

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

Copyright © by
Mykel Loren Brinkerhoff
2025

Contents

List of Figures	vi
List of Tables	ix
Abstract	xi
Dedication	xii
Acknowledgments	xiii
1 Introduction	1
1.1 What is Voice Quality	1
1.2 Voice Quality and Tone	2
1.3 Voice Quality and Phonation	2
1.4 Interactions between Voice Quality and Tone	2
1.4.1 Why Esposito & Khan 2020 matters to my research	2
1.4.2 Connecting the two	4
1.5 Esposito & Khan 2020	5
1.5.1 Type I languages	8
1.5.2 Type II languages	9
1.5.3 Type III languages	10
1.5.4 Type IV languages	12
2 Vowels and suprasegmentals in Santiago Laxopa Zapotec	16
2.1 Introduction	16
2.2 Vowels in Santiago Laxopa Zapotec	17
2.3 Phonation in Santiago Laxopa Zapotec	20
2.3.1 Phonation types in SLZ	20
2.3.2 Breathy vowels in SLZ	21

2.3.3	Checked vowels in SLZ	21
2.3.4	Rearticulated vowels in SLZ	22
2.4	Tonal contrasts in Santiago Laxopa Zapotec	22
2.5	Interactions between tone and voice quality	23
3	On using Residual H1* for voice quality research	24
3.1	Introduction	24
3.2	Santiago Laxopa Zapotec	28
3.3	Methods	30
3.3.1	Elicitation	30
3.3.2	Data Processing	31
3.3.3	Statistical modeling	33
3.4	Results	35
3.4.1	H1*–H2*	35
3.4.2	Residual H1*	36
3.4.3	Model Comparison	37
3.4.4	GAMM analysis and model comparisons	39
3.5	Conclusion	42
4	The acoustic space of voice quality in Santiago Laxopa Zapotec	43
4.1	Introduction	43
4.2	Methods	44
4.2.1	Participants	44
4.2.2	Recordings	44
4.2.3	Acoustic measuring	45
4.2.4	Data processing	46
4.2.5	Statistical analysis	48
4.3	Results	50
4.3.1	Acoustic space of voice quality	50
4.3.2	Acoustic correlates of voice quality	57
4.4	Discussion	64
4.5	Conclusion	67
5	Trees reveal the importance of measures in SLZ	69
5.1	Introduction	69
5.2	What are Decision Trees	70
5.3	Decision trees in linguistics	72
5.4	Growing a forest of decision trees	74
5.4.1	Bagging trees	75
5.4.2	Random Forests	76
5.4.3	How to interpret the results	77

5.5	Random Forests in SLZ	81
5.5.1	Methods	81
5.5.2	Parameter selection	85
5.6	Results	88
5.7	Discussion of the results	90
5.7.1	Comparing the MDS and bagging trees	90
5.7.2	Importance of duration	92
5.7.3	Importance of A1*	94
5.7.4	Importance of H1*—A1*	97
5.7.5	Importance of Residual H1*	98
5.7.6	Importance of the HNR <1500 Hz	100
5.7.7	Importance of Strength of Excitation	102
5.8	Conclusion	102
6	Testing the laryngeal complexity in SLZ	103
6.1	What is Laryngeal Complexity?	103
6.1.1	Phasing and recoverability	104
6.1.2	Implicational hierarchy of laryngealization	108
6.2	Previous analyses of laryngeal complexity	110
6.3	Analysis of laryngeal complexity	114
6.4	Results	114
6.5	Discussion	114
6.6	Conclusion	114
7	Modeling laryngeal complexity	115
7.1	The Laryngeal Articulator Model	115
7.2	Modeling laryngeal complexity	117
7.3	Alternative accounts	117
7.3.1	Articulatory Phonology account	117
7.3.2	Q-theory account	117
7.3.3	Radical CV Phonology account	117
7.4	Conclusion	117
8	Conclusion	118
	Bibliography	119
A	Some Ancillary Stuff	146

List of Figures

2.1	Santiago Laxopa taken by Beto Diaz, a resident of Santiago Laxopa. .	17
2.2	Vowel space of Santiago Laxopa Zapotec. The ellipses around each vowel mean represents 1 standard. The scale of the axes are in barks with their corresponding Hz values.	19
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.	36
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.	37
3.3	GAMM smooths and difference plots for H1*–H2* across the duration of the vowel.	40
3.4	GAMM smooths and difference plots for residual H1* across the duration of the vowel.	41
4.1	Scree plot showing the stress for each dimension for the MDS analysis.	50
4.2	Two-dimensional MDS solution showing the first and second dimensions.	51
4.3	Two-dimensional MDS solution showing the first and third dimensions.	52
4.4	Two-dimensional MDS solution showing the first and fourth dimensions.	53
4.5	Two-dimensional MDS solution showing the second and third dimensions.	54
4.6	Two-dimensional MDS solution showing the second and fourth dimensions.	55

4.7	Two-dimensional MDS solution showing the third and fourth dimensions.	56
4.8	A diagram showing the relationship between breathy, modal, and creaky phonation types from Gordon & Ladefoged (2001).	56
4.9	A diagram showing the relationship between breathy, modal, and creaky phonation types. Based on Gordon & Ladefoged (2001).	65
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.	67
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.	73
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.	80
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.	87
5.4	Variable importance plots showing the impurity importance and permutation importance of each acoustic measure.	89
5.5	Plot showing the distribution of duration across the different voice qualities in SLZ.	93
5.6	Plot showing the distribution of $A1^*$ across the different voice qualities in SLZ.	95
5.7	Plot showing the distribution of $H1^* - A1^*$ across the different voice qualities in SLZ.	99
5.8	Plot showing the distribution of residual $H1^*$ across the different voice qualities in SLZ.	100
5.9	Plot showing the distribution of $HNR < 1500$ Hz across the different voice qualities in SLZ.	101
6.1	A diagram showing the relationship between breathy, modal, and creaky phonation types. Based on Gordon & Ladefoged (2001).	107

7.1	The Laryngeal Articulator Model from Esling et al. (2019). This model shows the interactions between the laryngeal articulators (labeled circles). Syngeristic interactions are shown with solid lines, while anti-syngeristic interactions are shown with dotted lines.	116
-----	--	-----

List of Tables

1.1	Language types based on the lexically contrastive nature of <i>f0</i> /tone/pitch accents (rows) and voice quality on vowels (columns) and their interactions (Esposito & Khan 2020)	2
1.2	Language types based on the lexically contrastive nature of <i>f0</i> /tone/pitch accents (rows) and voice quality on vowels (columns) and their interactions (Esposito & Khan 2020)	8
1.3	Updated language types based on the lexically contrastive nature of <i>f0</i> /tone/pitch accents (rows) and voice quality on vowels (columns) and their interactions	15
2.1	Vowel qualities in Santiago Laxopa Zapotec.	18
2.2	Pillai scores and Bhattacharyya's Affinity for /o/ and /u/ in SLZ. . . .	19
2.3	SLZ tone and voice quality combinations.	23
3.1	Distribution of tone and voice quality in the wordlist	31
3.2	Model comparison between H1*–H2* and Residual H1* in distinguishing Santiago Laxopa Zapotec voice quality.	38
3.3	AIC for the H1*–H2* and residual H1* models.	38
4.1	Correlations for each acoustic measure to the four dimensions (NMDS1, NMDS2, NMDS3, NMDS4). The four largest correlations in each dimension are bolded.	59
5.1	Correlations for each acoustic measure to the four dimensions (NMDS1, NMDS2, NMDS3, NMDS4). The four largest correlations in each dimension are bolded.	91
5.2	Possible phonetic correlates of register. From Brunelle & Kirby (2016).	96

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).	108
-----	---	-----

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 in 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

...

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 describes the state of the larynx during phonation, when the vocal folds are set in motion. Languages make use of voice quality for paralinguistic purposes, such as conveying indexation of “biological, psychological, and social characteristics of the speaker” (Laver 1968) and racial identity (Podesva 2016).

Voice quality is also used linguistically. In English, it is often the case that we use creaky voice to indicate that we are at the end of an utterance (e.g., Garellek 2013). In many other languages, voice quality is used as part of the phonological system. Most famously, Gujarati has a phonemic contrast between breathy and modal voice in vowels (e.g., Fischer-Jørgensen 1968, Esposito & Khan 2012, Khan 2012, Esposito et al.

2019).

Esposito & Khan (2020)

1.2 Voice Quality and Tone

1.3 Voice Quality and Phonation

1.4 Interactions between Voice Quality and Tone

1.4.1 Why Esposito & Khan 2020 matters to my research

- Esposito & Khan (2020) is primarily important for my research because of their typology of the interactions between tone and phonation, which are summarized in Table 1.1.

Table 1.1: Language types based on the lexically contrastive nature of f_0 /tone/pitch accents (rows) and voice quality on vowels (columns) and their interactions (Esposito & Khan 2020)

	No VQ contrast		VQ contrast	
No tone contrast	No contrasts (Type I)		f_0 not a cue (Type IIa)	f_0 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)

- Another aspect of my research into the tone and phonation interactions is why we see gaps/restrictions for which tones and phonations are allowed to combine.

- This is important because of the gaps I observe in SLZ where some we never see breathy with high tone and we do not see checked vowels with rising tone.
 - * This is true for nominals.
 - * I have not looked at verbs and whether or not we see this same gap.
 - I expect that the gap is also present in the verbal paradigms similar to what Uchihara (2016) observed where breathy phonation fails to appear in some parts of the paradigm.
 - Especially with the Potential Aspect, which is always realized as a high tone on the verbal root.
- These four types of languages offer an excellent way for me to characterize what we see cross-linguistically in the interactions between tone and phonation.
- This primarily is useful for me as a way to zero in on which languages I need to look at.
 - This means that I need to focus my search into types IIb, IIIb, and IV languages.

1.4.2 Connecting the two

- The connection here between Silverman (1997a) and the typology is that these both have to deal with the interactions between tone and phonation.
- Silverman offers an account for what we expect to see when a language has both tone and phonation.
 - It additionally offers three motivating factors for interactions between tone and phonation.
 - These three factors also offer ways to account for the patterns that we observe and don't observe in the typology
- The typology allows us to see more generally what combinations we have and don't have.
- If there are any cross-linguistic gaps that we observe, then do Silverman's (1997a) three factors of (i) sufficient acoustic distance, (ii) sufficient articulatory compatibility, and (iii) optimal auditory salience provide the answers to how and why.

1.5 Esposito & Khan 2020

- Esposito & Khan (2020) is a paper that explains the various cross-linguistic patterns that exist in the world's languages for phonation.
- They describe that for the majority of the world's languages there is only modal voice.
- However, a subset of the languages have non-modal phonation.
- They state that languages can either associate phonation, which they define as production of sound by the vocal folds, with consonants or vowels.
 - For consonants this takes the form of breathy or creaky voice (4ff).
- I am going to ignore their discussion of phonation associated with consonants for the most part and focus on what they have to say about vowels.
 - According to them only five languages contrast phonation on both vowels and consonants.
 - These are !Xóõ, Ju|'hoansi, Wa, White Hmong, and Gujarati
- They describe that there are two main avenues for researching phonation's production.
 - This is done through acoustic measurements and electroglottography

- The most common way that this is done is with spectral tilt measures which reflect the open quotient.
 - The open quotient is the proportion of the glottal cycle during which the glottis is open (Holmberg et al. 1995).
- Other measures that reflect the periodicity of the signal are also consulted such as HNR and CPP. Typically nonmodal phonation is also aperiodic and will have a lower score compared to the modal.
- They also state that “H1-H2 may be a (near-)universal acoustic measure of phonation” (8).
- They further say that this measure works the best the most often but there are several exceptions where other H1-X measures are more robust at capturing the phonation contrasts in the language.
 - As discussed in Chai & Garellek (2022), what they are actually trying to do is get at the amplitude differences in H1 as first observed by Fischer-Jørgensen (1968) and Laver (1968).
 - I think this is where, I and other researchers fell into a fallacy where we think that just the spectral-tilt is what matters instead of realizing what those measures were trying to accomplish which was to normalize H1 for comparison.

