocn.2004.11.002.

Gerratt, Bruce R. & Jody Kreiman. 2001. Toward a taxonomy of nonmodal phonation. *Journal of Phonetics* 29(4). 365–381. https://doi.org/10.1006/jpho.2001.0149.

Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29(4). 383–406. https://doi.org/10.1006/jpho.2001.0147.

Grønnum, Nina, Miguel Vazquez-Larruscaín & Hans Basbøll. 2013. Danish Stød: Laryngealization or Tone. *Phonetica* 70(1–2). 66–92. https://doi.org/10.1159/000354640.

Hanson, Helen M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America* 101(1). 466–481. https://doi.org/10.1121/1.417991.

Hastie, Trevor & Robert Tibshirani. 1986. Generalized Additive Models. *Statistical Science* 1(3). 297–310. https://doi.org/10.1214/ss/1177013604.

Hastie, Trevor, Robert Tibshirani & Jerome Friedman. 2009. *The Elements of Statistical Learning* (Springer Series in Statistics). New York, NY: Springer. https://doi.org/10.1007/978-0-387-84858-7.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*. Modelling Sociophonetic Variation 34(4). 458–484. https://doi.org/10.1016/j.wocn.2005.10.001.

Henderson, Eugénie J. A. 1952. The Main Features of Cambodian Pronunciation. *Bulletin of SOAS* 14(1). 149–174. https://doi.org/10.1017/S0041977X00084251.

Herbst, Christian T. 2021. Performance Evaluation of Subharmonic-to-Harmonic Ratio (SHR) Computation. *Journal of Voice* 35(3). 365–375. https://doi.org/10.1016/j.jvoice.2019.11.005.

Herrera Zendejas, Esther. 2000. Amuzgo and Zapotec: Two More Cases of Laryngeally Complex Languages. *Anthropological Linguistics* 42(4). 545–563. http://www.jstor.org/stable/30028566.

Hillenbrand, James & Robert A. Houde. 1996a. Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech. *Journal of Speech, Language, and Hearing Research* 39(2). 311–321. https://doi.org/10.1044/jshr.3902.311.

Hillenbrand, James & Robert A. Houde. 1996b. Role of F0 and Amplitude in the Perception of Intervocalic Glottal Stops. *Journal of Speech, Language, and Hearing Research* 39(6). 1182–1190. https://doi.org/10.1044/jshr.3906.1182.

Hockett, Charles F. 1947. Peiping Phonology. *Journal of the American Oriental Society* 67(4). 253–267. https://doi.org/10.2307/596062.

Hollenbach, Barbara. 1984. *The phonology and morphology of tone and laryngeals in Copala Trique*. University of Arizona.

Holmberg, Eva B., Robert E. Hillman, Joseph S. Perkell, Peter C. Guiod & Susan L. Goldman. 1995. Comparisons Among Aerodynamic, Electroglottographic, and Acoustic

Spectral Measures of Female Voice. *Journal of Speech, Language, and Hearing Research* 38(6). 1212–1223. https://doi.org/10.1044/jshr.3806.1212.

Iseli, Markus, Yen-Liang Shue & Abeer Alwan. 2007. Age, sex, and vowel dependencies of acoustic measures related to the voice source. *The Journal of the Acoustical Society of America* 121(4). 2283–2295. https://doi.org/10.1121/1.2697522.

Jaeger, Jeri J. & Robert D. Van Valin. 1982. Initial Consonant Clusters in Yateé Zapotec. *International Journal of American Linguistics* 48(2). 125–138. https://doi.org/10.1086/465724.

James, Gareth, Daniela Witten, Trevor Hastie & Robert Tibshirani. 2021. *An introduction to statistical learning: with applications in R*. Second edition (Springer Texts in Statistics). New York: Springer. 1 p. https://doi.org/10.1007/978-1-0716-1418-1.

Janitza, Silke & Roman Hornung. 2018. On the overestimation of random forest's out-of-bag error. *PLOS ONE* 13(8). e0201904. https://doi.org/10.1371/journal.pone.0201904.

Johnson, Daniel Ezra. 2015. Quantifying Overlap with Bhattacharyya's affinity and other measures!

Josserand, Judy Kathryn. 1983. *Mixtec Dialect History*. New Orleans, LA: Tulane University dissertation. 727 pp.

Kawahara, Hideki, Alain De Cheveigne & Roy D. Patterson. 1998. An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite. In *5th International Conference on Spoken Language Processing (ICSLP 1998)*, paper 0659–. ISCA. https://doi.org/10.21437/ICSLP.1998-555.