- Esposito & Khan (2020) further discuss the role that EGG play.
- Two things that I think are relevant for my dissertation is the discussion around localization of non-modal phonation and how phonation relates to tone.
- For the discussion around localization, they keep in line with Silverman (1997a) who says that phonation is located in certain portions of the vowel.
- Localization of non-modal phonation is also important in languages with complex “clusters” (i.e., phonetic sequences) of phonation type, which involve at least one non-modal phonation localized to a portion of the vowel.
 - !Xóõ distinguishes clusters of breathy-creaky, pharyngealized-creaky, and breathy-pharyngealized phonation in addition to modal, breathy, creaky, and pharyngealized voice
- In terms of tone and phonation interactions, Esposito & Khan (2020) claim that there are four different types of interactions based on whether or not tone and phonation is contrastive in the language.
- Table 1.2 shows each of the different interactions between tone and phonation. In this table each row represents whether or not tone/*f₀*/pitch accent is contrastive and each column represents whether or not phonation is contrastive.
- Each cell represent each of the four different types of languages (I-IV).

- Types II-IV are further subdivided based on whether or not tone and phonation are cues for each other.

Table 1.2: Language types based on the lexically contrastive nature of $f\theta$ /tone/pitch accents (rows) and voice quality on vowels (columns) and their interactions (Esposito & Khan 2020)

	No VQ contrast		VQ contrast	
No tone contrast	No contrasts (Type I)		$f\theta$ not a cue (Type IIa)	$f\theta$ 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)

1.5.1 Type I languages

- Type I languages are those languages that lack tonal contrasts and phonation contrasts.
- According to Esposito & Khan (2020) these are the languages that belong to this type:
 - most Australian Aboriginal languages,
 - most Austronesian languages,
 - most Indo-European languages
 - Afro-Asiatic languages,
 - Standard Khmer,

- Turkic languages
- They do not go into much detail about these languages

1.5.2 Type II languages

- Type II languages are languages that lack tonal contrasts but do have phonation contrasts on the vowel.
- This is further subdivided into two subtypes (IIa and IIb) based on whether or not changes in f_0 are a cue to the phonation.
- Type IIa languages are those where f_0 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 (Grønnum, Vazquez-Larruscaín & Basbøll 2013)
 - Gujarati (Khan 2012)
- Type IIb languages are those where f_0 is a cue for the different phonation types.
- These languages are sometimes called “register languages”
- These are languages such as:
 - Chanthaburi Khmer

- Chong
 - Javanese
 - Kedang
 - Mon
 - Suai
 - Wa
- In type IIb languages the way in which f_0 is a cue is language specific.
 - They describe that breathy vowels in Wa and Kedang both begin with a lower f_0 than their clear/modal counterparts
 - However, breathy vowels in Chanthaburi Khmer have a higher f_0 than their clear counterparts.

1.5.3 Type III languages

- Type III languages are characterized by languages that have a tonal contrast but lack a phonation contrast.
- These languages are further subdivided into two subtypes based on whether or not phonation is a cue for the tonal contrasts (Type IIIa vs. Type IIIb)

- In Type IIIa languages, tone is contrastive and phonation is not a cue for those tonal contrasts.
- These are languages like:
 - Japanese,
 - Navajo
 - Punjabi
 - Manange
 - Most W. African languages
 - Swedish
 - Central Thai
- Type IIIb languages are those that have lexical tones and no lexical vowel phonation. However, subsets of the tone categories have been reported to have optional phonation which helps cue the tonal contrasts.
- For example, tone 3 of Mandarin Chinese which is usually accompanied with creak (Kuang 2017).
- These are languages such as:
 - Cantonese Yue,
 - Khmu' Rawk

- Mandarin
- Pakphanang Thai
- Ph.Penh Khmer
- Yueyang Xiang

1.5.4 Type IV languages

- Type IV languages are those languages that have both a contrast in tone and phonation and is also subdivided into two subtypes (IVa and IVb).
- Type IVa are those languages where tone and phonation are completely orthogonal to each other. This means that all tone contrasts can appear with any of the phonation contrasts.
- Type IVa languages include:
 - Dinka,
 - Mazatec language,
 - Mpi
 - Yalálag Zapotec
 - Yi languages such as Bo

- Type IVb languages are often called “register languages” because tone and phonation are so closely tied to one another that it is impossible to state whether or not they have tonal contrasts or phonation contrasts.
- In these languages suprasegmental categories (i.e., tone) are consistently produced with both a specific f_0 as well as a specific phonation.
- Examples for this include the behavior where specific tones only arise with specific phonation types. This is most commonly found among languages in Mainland Southeast Asia.
- Esposito & Khan (2020) additionally describe that most Zapotec languages also fall into this type of language.
 - I do not agree with the strong version of this statement. The evidence that they cite for this comes from Esposito (2010) where she describes that SAV Zapotec syllables fall into one of four categories: breathy phonation with falling tone, creaky phonation with low falling tone, modal phonation with a high tone, and modal phonation with rising tone
 - It is true that some phonation types are closely tied to one tonal pattern in Zapotec languages.
 - For example, in Isthmus Zapotec (Pickett, Villalobos & Marlett 2010) is described as having five tonal melodies (H, L, LH, HL, and LHL) and three

phonation types (modal, checked, and laryngealized).

- In monosyllabic nouns L and LH can surface with any of the three phonation types but the combinations of H and HL with any of the phonation types do not appear to be present in the lexicon
- When we consider disyllabic nouns we see the following combinations:
 - * L appears with any of the phonation types.
 - * H appears with modal and laryngealized
 - * LH appears with modal and checked
 - * HL appears with modal and checked
 - * LHL only appears with modal.
- This is also true in SLZ where the only combination of tone and phonation that we fail to see is H with breathy phonation and MH with checked.
- This leads me to believe that maybe it would be better to think of Zapotec languages as a third subtype of languages where some of the tone and phonation contrasts are orthogonal and other tone and phonation contrasts are fused.
- In other words a hybrid between Type IVa and Type IVb which I would call mixed languages or Type IVc.
- This is illustrated in a new updated table in Table 1.3.

Table 1.3: Updated language types based on the lexically contrastive nature of *f* θ /tone/pitch accents (rows) and voice quality on vowels (columns) and their interactions

	No VQ contrast		VQ contrast		
No tone contrast	No contrasts (Type I)		<i>f</i> θ not a cue (Type IIa)	<i>f</i> θ 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)	Mixed (Type IVc)

Chapter 2

Vowels and suprasegmentals in Santiago Laxopa Zapotec

2.1 Introduction

Santiago Laxopa Zapotec (SLZ; *Dilla'xhunh Laxup* [diz̥a'z̥un l:aʃup^h]) 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: Santiago Laxopa taken by Beto Diaz, a resident of Santiago Laxopa.

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 Bhattacharyya's Affinity (Bhattacharyya 1943, Johnson 2015, Warren 2018, Strelluf 2018). Both of these measures show what degree of overlap exists between two different items in some space. Their use in linguistics has been used mainly to

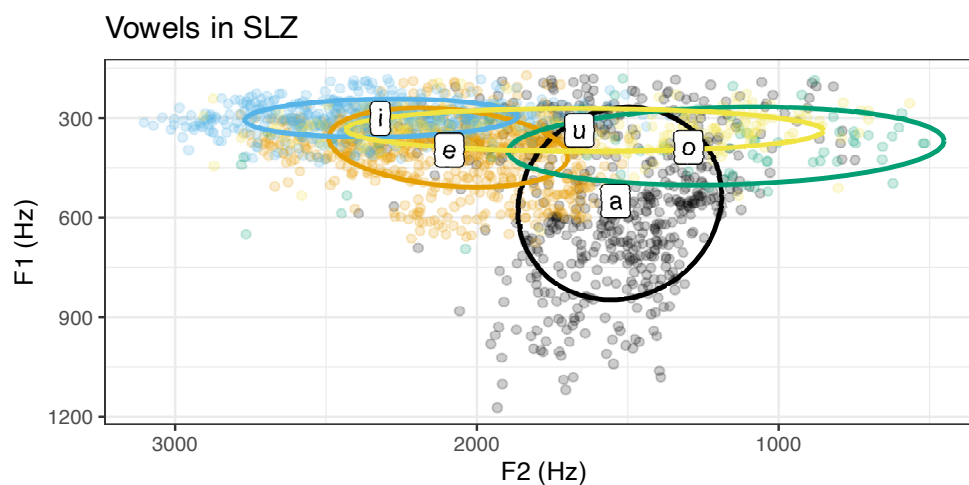


Figure 2.2: Vowel space of Santiago Laxopa Zapotec. The ellipses around each vowel mean represents 1 standard. The scale of the axes are in barks with their corresponding Hz values.

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

2.3.1 Phonation types in SLZ

Most Zapotec languages also make use of contrastive voice qualities (see Ariza-García 2018 for an overview and typology of the voice quality contrasts in the Zapotec language family), with SLZ being no exception. SLZ has a four-way voice quality contrast: modal, breathy, checked, and rearticulated. These contrasts are exemplified in the minimal quadruple in (1).

(1) Four-way near minimal phonation contrast

- a. *yag* /çag^L/ ‘tree; wood; almúd (unit of measurement ~4kg)’
- b. *yah* /ça^L/ ‘metal; rifle; bell’

c. *cha'* /tʃaʔ^L/ ‘cooking pot’

d. *ya'a* /çaʔa^L/ ‘market’

2.3.2 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).¹ Breathy voice is characterized by a raspiness throughout the whole vowel or a portion of the vowel depending on the speaker.

[INSERT SPECTROGRAM OF BREATHY VOWEL]

2.3.3 Checked vowels in SLZ

SLZ shares with most Zapotec varieties, two types of creaky voice: checked and rearticulated. Checked vowels are characterized by an abrupt glottal closure which cuts the vowel short. This phonation is sometimes realized as a period of creakiness at the end of the vowel.

[INSERT SPECTROGRAM OF CHECKED VOWEL]

¹Breathy voice in Zapotec languages, however, is common in Central Valley Zapotecs (Munro & Lopez 1999, Esposito 2004, 2010, Uchihara 2016, Ariza-García 2018).

2.3.4 Rearticulated vowels in SLZ

Among speakers of SLZ, there is a large amount of inter- and intra-speaker variability in how the rearticulated vowels are produced. Some speakers produce these vowels with a full glottal stop in the middle of the vowel, 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 creaky voice throughout the entire vowel. Some speakers produce a combination of these unique productions. Overall, these rearticulated vowels are produced with some form of manipulation of glottal closure or amplitude drop in the middle of the vowel.

[INSERT SPECTROGRAMS OF REARTICULATED VOWELS]

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).

Brinkerhoff, Duff & Wax Cavallaro (2022)

2.5 Interactions between tone and voice quality

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 2.3.

Table 2.3: SLZ tone and voice quality combinations.

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

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).

One measure that Fischer-Jorgensen established, the fundamental’s relative strength-

ive amplitudes of the first harmonic and second harmonic. As established by Fischer-Jørgensen, 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 (2010) 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 was also found to be the case in Betaza Zapotec (Crowhurst, Kelly & Teodocio 2016).

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.

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 (/ʔ/) 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 (/Vʔ/ or /Ṿ/). 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 (/VʔV/ or /ṾV/). 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 a phonetic underpinning.

²This 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).

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

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* [ʃn:ia' X tʃone ras] “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

⁴There is ongoing debate about the status of strong syllables in Zapotec. The majority of evidence

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

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.

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 (f_0) 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 f_0 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 f_0 , F1, or F2 values were outliers were also excluded from $H1^*$

and $H1^* - H2^*$ analyzes because $H1^*$ and $H1^* - H2^*$ are calculated based on f_0 , 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

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

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

⁶Tone 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.

Additional statistical modeling was performed with two generalized additive mixed models (GAMM) using the `mgcv` package in R (Wood 2017). The GAMMs were fitted to account for any non-linear effects that may be present in the data. In each case the response variable was the same as in the linear mixed-effects models with voice quality as a fixed effect. Splines were used to model the non-linear effects of position and position by phonation. Speaker and word were treated as random intercepts.

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.

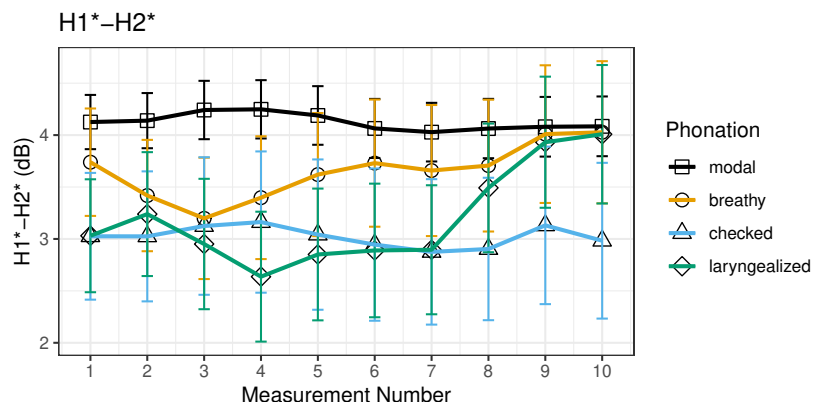


Figure 3.1: $H1^*-H2^*$ across the duration of the vowel. Points represent the mean of each measure across the ten intervals. The error bars around each point represent a 95% confidence interval. A line was plotted over each to show how the acoustic measure functions across the ten intervals.

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. 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.

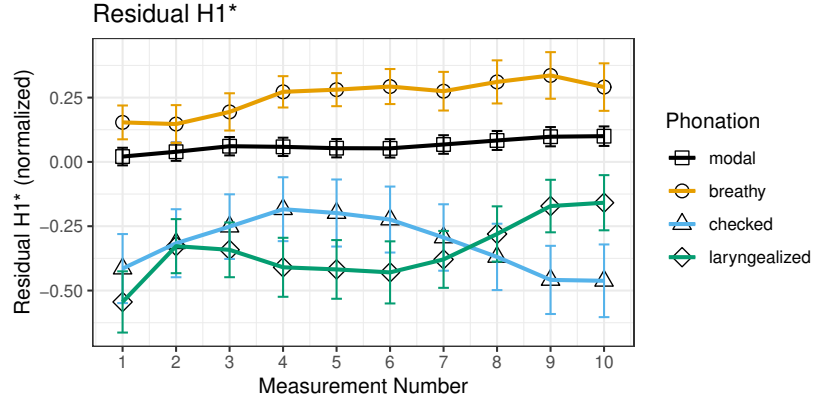


Figure 3.2: Residual $H1^*$ across the duration of the vowel. Points represent the mean of each measure across the ten intervals. The error bars around each point represent a 95% confidence interval. A line was plotted over each to show how the acoustic measure functions across the ten intervals.

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 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 3.2: 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	***

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 $H1^* - H2^*$ 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 non-linear effects present in the data. This was expected because of the dynamic nature of the voice quality in SLZ. Figures 3.3 and 3.4 show the GAMM smooths and difference plots for $H1^* - H2^*$ and residual $H1^*$, respectively. The difference plots in each figure show the difference between the smooths for each voice quality compared to modal.

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.

In Figure 3.3, we see that the different voice qualities are very difficult to observe clearly. In the top left panel, the smooth functions of the GAMM analysis are shown. In it 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. The difference plot in the top right shows that for breathy vs. modal there was no significant difference between the two voice qualities in terms of $H1^* - H2^*$. In the bottom left, the difference plot shows that the differences we observe between checked and modal voice is significant across the entire duration of the vowel. The difference plot in the bottom right shows that there is a significant difference between rearticulated and modal voice in the first two thirds of the vowel, but not in the final third of the vowel.

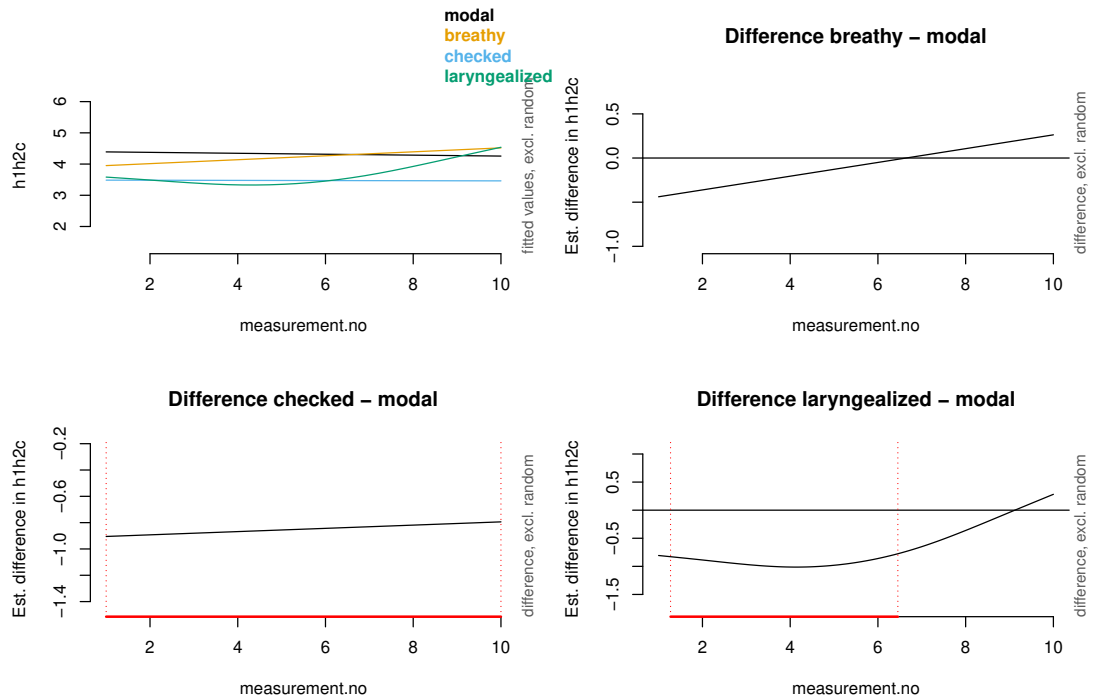


Figure 3.3: GAMM smooths and difference plots for $H1^* - H2^*$ across the duration of the vowel.

The residual $H1^*$ GAMM showed a clearer distinction between the different voice qualities than $H1^* - H2^*$. This is especially clear in the smooth plot where the voice quality distinctions match with the predictions stated above. The difference plot in the top right shows that breathy and modal voice are only significantly different in the final two thirds of the vowel. The bottom left shows that checked and modal voice are significantly different across the entire duration of the vowel. Finally, the bottom right shows that rearticulated and modal voice are significantly different across the entire duration of the vowel.

In comparing the two GAMM analyses, we find that the residual $H1^*$ model pro-

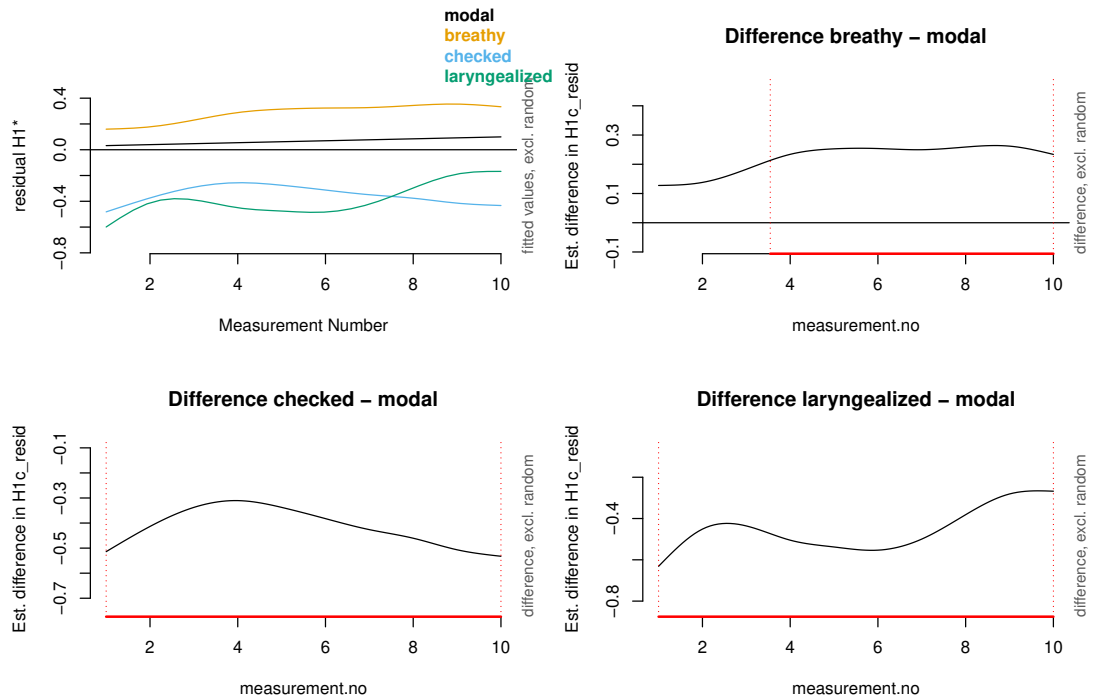


Figure 3.4: GAMM smooths and difference plots for residual $H1^*$ across the duration of the vowel.

vides 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 breathy voice is only significantly different from modal in the final two thirds of the vowel. This suggests that breathy voice is aligned with the end of the vowel. This is similar to the predictions made by Silverman (1997a,b) about the alignment of non-modal phonation in laryngeally complex

languages. This will be further discussed in Chapters 6 and 7.

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 esti-

mate 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.

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.

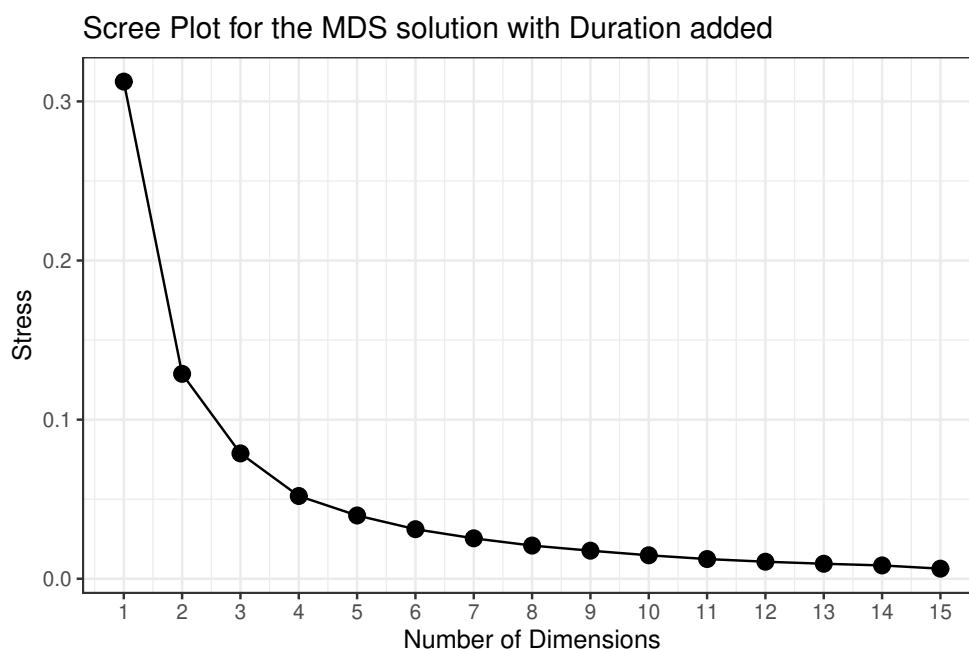


Figure 4.1: Scree plot showing the stress for each dimension for the MDS analysis.