Keating, Patricia, Christina M. Esposito, Marc Garellek, Sameer ud Dowla Khan & Jianjing Kuang. 2010. Phonation Contrasts Across Languages*. *UCLA Working Papers in Phonetics* (108). 188–202.

Keating, Patricia, Marc Garellek & Jody Kreiman. 2015. Acoustic properties of different kinds of creaky voice. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

Keating, Patricia, Jianjing Kuang, Marc Garellek, Christina M. Esposito & Sameer ud Dowla Khan. 2023. A cross-language acoustic space for vocalic phonation distinctions. *Language* 99(2). 351–389. https://doi.org/10.1353/lan.2023.a900090.

Kelterer, Anneliese & Barbara Schuppler. 2020. Phonation type contrasts and tone in Chichimec. *The Journal of the Acoustical Society of America* 147(4). 3043–3059. https://doi.org/10.1121/10.0001015.

Kent, Raymond D. & Martin J. Ball (eds.). 1999. *Voice quality measurement*. San Diego: Singular Pub. Group. 492 pp.

Khan, Sameer ud Dowla. 2012. The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics* 40(6). 780–795. https://doi.org/10.1016/j.wocn.2012.07.001.

Kirk, Paul L, Jenny Ladefoged & Peter Ladefoged. 1993. Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec. In Anthony Mattina & Timothy Montler (eds.), *American Indian linguistics and ethnography in honor of Laurence C. Thompson*, 435–450. Ann Arbor, MI: University of Michigan Press.

Klatt, Dennis H. & Laura C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 87(2). 820–857. https://doi.org/10.1121/1.398894.

Kreiman, Jody, Bruce R. Gerratt & Norma Antoñanzas-Barroso. 2007. Measures of the Glottal Source Spectrum. *Journal of Speech, Language, and Hearing Research* 50(3). 595–610. https://doi.org/10.1044/1092-4388(2007/042).

Kreiman, Jody, Bruce R. Gerratt, Marc Garellek, Robin Samlan & Zhaoyan Zhang. 2014. Toward a unified theory of voice production and perception. *Loquens* 1(1). e009. https://doi.org/10.3989/loquens.2014.009.

Kreiman, Jody, Yoonjeong Lee, Marc Garellek, Robin Samlan & Bruce R. Gerratt. 2021. Validating a psychoacoustic model of voice quality. *The Journal of the Acoustical Society of America* 149(1). 457–465. https://doi.org/10.1121/10.0003331.

de Krom, Guus. 1993. A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals. *Journal of Speech, Language, and Hearing Research* 36(2). 254–266. https://doi.org/10.1044/jshr.3602.254.

Kruskal, Joseph & Myron Wish. 1978. *Multidimensional Scaling*. SAGE Publications, Inc. https://doi.org/10.4135/9781412985130.

Kuang, Jianjing. 2017. Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America* 142(3). 1693–1706. https://doi.org/10.1121/1.5003649.

Ladefoged, Peter & Ian Maddieson. 1996. *The sounds of the world's languages* (Phonological Theory). Oxford, OX, UK ; Cambridge, Mass., USA: Blackwell Publishers. 425 pp.

Laver, John D. M. 1968. Voice Quality and Indexical Information. *British Journal of Disorders of Communication* 3(1). 43–54. https://doi.org/10.3109/13682826809011440.

Laver, John D. M. 1980. The Phonetic Description of Voice Quality. *Cambridge Studies in Linguistics London* 31. 1–186.

Lillehaugen, Brook Danielle. 2019. Otomanguean Languages. In *The Routledge Handbook of North American Languages*. Routledge.

Lindblom, Björn. 1983. Economy of Speech Gestures. In Peter F. MacNeilage (ed.), *The Production of Speech*, 217–245. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4613-8202-7_10.

Long, Rebecca & Sofronio Cruz. 2005. *Diccionario zapoteco de San Bartolomé Zoogocho, Oaxaca*. 2nd edition (Serie de vocabularios y diccionarios indígenas "Mariano Silva y Aceves" no. 38). Coyoacán, D.F. [Mexico]: Instituto Lingüístico de Verano. 531 pp.

López Nicolás, Oscar. 2016. *Estudios de la fonología y gramática del zapoteco de Zoochina*. Mexico City, Mexico: Ciudad de México: Centro de Investigaciones y Estudios Superiores en Antropología Social Doctoral thesis.

Lotto, A. J., L. L. Holt & K. R. Kluender. 1997. Effect of Voice Quality on Perceived Height of English Vowels. *Phonetica* 54(2). 76–93. https://doi.org/10.1159/000262212.

Macmillan, Neil A., John Kingston, Rachel Thorburn, Laura Walsh Dickey & Christine Bartels. 1999. Integrality of nasalization and F1. II. Basic sensitivity and phonetic labeling measure distinct sensory and decision–rule interactions. *The Journal of the Acoustical Society of America* 106(5). 2913–2932. https://doi.org/10.1121/1.428113.

Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Chicago: The University of Chicago Press.

Matisoff, James A. 1975. Rhinoglottophilia: The mysterious connection between nasality and glottality. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*, 265–287. Stanford: Stanford University Language Universals Project.

Merrill, Elizabeth D. 2008. Tilquiapan Zapotec. *Journal of the International Phonetic Association* 38(01). https://doi.org/10.1017/S0025100308003344.

Michaud, Alexis. 2012. The Complex Tones of East/Southeast Asian Languages: Current Challenges for Typology and Modelling. 8.

Mittal, Vinay Kumar, B. Yegnanarayana & Peri Bhaskararao. 2014. Study of the effects of vocal tract constriction on glottal vibration. *The Journal of the Acoustical Society of America* 136(4). 1932–1941. https://doi.org/10.1121/1.4894789.

Mock, Carol. 1988. Pitch accent and stress in Isthmus Zapotec. In Harry van der Hulst (ed.), *Autosegmental studies on pitch accent*, 197–223. Dordrecht: Foris.

Munro, Pamela & Felipe H. Lopez. 1999. *Di'csyonaary x:tèe'n dìi'zh sah Sann Lu'uc (San Lucas Quiaviní Zapotec dictionary/Diccionario Zapoteco de San Lucas Quiaviní)*. 2 vols. Los Angeles: (UCLA) Chicano Studies Research Center Publications.

Murty, K. Sri Rama & B. Yegnanarayana. 2008. Epoch Extraction from Speech Signals. *IEEE Transactions on Audio, Speech, and Language Processing* 16(8). 1602–1613. https://doi.org/10.1109/TASL.2008.2004526.

Nellis, Donald G. & Barbara E. Hollenbach. 1980. Fortis versus Lenis in Cajonos Zapotec Phonology. *International Journal of American Linguistics* 46(2). 92–105. https://doi.org/10.1086/465639.

Nycz, Jennifer & Lauren Hall-Lew. 2014. Best practices in measuring vowel merger. *Proceedings of Meetings on Acoustics* 20(1). 060008. https://doi.org/10.1121/1.4894063.

Ohala, John J. 1975. Phonetic explanations for nasal sound patterns. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*, 289–316. Stanford: Stanford University Language Universals Project.

Ohala, John J. & Manjari Ohala. 1993. The phonetics of nasal phonology: Theorems and data. In Marie K. Huffman & Rena A. Krakow (eds.), *Nasals, Nasalization, and the Velum*, vol. 5 (Phonetics and Phonology), 225–249. San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-360380-7.50013-2.

Oksanen, Jari, Gavin L. Simpson, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R.B. O'Hara, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner, Matt Barbour, Michael Bedward, Ben Bolker, Daniel Borcard, Gustavo Carvalho, Michael Chirico, Miquel De Caceres, Sebastien Durand, Heloisa Beatriz Antoniazi Evangelista, Rich FitzJohn, Michael Friendly, Brendan Furneaux, Geoffrey Hannigan, Mark O. Hill, Leo Lahti, Dan McGlinn, Marie-

Helene Ouellette, Eduardo Ribeiro Cunha, Tyler Smith, Adrian Stier, Cajo J.F. Ter Braak, James Weedon & Tuomas Borman. 2025. *Vegan: Community Ecology Package*. manual.

Peña, Jailyn M. 2022. Stød Timing and Domain in Danish. *Languages* 7(1). 50. https://doi.org/10.3390/languages7010050.

Peña, Jailyn M. 2024. *The production and perception of stød in Danish*. New York City, NY: New York University dissertation.

Pickett, Velma B., María Villalobos Villalobos & Stephen A. Marlett. 2010. Isthmus (Juchitán) Zapotec. *Journal of the International Phonetic Association* 40(3). 365–372. https://doi.org/10.1017/S0025100310000174.

Pike, Eunice Victoria. 1948. Problems in Zapotec Tone Analysis. *International Journal of American Linguistics* 14(3). 161–170. https://doi.org/10.1086/463998.