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

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

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.

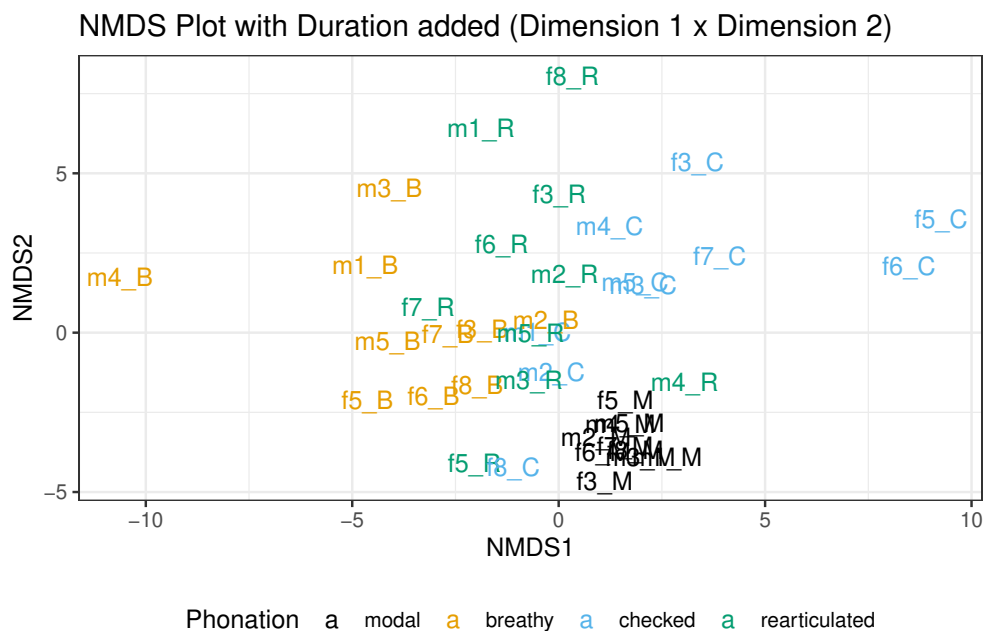


Figure 4.2: Two-dimensional MDS solution showing the first and second dimensions.

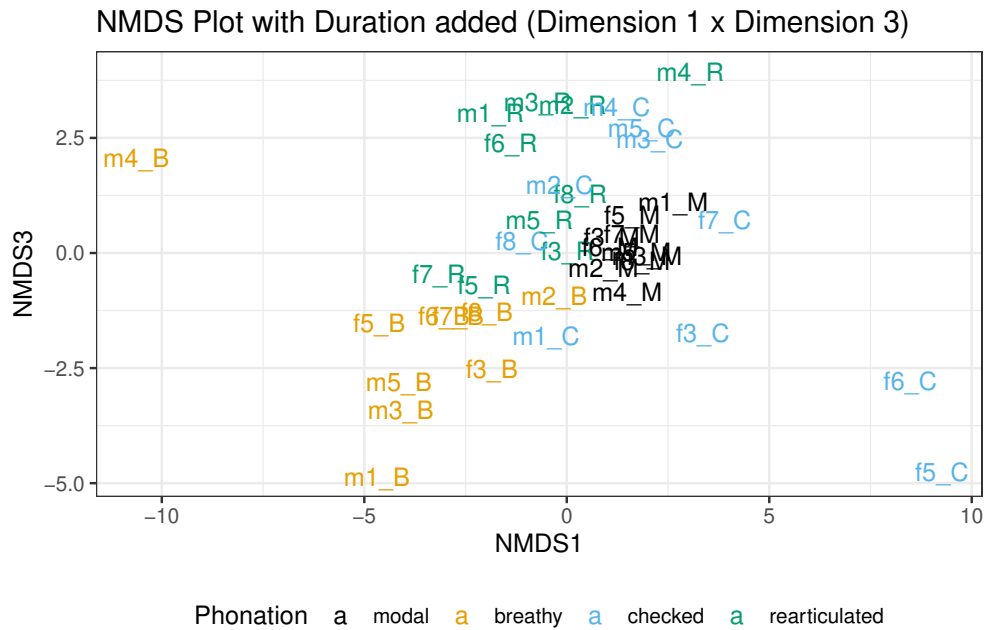


Figure 4.3: Two-dimensional MDS solution showing the first and third dimensions.

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

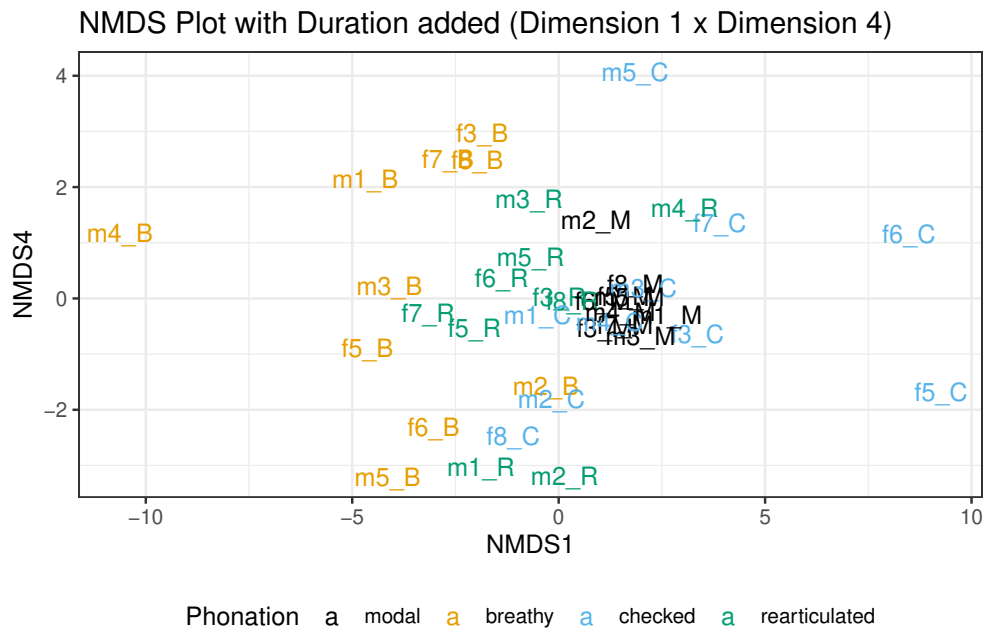


Figure 4.4: Two-dimensional MDS solution showing the first and fourth dimensions.

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

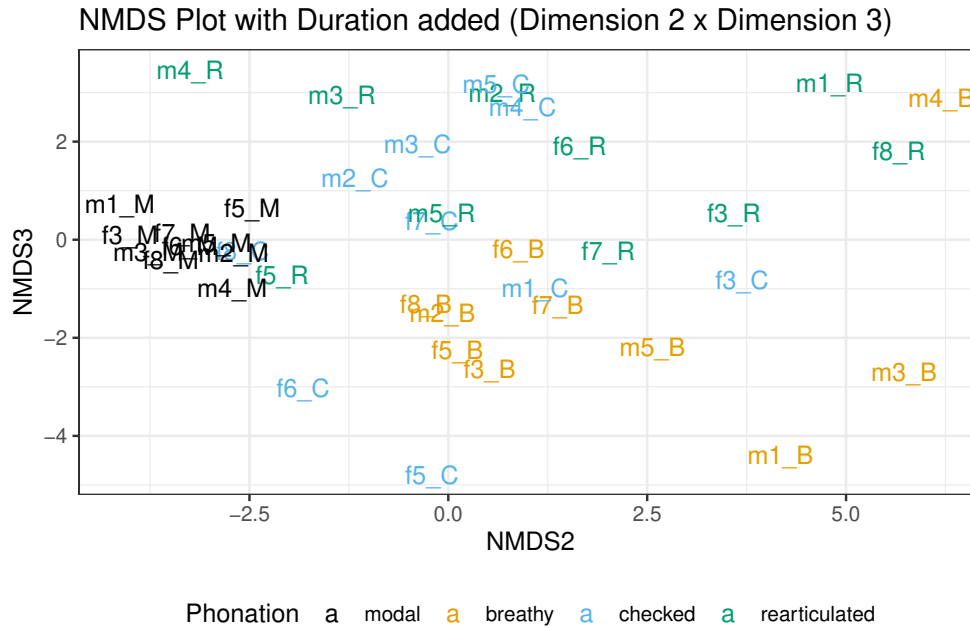


Figure 4.5: Two-dimensional MDS solution showing the second and third dimensions.

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 centralized 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,

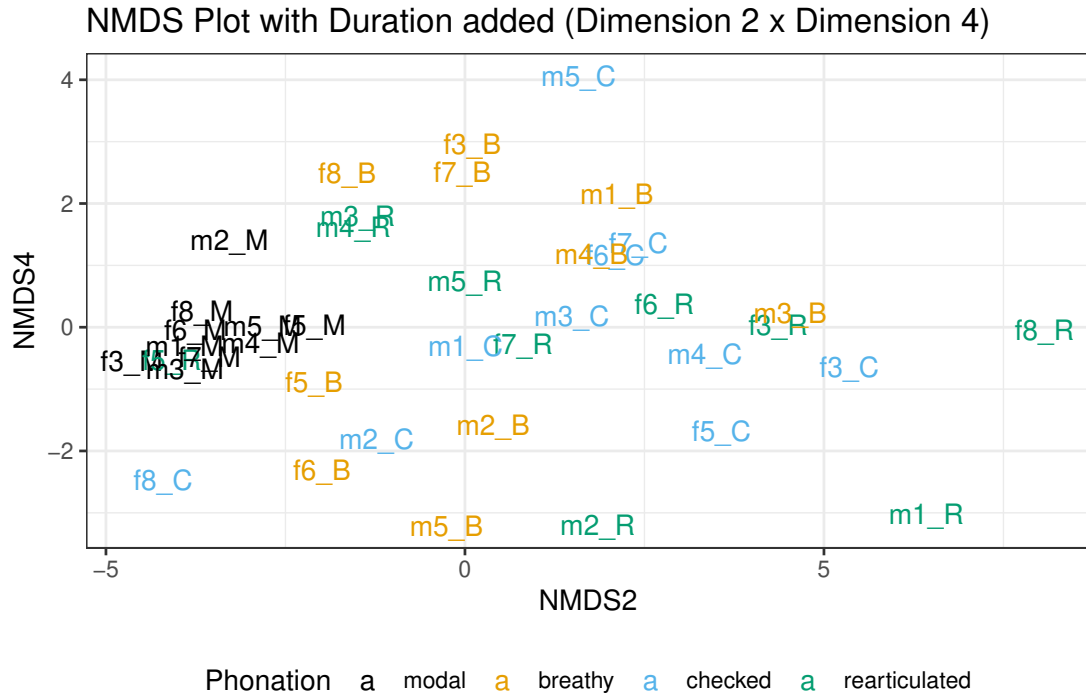


Figure 4.6: Two-dimensional MDS solution showing the second and fourth dimensions.

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-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

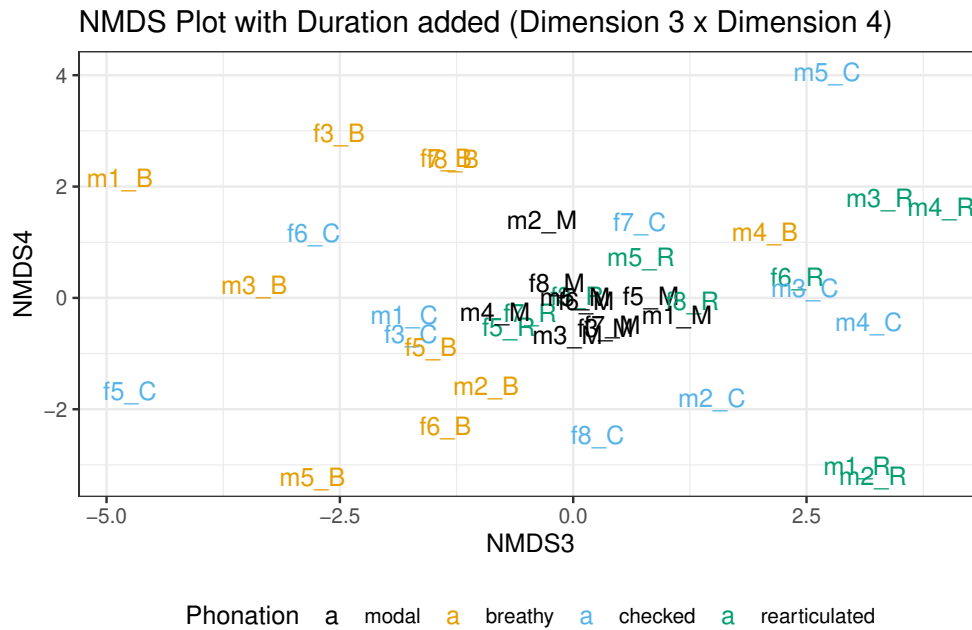


Figure 4.7: Two-dimensional MDS solution showing the third and fourth dimensions.



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

are all measures of the amount of noise present in the signal across various bandwidths (de Krom 1993, Hillenbrand & Houde 1996, 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 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 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.

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 di-

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

mension, 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 there 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.

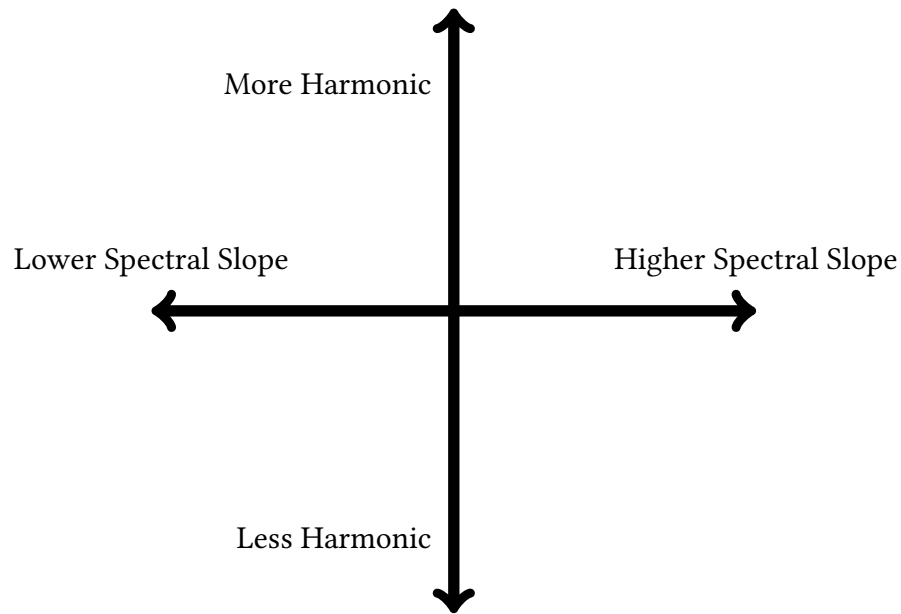


Figure 4.10: A two-dimensional representation of the acoustic space of voice quality in SLZ. The horizontal axis represents the spectral slope of the signal, while the vertical axis represents the amount of noise or energy in the signal.

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 the 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).

I show in this chapter 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 on the statistics and how to perform these analyses 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 of all potential subsets. Meaning that 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 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 the 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 Hz to 500 Hz for the middle third of each vowel. This split is made at the z-score of 0.48. If the value of the predictor is greater or equal to 0.48, the dominate 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

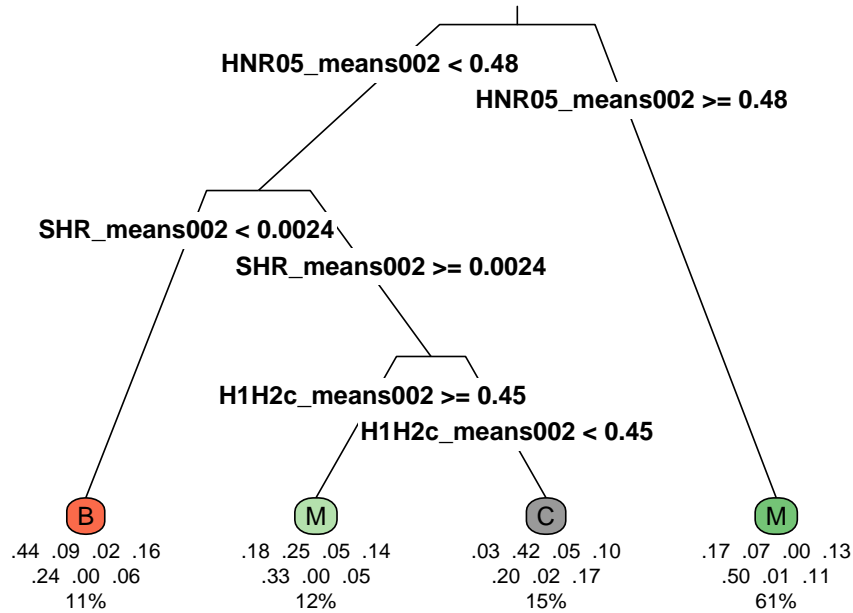


Figure 5.1: Classification tree of phonation categories from Keating et al. (2023). Abbreviations used in this figure are: HNR05_means002: harmonics-to-noise ratio over the frequency range from 0 Hz to 500 Hz for the middle third of each vowel; SHR_means002: subharmonic-to-harmonic ratio for the middle 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 measure, one can classify the voice quality of the data. This is a powerful tool for understanding the importance of the different acoustic measures in the acoustic space.

5.4 Growing a forest of decision trees

Simple decision trees, however, 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 that it was trained on. 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 discussion). The following two 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, number of predictors that are used. Bagging trees use all of the predictors at each split while Random Forests only use a random subset of the predictors at 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 help 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 using bagging, because of the large numbers 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 making the splits in the data. There are two measures that are commonly used to understand variable importance in bagging trees: 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 that the RSS is decreased due to the splits over a given predictor is recorded and averaged across all the trees. In classification trees, the total amount that the Gini

index is decreased for each predictor is recorded and averaged across all the 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 needed 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 that are 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 that is chosen.

5.4.2 Random Forests

However, subsequent research by Breiman (2001), showed that bagging is 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 of the predictors every time, we should only consider a random subset of the predictors at each split. This means that bagging and Random Forests are essentially the exact same thing except in 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 need to be considered at each split (i.e., 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 & 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 and ≈ 27 predictors for regression trees. 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 that is 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 make the splits in the data. This is not possible in bagging and random forests because of the large number of trees that are grown that are then used to average the predictions. Instead, we use variable importance plots to show how important a variable is across all of the trees in making the splits in the data. 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 of the 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 we are concerned 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 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 making the splits in the data.

Tagliamonte & Baayen (2012) made use of both the Gini Index and the permutation importance 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 plot that for the impurity importance (i.e., Gini index) the most important predictors are the ones that are at the top of the plot. 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, and so on. 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, and so on.

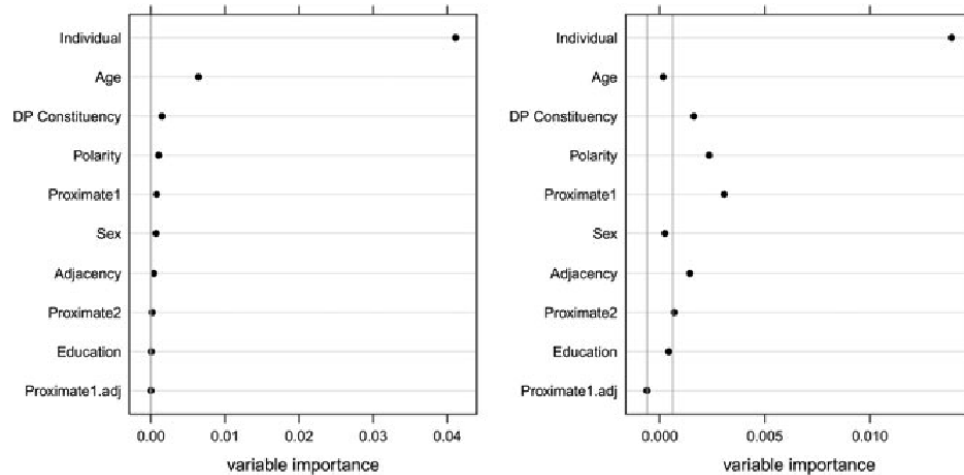


Figure 5.2: Variable importance plot from Tagliamonte & Baayen (2012). The plot on the left shows the impurity importance (i.e., Gini index) and the plot on the right shows the permutation importance. The y-axis shows the different predictors for both plots.

By looking at both the impurity importance and the 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 in 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 data from SLZ. The goal of this analysis is to understand which acoustic measures are the most important in making the splits in the acoustic space of SLZ. The methods for the data collection and analysis 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

The 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* [ʃn:ia’ X tʃone ras] “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

These 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 the 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 analyses in the dissertation. 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.

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.