Pillai, K. C. S. 1955. Some New Test Criteria in Multivariate Analysis. *The Annals of Mathematical Statistics* 26(1). 117–121. https://doi.org/10.1214/aoms/1177728599.

Podesva, Robert J. 2016. Stance as a Window into the Language-Race Connection: Evidence from African American and White Speakers in Washington, DC. In H. Samy Alim, John R. Rickford & Arnetha F. Ball (eds.), *Raciolinguistics: How Language Shapes Our Ideas About Race*, 203–219. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190625696.003.0012.

Pruthi, Tarun & Carol Y. Espy-Wilson. 2004. Acoustic parameters for automatic detection of nasal manner. *Speech Communication* 43(3). 225–239. https://doi.org/10.1016/j.specom.2004.06.001.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. manual. Vienna, Austria: R Foundation for Statistical Computing.

Ratliff, Martha Susan. 1992. *Meaningful tone: a study of tonal morphology in compounds, form classes and expressive phrases in White Hmong* (Special Report no. 27, 1992). DeKalb, Ill: Northern Illinois University, Center for Southeast Asian Studies. 269 pp.

Remez, Robert E. & Philip E. Rubin. 1993. On the intonation of sinusoidal sentences: Contour and pitch height. *The Journal of the Acoustical Society of America* 94(4). 1983–1988. https://doi.org/10.1121/1.407501.

Rensch, Calvin R. 1976. *Comparative Otomanguean phonology* (Language Science Monographs 14). Bloomington: Indiana University Press.

Ritsma, Roelof J. 1967. Frequencies Dominant in the Perception of the Pitch of Complex Sounds. *The Journal of the Acoustical Society of America* 42(1). 191–198. https://doi.org/10.1121/1.1910550.

Rosenberg, Aaron E. 1966. Pitch Discrimination of Jittered Pulse Trains. *The Journal of the Acoustical Society of America* 39. 920–928. https://doi.org/10.1121/1.1909973.

Samely, Ursula B. & Ursula Samely. 1991. *Kedang, (Eastern Indonesia): some aspects of its grammar* (Forum phoneticum 46). Hamburg: Buske. 235 pp.

Sandri, Marco & Paola and Zuccolotto. 2008. A Bias Correction Algorithm for the Gini Variable Importance Measure in Classification Trees. *Journal of Computational and Graphical Statistics* 17(3). 611–628. https://doi.org/10.1198/106186008X344522.

Schwartz, Martin F. 1968. The Acoustics of Normal and Nasal Vowel Production. *The Cleft Palate Journal* 5(2). 125–140.

Seyfarth, Scott & Marc Garellek. 2018. Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics* 71. 425–450. https://doi.org/10.1016/j.wocn.2018.09.001.

Shue, Yen-Liang, Patricia Keating, Chad Vicenik & Kristine Yu. 2011. VoiceSauce: A program for voice analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, 1846–1849. Hong Kong.

Sichel, Ivy & Maziar Toosarvandani. 2020a. Pronouns and Attraction in Sierra Zapotec. In Andrew Hedding & Morwenna Hoeks (eds.), *Syntax and semantics at Santa Cruz, Volume IV* (Syntax & Semantics at Santa Cruz (SASC) 4). Santa Cruz, CA: Linguistics Research Center.

Sichel, Ivy & Maziar Toosarvandani. 2020b. The featural life of nominals. lingbuzz/005523.

Silverman, Daniel. 1997a. Laryngeal complexity in Otomanguean vowels. *Phonology* 14(2). 235–261. https://doi.org/10.1017/S0952675797003412.

Silverman, Daniel. 1997b. *Phasing and recoverability* (Outstanding Dissertations in Linguistics). New York: Garland Pub. 242 pp.

Simpson, Adrian 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. https://doi.org/10.1016/j.wocn.2012.02.001.

Sjölander, Kåre. 2004. *Snack Sound Toolkit*. Stockholm, Sweden: KTH.

Smalley, William A. 1976. The problems of consonants and tone: Hmong (Meo, Miao). In W. A. Smalley (ed.), *Phonemes and orthography: Language planning in ten minority languages of thailand*, vol. 43 (Pacific Linguistics, Series C), 85–123. Canberra: Australian National University.

Smith-Stark, Thomas. 2007. Algunas isoglosas zapotecas. In *Clasidicación de las lenguas indígenas de México: Memorias del III Coloquio Internacional de Lingüística Mauricio Swadesh*, 69–133. Mexico City: UNAM: Instituto de Investigaciones Antropológicas.