I then 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 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 through 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 inaccurately 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 is using all of the predictors at each split and is therefore a bagging model. The other values of m_{try} are potential values corresponding to 5%, 15%, 25%, and 40% of the total number of predictors. Additionally, 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 out-of-bag error percentage when compared with the other models. The model that performed the best was the model that was trained on 300 trees and whose $m_{try} = 5$ had the lowest out-of-bag error percentages. Additionally it was found that sampling only 80% of the data with replacement produced the best model.¹

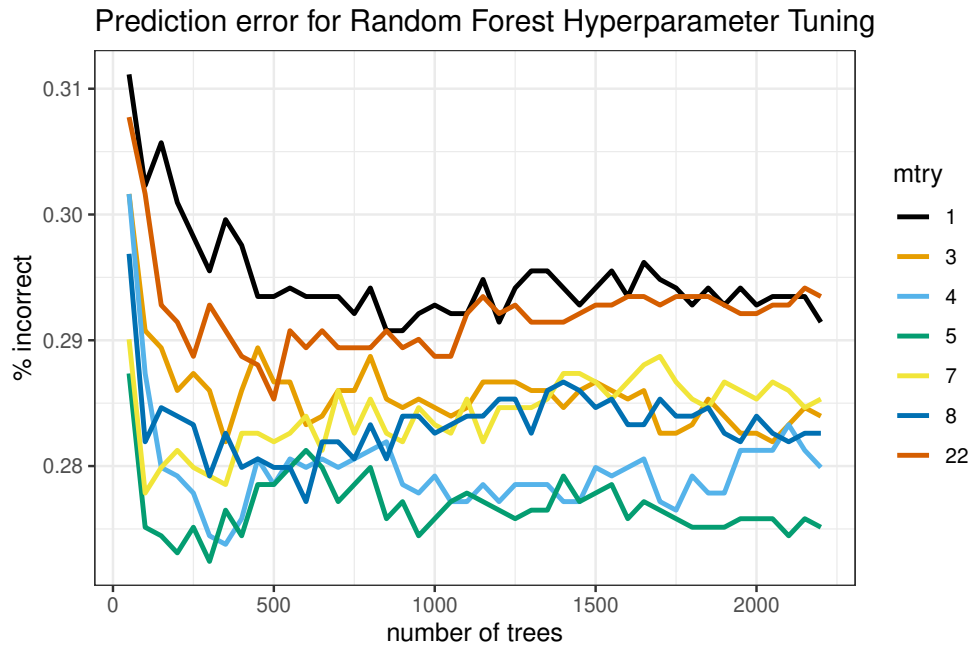


Figure 5.3: Plot showing the percent of inaccurately classified phonation types as a function of the number of trees ran. The different colored lines indicate the different m_{try} values.

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

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 Gini index 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 of the 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 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 making the splits in 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.

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.

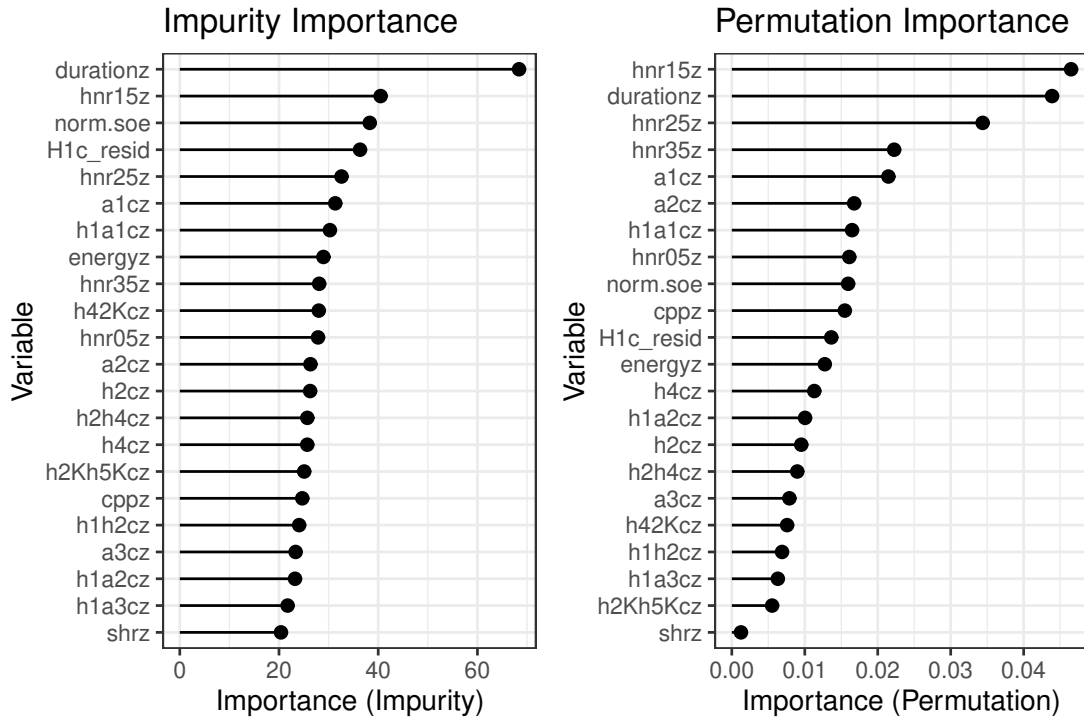


Figure 5.4: Variable importance plots showing the impurity importance and permutation importance of each acoustic measure.

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.

Additionally, 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. This same reasoning holds for $H1^*-A1^*$ which is the seventh most important predictor across 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 the 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. Additionally, I will discuss duration and $A1^*$ because of how important

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

the Random Forest model showed these measures to be. 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 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.

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 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 cross-linguistically (Gao & Kuang 2022, Chai & Ye 2022).

In SLZ, the distribution of duration across 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 breathy voice has the longest duration and rearticulated voice have the longest durations, followed by checked voice. The shortest duration is found in modal voice.

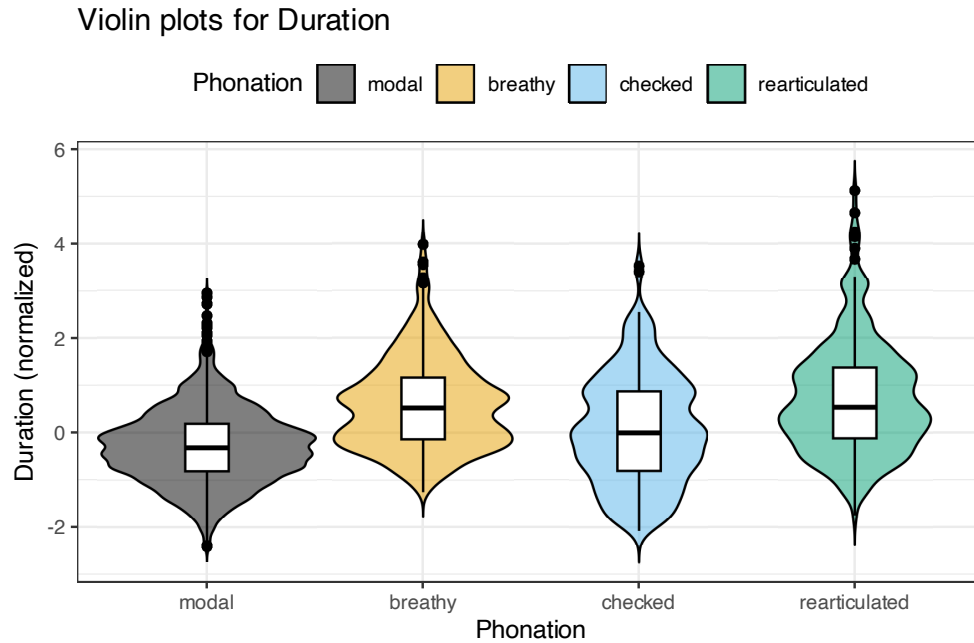


Figure 5.5: Plot showing the distribution of duration across the different voice qualities in SLZ.

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 word final open 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 &

5.7.3 Importance of A1*

The measure that was found to be the most important for classifying SLZ's 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 as Figure 5.6, we see that 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 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).

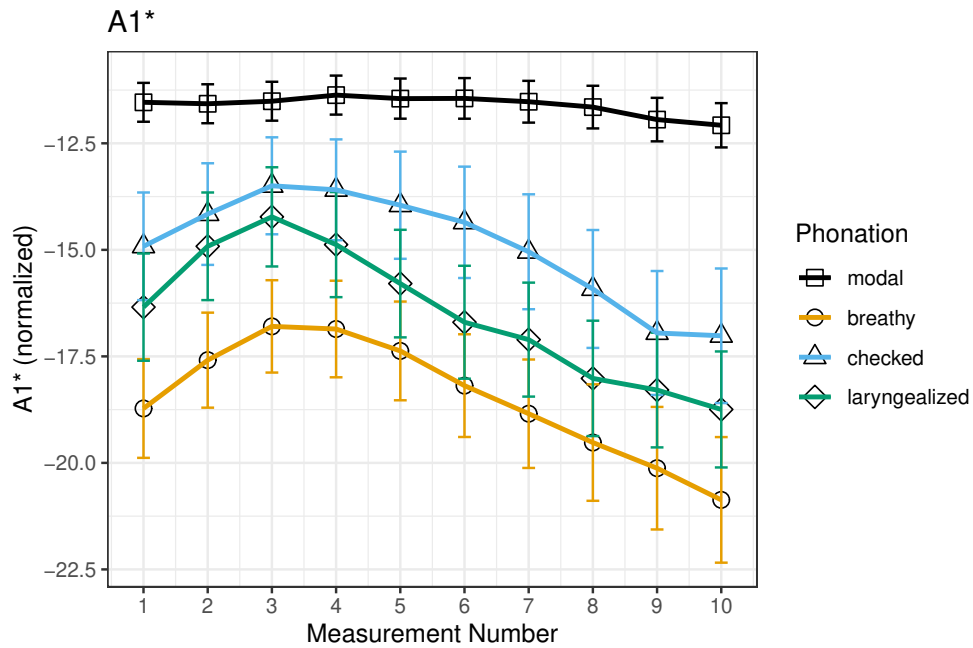


Figure 5.6: Plot showing the distribution of A1* across the different voice qualities in SLZ.

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 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

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

discuss the amplitude of the first formant but rather the frequency of the first formant.

Another comparison can be found in nasality research, where this measure is discussed quite 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 ones, 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 the 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 in nasalization in three different Yi languages were associated with breathy voice. Leading the authors to suggest that breathy voice during nasalization can arise from misperception or as a type of phonetic enhancement.

In regards to 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 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 sub-glottal resonances which is much harder to test than the other two possibilities.

5.7.4 Importance of $H1^* - A1^*$

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 under the category of spectral slope measures. As previously discussed in Chapter 3, these measures capture the differences in the amplitude of the harmonics of the spectrum. The higher the spectral slopes' values

are the more breathy the phonation is (Fischer-Jørgensen 1968).

In the Random Forest analysis, $H1^* - A1^*$ was found 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 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 across the duration of the vowel. The plot shows that breathy voice has the highest $H1^* - A1^*$ score and checked voice has the lowest $H1^* - A1^*$ score. This is consistent with what we would expect from a spectral slope measure.

5.7.5 Importance of Residual $H1^*$

As residual $H1^*$ is a type of spectral slope measure it is expected to show similar behavior to other spectral slope measures. Namely, breathy voice should be associated with a higher score than 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

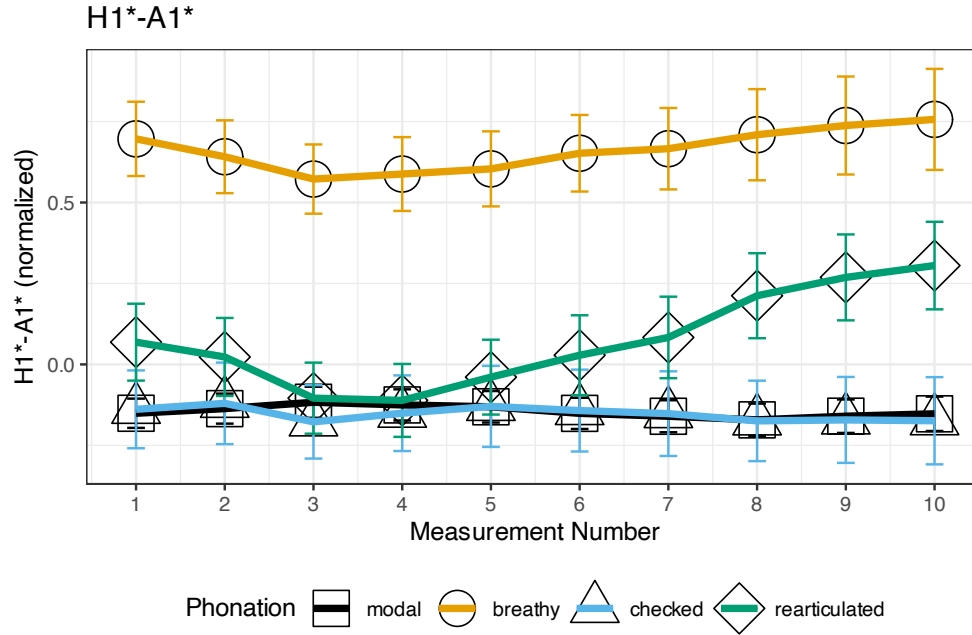


Figure 5.7: Plot showing the distribution of $H1^*-A1^*$ across the different voice qualities in SLZ.

the vowel.

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

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

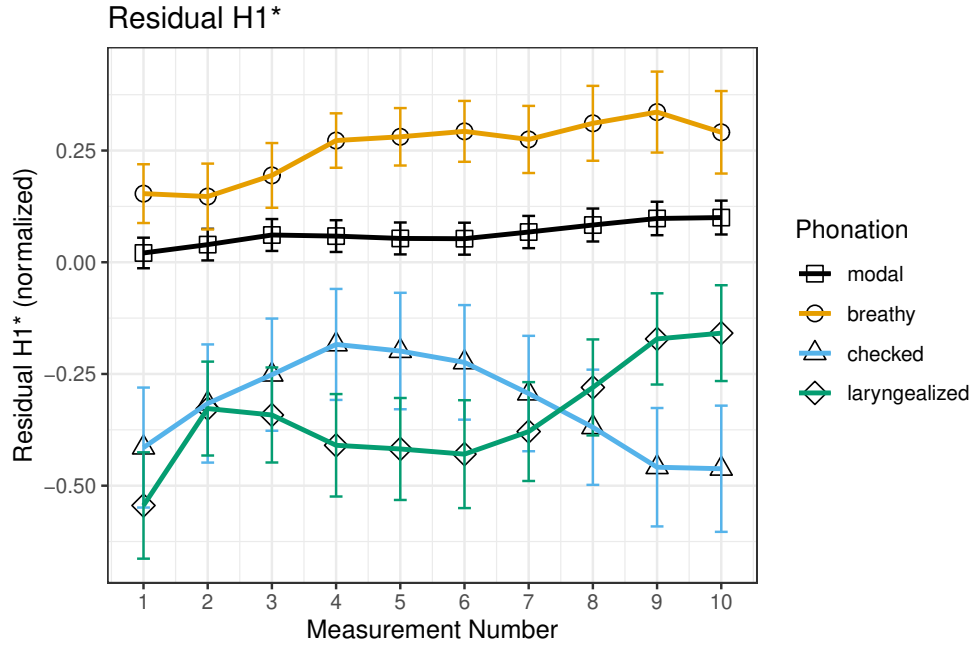


Figure 5.8: Plot showing the distribution of residual $H1^*$ across the different voice qualities in SLZ.

is no significant difference between breathy and modal voice. Suggesting that breathy voice is more closely associated with the latter part of the vowel than 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 measure of periodicity. In these measures, modal voice is associated with a higher score than non-modal phonations. This is because modal voice is associated with a higher degree of periodicity than non-

modal phonations as a-periodicity is a defining feature of non-modal phonations (e.g., Hillenbrand & Houde 1996, 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 the non-modal phonations, breathy and rearticulated voice have a higher HNR < 1500 Hz than checked voice.

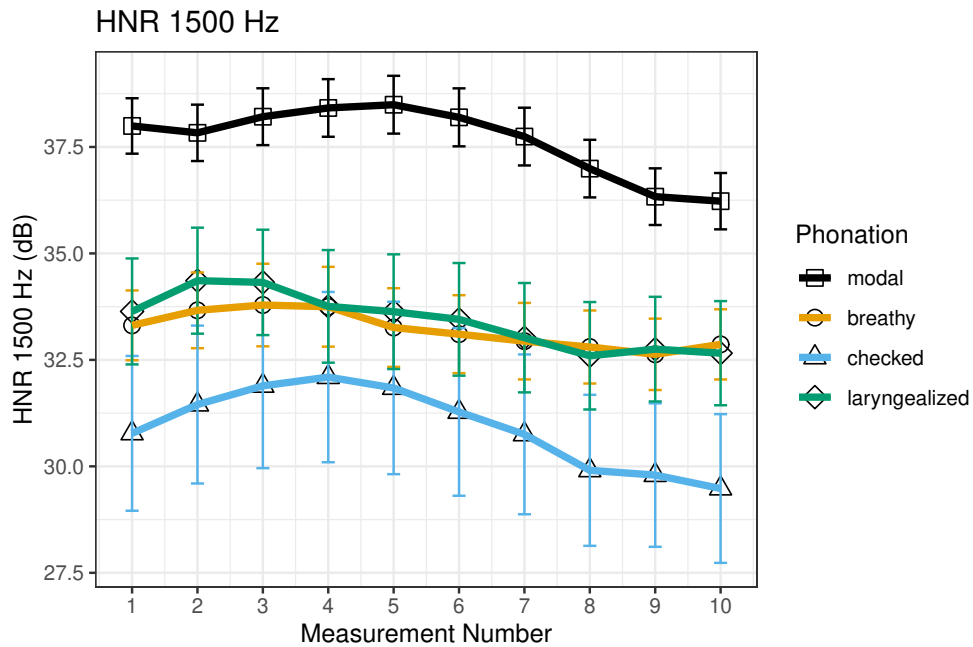


Figure 5.9: Plot showing the distribution of HNR < 1500 Hz across the different voice qualities in SLZ.

The fact that breathy and rearticulated voice have a higher HNR < 1500 Hz than checked voice is not surprising. As discussed in Chapter 2, breathy and rearticulated 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 typically the shortest of the non-modal phonations and is associated 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 $\text{HNR} < 1500 \text{ Hz}$ of the non-modal phonations.

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

5.7.7 Importance of Strength of Excitation

5.8 Conclusion

In conclusion, the results of the bagging analysis show that the acoustic measures $A1^*$, residual $H1^*$, and $\text{HNR} < 1500 \text{ Hz}$ are the most important in making the splits in the acoustic space of SLZ. These three measures are also the measures that were found to be the important in the MDS analysis in Chapter 4. This suggests that these three measures play an important role in defining and characterizing the different voice qualities from one another in SLZ.

Chapter 6

Testing the laryngeal complexity in SLZ

6.1 What is Laryngeal Complexity?

Laryngeal complexity is defined as the contrastive use of tone and phonation on the same syllabic nucleus (Silverman 1997a,b). This use of both tone and phonation on the same nucleus 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 or pitch in languages outside of the Oto-Manguean languages; such as the Tibeto-Burman languages of Mpi and Tamang (Silverman 1997a,b) and the Mayan language Yucatec Mayan (Frazier 2013).

6.1.1 Phasing and recoverability

According to Silverman (1997a,b), one of the defining aspects of laryngeal complexity is the concept of phasing and recoverability. Under this idea, in laryngeally complex vowels the phonation and tone are phased with respect to one another in way that lends itself to a listener's ability to recover the underlying phonation and tone. In practical terms this means that laryngeally complex vowels are composed of two components: a modal voice portion of the vowel where tone is realized, and a non-modal voice portion of the vowel where phonation is realized. For the researcher that means that there are two distinct portions of the vowel that can be analyzed separately and that need to be analyzed temporally rather than 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 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. This means that the breathiness or creakiness is phased with respect to the tone.

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

6.1.1.1 Sufficient acoustic distance

Silverman (1997a) argues that sufficient acoustic distance is necessary for the recoverability of the phonation and tone. As Silverman explain, listeners do not rely on the fundamental frequency alone to perceive pitch. Instead, listeners use the harmonic spacing and the pulse period in the signal to perceive pitch (Ritsma 1967, Remez & Rubin 1993). For modal phonation, this means that the harmonic spacing and pulse periods are present and encode a salient pitch value. However, during non-modal phonation, the harmonic spacing and pulse periods are often obscured or not present.

For breathy voice, this means that there is a general weakening of the harmonic structure which makes it difficult to recover the pitch by the listener (Silverman 1997b). Creaky voice on the other hand obscures the pulse periods due to its aperiodic and unstable glottal vibration (Ladefoged & Maddieson 1996). This is what was observed in Mazatec where the harmonic structure is gone and the pulses are indiscernible (Kirk, Ladefoged & Ladefoged 1993). Additionally, the perception of pitch is rendered indiscernible when the 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. This means that the phonation and tone must be phased with respect to one another in order for the listener to recover the underlying phonation and tone. I will return to this point in Section 6.5.

6.1.1.2 Sufficient articulatory compatibility

Another important point for the Silverman's theory about laryngeal complexity has to deal 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 with speech production (Lindblom 1983). According to Lindblom (1983), speech gestures can be thought of as distinct motor goals in our speech production system. These goals are achieved by the speaker through the coordination of the articulators with the least amount of effort. This means that that the gestures are coordinated in such a way that they are compatible with one another. This manifests itself either through sequencing or coarticulation of the gestures.

A good example of this comes from nasalization. In nasal contexts, the velum is lowered to allow air to pass through the nasal cavity, creating a nasal sound. This velum lowering gesture is compatible with the gestures needed in the oral cavity to produce different vowel qualities. This is what leads to the production of nasal vowels in languages like French or Portuguese. Additionally, 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. This lowering of the velum is compatible with the gestures needed to produce the vowels in the oral cavity (e.g., Ohala 1975, Chen 1997, Styler 2015).

According to Silverman's theory of laryngeal complexity, this idea of articulatory

compatibility is also driving the need to phase phonation and tone. In the case of laryngeal complexity this is because it is assumed that both tone and phonation are produced by the larynx, more specifically the vocal folds and the glottis. This comes from early work on phonation and tone. For tone, Ohala (1978) showed that pitch was controlled primarily by the tensing or laxing of the vocal folds which changed the rate at which the vocal folds vibrate. For phonation, it was similarly shown that the amount the vocal folds were held open or closed determined the type of phonation that was produced (Ladefoged & Maddieson 1996). This aspect of phonation was shown in Figure 4.8 and repeated here as Figure 6.1.



Figure 6.1: A diagram showing the relationship between breathy, modal, and creaky phonation types. Based on Gordon & Ladefoged (2001).

For Silverman's 1997 theory of laryngeal complexity, the articulatory mechanisms for tone and phonation are exactly the same which leads to a need to phase the two in order to optimally make use of the same articulatory gestures. However, there is a growing body of literature that shows that tone and phonation is much more complex and is reliant on the entire larynx not just the vocal folds (e.g., Esling et al. 2019). This matter will be picked up again in Section 6.5 and Chapter 7.

6.1.1.3 Optimal auditory salience

6.1.2 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 how 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 case, the laryngealization appears before the vowel. In the postvocalic case, the laryngealization appears after the vowel. In the interrupted case, the laryngealization interrupts a vowel and appears in the middle.

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 prevocalic 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 ? 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	hVɿ, ?Vɿ	—	—
Comaltepec Chinantec	hVɿ, ?Vɿ	Vhɿ, V?ɿ	—
Copala Trique	hVɿ, ?Vɿ	Vhɿ, V?ɿ	VhVɿ, V?Vɿ

For many descriptions of languages with laryngeal complexity, the implicational hierarchy seems to hold. This is certainly 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 by Frazier (2013), it is not clear how accurate or robust this implicational hierarchy actually is. The reason for this is because it is not always clear if something is a laryngeal consonant or a laryngeal feature on the vowel. Indeed, Silverman (1997a,b) treats laryngeal consonants and laryngeal features on vowels as the same thing. For example, in many of the 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), but 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. This is a rather small sample size and doesn't 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 laryngealization can appear either in the middle of the vowel, what 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, it is also the case in other branches of the Oto-Manguean languages. It is often the case that Zapotec languages only have 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 prevocalic laryngealization in the language. This is in direct contrast laryngeal complexity's implicational hierarchy. This is also true for laryngealization in Santa Ana del Valle Zapotec (Esposito 2004, 2012). This is also true for Santiago Laxopa Zapotec, where there is no prevocalic laryngealization despite having both interrupted and postvocalic laryngealization, see Chapter 2 for more detailed information.