Sonnenschein, Aaron H. 2005. *A descriptive grammar of San Bartolomé Zoogocho Zapotec*. Munich: Lincom Europa.

Sóskuthy, Márton. 2017. *Generalised Additive Mixed Models for Dynamic Analysis in Linguistics: A Practical Introduction*. arXiv: 1703.05339 [stat]. Pre-published.

Sóskuthy, Márton. 2021. Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics* 84. 101017. https://doi.org/10.1016/j.wocn.2020.101017.

Stevens, Kenneth N. 2000. *Acoustic phonetics*. 1. paperback ed (Current Studies in Linguistics Series 30). Cambridge, Mass.: MIT Press. 607 pp.

Strelluf, Christopher. 2018. Chapter 3: The Low Back Vowel(s). *The Publication of the American Dialect Society* 103(1). 43–64. https://doi.org/10.1215/00031283-7318737.

Strobl, Carolin, Anne-Laure Boulesteix, Achim Zeileis & Torsten Hothorn. 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8(1). 25. https://doi.org/10.1186/1471-2105-8-25.

Styler, Will. 2015. *On the acoustical and perceptual features of vowel nasality*. University of Colorado at Boulder dissertation.

Styler, Will. 2017. On the acoustical features of vowel nasality in English and French. *The Journal of the Acoustical Society of America* 142(4). 2469–2482. https://doi.org/10.1121/1.5008854.

Sun, Xuejing. 2002. Pitch determination and voice quality analysis using Subharmonic-to-Harmonic Ratio. *2002 IEEE International Conference on Acoustics, Speech, and*

Signal Processing 1. I-333-I–336. https://doi.org/10.1109/ICASSP.2002.5743722.

Sundberg, Johan. 2022. Objective Characterization of Phonation Type Using Amplitude of Flow Glottogram Pulse and of Voice Source Fundamental. *Journal of Voice* 36(1). 4–14. https://doi.org/10.1016/j.jvoice.2020.03.018.

Tagliamonte, Sali A. & R. Harald Baayen. 2012. Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change* 24(2). 135–178. https://doi.org/10.1017/S0954394512000129.

Thurgood, Graham. 2002. Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica* 19(2). 333–363. https://doi.org/10.1075/dia.19.2.04thu.

Tyler, Richard S., Quentin Summerfield, Elizabeth J. Wood & Mariano A. Fernandes. 1982. Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America* 72(3). 740–752. https://doi.org/10.1121/1.388254.

Uchihara, Hiroto. 2016. Tone and registrogenesis in Quiaviní Zapotec. *Diachronica* 33(2). 220–254. https://doi.org/10.1075/dia.33.2.03uch.

Uchihara, Hiroto & Gabriela Pérez Báez. 2016. Fortis/lenis, glides and vowels in Quiaviní Zapotec. *Glossa: A Journal of General Linguistics* 1(1). 27. https://doi.org/10.5334/gjgl.13.

Villard, Stéphanie. 2015. *The Phonology and Morphology of Zacatepec Eastern Chatino*. Austin, TX: University of Texas at Austin dissertation.

Warren, Paul. 2018. Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association* 48(3). 305–330. https://doi.org/10.1017/S0025100317000329.

Wayland, Ratree & Allard Jongman. 2003. Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics* 31(2). 181–201. https://doi.org/10.1016/S0095-4470(02)00086-4.

Weller, Jae, Jeremy Steffman, Félix Cortés & Iara Mantenuto. 2023a. Interactions of tone, glottalization, and word shape in San Sebastián Del Monte Mixtec.

Weller, Jae, Jeremy Steffman, Félix Cortés & Iara Mantenuto. 2023b. Lexical tone and vowel duration in San Sebastián del Monte Mixtec. *The Journal of the Acoustical Society of America* 153. A293. https://doi.org/10.1121/10.0018900.

Weller, Jae, Jeremy Steffman, Félix Cortés & Iara Mantenuto. 2024. Voice Quality and Tone in San Sebastián del Monte Rearticulated and Modal Vowels.

Wieling, Martijn. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and

L2 speakers of English. *Journal of Phonetics* 70. 86–116. https://doi.org/10.1016/j.wocn.2018.03.002.

Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd edn. (Texts in Statistical Science Series). Boca Raton, FL: Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279.

Yip, Moira Jean Winsland. 2002. *Tone* (Cambridge Textbooks in Linguistics). Cambridge ; New York: Cambridge University Press. 341 pp.

Zapotec Language Project — University of California, Santa Cruz. 2022. https://zapotec.ucsc.edu/ (6 December, 2022).