6.2 Previous analyses of laryngeal complexity

Previous analyses about laryngeal complexity fall into two categories: (i) descriptive studies and (ii) instrumental studies. In most descriptive studies, the focus has been on describing the patterns for tone and voice quality and how they interact with

one another. 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 vowel that have high tone with glottalized that variably surface as either rearticulated vowels¹ or as a vowel with creaky voice. Frazier notes that for most speakers these vowels show clear evidence of phasing between the tone and the voice quality. In these vowels the first portion of the vowel is always modal and produced with the high tone. The second portion of the vowel is produced with creaky voice which greatly 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 (Fischer-Jørgensen 1989, Grønnum, Vazquez-Larruscaín & Basbøll 2013, Peña 2022, 2024). In Danish, there is a phonation contrast that exists between modal voice and a type of creaky voice which is called *stød*. Research has shown that *stød* is also associated with a secondary cue of a heightened *f*₀ (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 syllabic rhyme and the

¹This 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.

secondary cue of pitch being produced in the first half of the syllabic rhyme.

In contrast to the descriptive studies, instrumental studies have focused on the acoustic properties of laryngeal complexity, primarily how laryngealization affects f_0 and the harmonic structure of the vowel. For example, Garellek & Keating (2011) found that in Jalapa Mazatec, the laryngealization causes f_0 perturbation in the first portion 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 f_0 perturbation has been shown to be the case with laryngealization in some varieties of Trique, again showing that there is phasing between modal and non-modal phonation (DiCanio 2012a). For these previous studies the main piece of evidence for laryngeal complexity has been the perturbation of the f_0 signal in the portion of the vowel that is affected by the non-modal phonation.

In more recent studies, researchers have also looked at other measures besides f_0 to determine if there is phasing. For example Weller et al. (2023a,b, 2024) have investigated laryngeal complexity in San Sebastián del Monte Mixtec using a combination of f_0 and Strength of Excitation (SoE), a measure that correlates to the strength of voicing, measures. They found that there is a clear phasing between modal and non-modal phonation in the rearticulated vowels of the language.

However, these accounts only offer a limited window into the question of laryngeal

complexity. Because there has been a focus on determining whether or not there are f_0 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 by showing that by looking at SoE, in addition to f_0 , we can get a better understanding of the phasing between tone and voice quality. However, there is still a need to look at other acoustic properties of the signal to get a full understanding of laryngeal complexity.

The class of harmonic-to-noise ratio measures is one such class that promises to give this 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 particularly helpful in determining whether or not 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 non-modal phonation (Ladefoged & Maddieson 1996). This means that harmonic-to-noise ratio measures can be used to determine if there is aperiodicity in the signal and if there is a clear phasing between modal and non-modal phonation.

The rest of this chapter will focus on analyzing SoE, a measure of the strength of voicing, and two harmonic-to-noise ratio measures, Harmonic-to-Noise Ratio (HNR) < 1500 Hz, which measures the noise in frequency band of 0 Hz to 1500 Hz, and Cepstral Peak Prominence (CPP), a measure of noise across the entire spectrum, in order to

determine if there is a clear phasing between modal and non-modal phonation in SLZ using a generalized additive mixed model (GAMM) analysis (Hastie & Tibshirani 1986, Wood 2017, Sóskuthy 2017, Wieling 2018).

6.3 Analysis of laryngeal complexity

6.4 Results

6.5 Discussion

Humbert (1978)

6.6 Conclusion

Chapter 7

Modeling laryngeal complexity

7.1 The Laryngeal Articulator Model

Esling (2005), Esling et al. (2019), Moisik, Czaykowska-Higgins & Esling (2021),
Moisik et al. (2015), and Moisik & Esling (2014)

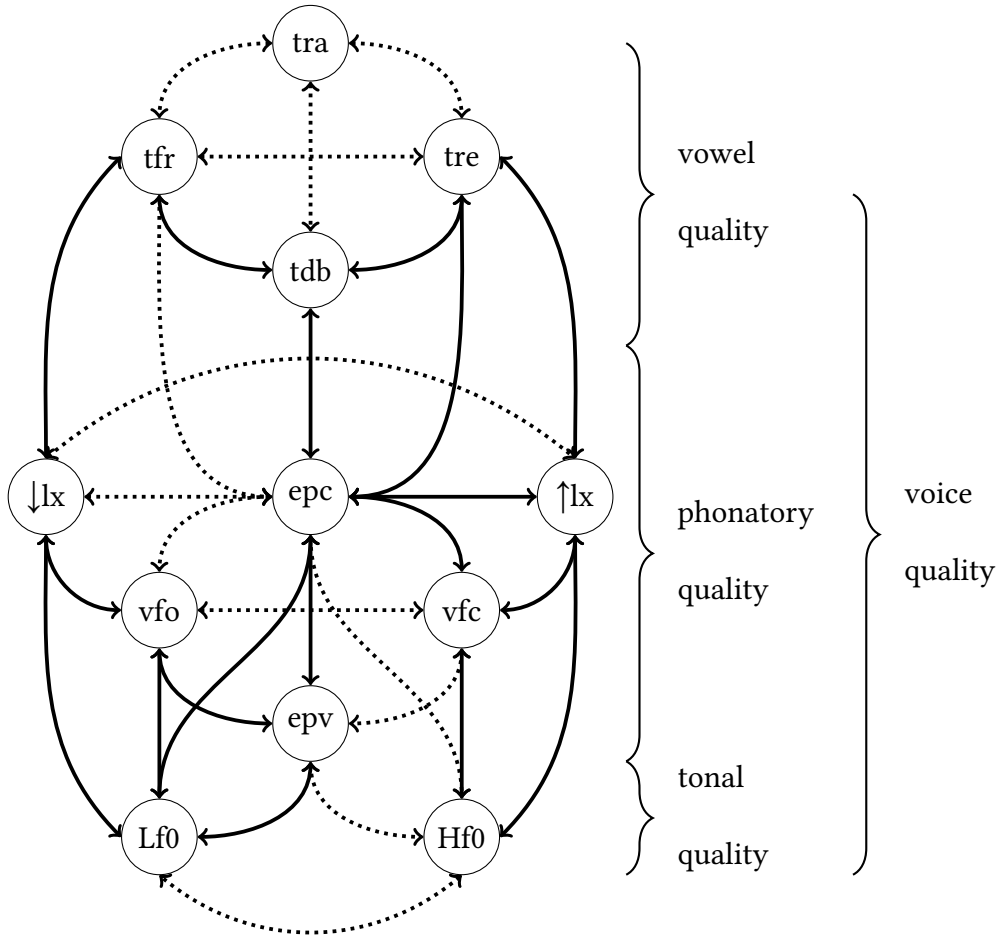


Figure 7.1: The Laryngeal Articulator Model from Esling et al. (2019). This model shows the interactions between the laryngeal articulators (labeled circles). Synergetic interactions are shown with solid lines, while anti-synergetic interactions are shown with dotted lines.

7.2 Modeling laryngeal complexity

7.3 Alternative accounts

7.3.1 Articulatory Phonology account

7.3.2 Q-theory account

7.3.3 Radical CV Phonology account

7.4 Conclusion

Chapter 8

Conclusion

Bibliography

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 Electroglossographic 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.), *Proceedings 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.
- 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 Santiago 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. 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 & 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. 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. 2025. Perception of checked and rearticulated phonations: Effect of duration and glottalization phasing.
- 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 Quiavini 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>.
- 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>.
- 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. 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>.
- Esling, John H. 2005. There Are No Back Vowels: The Larygeal Articulator Model. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 50(1–4). 13–44. <https://doi.org/10.1017/S0008413100003650>.

- 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. 2010. 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. 2012. Contrastive breathiness across consonants and vowels: A comparative study of Gujarati and White Hmong. *Journal of the International Phonetic Association* 42(2). 123–143. <https://doi.org/10.1017/S0025100312000047>.
- 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>.

- Esposito, Christina M., Sameer ud Dowla Khan, Kelly H. Berkson & Max Nelson. 2019. Distinguishing breathy consonants and vowels in Gujarati. *Journal of South Asian Languages and Linguistics* 6(2). 215–243. <https://doi.org/10.1515/jsall-2019-2011>.
- 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. 2013. *Production and perception of glottal stops*. Los Angeles, CA: University of California, Los Angeles dissertation.
- 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, 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>.

- 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>.
- Hillenbrand, James & Robert A. Houde. 1996. Acoustic Correlates of Breathly 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>.
- 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, Electrolottographic, 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>.
- Humbert, Jean-Marie. 1978. Consonant Types, Vowel Quality, and Tone. In Victoria A. Fromkin (ed.), *Tone*, 77–111. Academic Press. <https://doi.org/10.1016/B978-0-12-267350-4.50008-X>.
- 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, 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>.
- 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\(2007/042\)](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>.

- 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>.

- 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>.
- 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.
- Moisik, Scott R, John H Esling, Lise Crevier-Buchman, Angélique Amelot & Philippe Halimi. 2015. *Multimodal imaging of glottal stop and creaky voice: Evaluating the role of epilaryngeal constriction*. Scottish Consortium (ed.). Vol. paper 247.
- Moisik, Scott R., Ewa Czaykowska-Higgins & John H. Esling. 2021. Phonological potentials and the lower vocal tract. *Journal of the International Phonetic Association* 51(1). 1–35. <https://doi.org/10.1017/S0025100318000403>.
- Moisik, Scott R. & John H. Esling. 2014. Modeling the Biomechanical Influence of Epilaryngeal Stricture on the Vocal Folds: A Low-Dimensional Model of Vocal–Ven-

- tricular Fold Coupling. *Journal of Speech, Language, and Hearing Research* 57(2). S687–S704. https://doi.org/10.1044/2014_JSLHR-S-12-0279.
- Munro, Pamela & Felipe H. Lopez. 1999. *Di'csyonaary x:tèe'n dii'zh sah Sann Lu'uc* (*San Lucas Quiavini Zapotec dictionary/Diccionario Zapoteco de San Lucas Quiavini*). 2 vols. Los Angeles: (UCLA) Chicano Studies Research Center Publications.
- 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. 1978. Production of Tone. In Victoria A. Fromkin (ed.), *Tone: A linguistics survey*, 5–39. New York: Academic Press. <https://doi.org/10.1016/B978-0-12-267350-4.50006-6>.
- 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, Edward Szoecs, Helene Wagner, Matt Barbour, Michael Bedward, Ben Bolker, Daniel Borcard, Gustavo Carvalho, Michael Chirico, Miquel De Cáceres, 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, Jaily M. 2022. Stød Timing and Domain in Danish. *Languages* 7(1). 50. <https://doi.org/10.3390/languages7010050>.
- Peña, Jaily 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 & Arnetta 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.
- 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>.
- 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.
- Smith-Stark, Thomas. 2007. Algunas isoglosas zapotecas. In *Clasificación 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](https://arxiv.org/abs/1703.05339) [stat]. Pre-published.

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

Appendix A

Some Ancillary Stuff

Ancillary material should be put in appendices, which appear after the bibliography.