UNIVERSITY OF CALIFORNIA
SANTA CRUZ

VOICE QUALITY AND TONE AT THE PHONETICS-PHONOLOGY INTERFACE

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

DOCTOR OF PHILOSOPHY

in

LINGUISTICS

by

Mykel Loren Brinkerhoff

June 2025

The Dissertation of Mykel Loren Brinkerhoff
is approved:

Professor Grant McGuire, Chair

Professor Jaye Padgett

Professor Ryan Bennett

Professor Marc Garellek

Peter Biehl
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

List of Tables

Abstract

Voice Quality and Tone at the Phonetics-Phonology Interface

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

When it comes to acknowledgements, there are so many words that could be said, but they often feel awkward and do not actually . There are many people who have helped me along the way, from my family to my friends, to my colleagues and mentors. I am grateful to all of them.

...

Chapter 1

Introduction

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)

Chapter 2

Vowels and suprasegmentals in Santiago Laxopa Zapotec

2.1 Introduction

Santiago Laxopa Zapotec (SLZ; *Dilla'xhunh Laxup* [diʒaˀʐun lːaʂupʰ]) 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 2020). 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 2.2. This type of vowel inventory is very common among Sierra Norte Zapotecs. Most varieties have the vowels /i/, /e/, /a/, and /o/ (Nellis & Hollenbach 1980, Jaeger & Van Valin 1982, Butler H. 1997, Avelino 2004, Long & Cruz 2005, Sonnenschein 2005).

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

Table 2.1: Vowel qualities in Santiago Laxopa Zapotec.

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

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

Additional evidence for the overlap of /o/ and /u/ can be measured with a combination of Pillai scores (Pillai 1955, Hay, Warren & Drager 2006, Nycz & Hall-Lew 2014) and Bhattarcharyya's Affinity (Bhattacharyya 1943, Johnson 2015, Warren 2018, Strelluf 2018). Both of these measures show what degree of overlap exists between two different items in some space. Their use in linguistics has been used mainly to

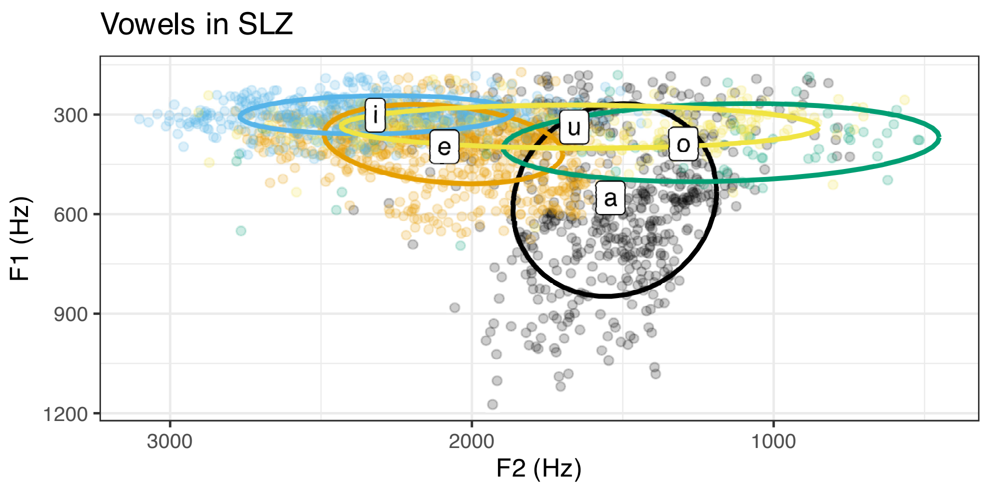

Figure 2.2: Vowel space of Santiago Laxopa Zapotec. The ellipses around each vowel mean represents 1 standard. The scale of the axes are in barks with their corresponding Hz values.

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 Voice Quality contrasts in Santiago Laxopa Zapotec

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$^{\mathrm{L}}$/ ‘tree; wood; almúd (unit of measurement approximately 4kg)’

b. *yah* /ça̤$^{\mathrm{L}}$/ ‘metal; rifle; bell’

c. *yu'* /çuˀᴸ/ 'cooking pot'

d. *ya'a* /çaˀaᴸ/ 'market'

SLZ shares with most Zpaotec varities, 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.

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 varities), while others produce creaky voice throughout the entire vowel. Some speakers produce a combination of these unique productions. Overall, these rearticulated vowels are proiduced with some form of manipulation of glottal closure or apmlitude drop in the middle of the vowel.

[INSERT SPECTROGRAMS OF CHECKED AND REARTICULATED VOWELS]

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

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

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]

2.4 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, 2012, 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

It is well understood that the term voice quality refers to and describes the manner in which the vocal folds vibrate during speech production. Many languages make use of voice quality to convey paralinguistic information by “indexing the biological, psychological, and social characteristics of the speaker” (Laver 1968, Podesva 2016). In addition to using voice quality to convey paralinguistic information, many languages use voice quality to convey phonemic distinctions (Garellek 2019).

It has long been established that voice quality contrasts have correlates in the acoustic signal (Fischer-Jørgensen 1968, Buder 1999, Kent & Ball 1999). For example, Fischer-Jørgensen (1968) found that a strengthened fundamental correlated with breathy voice in Gujarati. In order to normalize the amplitude of the fundamental and counter-act 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). This measure, H1-H2, has since 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 (Chai & Garellek 2022). One of the main problems is that using H2 and other normalizations (e.g., H1-A3) are really just attempts at trying to understand the relative strength of the fundamental to higher harmonic energy. Furthermore, 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). Furthermore, it is not uncommon for researchers to find that H1-H2 is not always the best measure to distinguish voice quality contrasts. For example, Esposito (2010) found that in Santa Ana del Valle Zapotec H1-H2 was only effective in distinguishing the voice quality contrasts in female speakers of the language and male speakers were better distinguished by H1-A3. Furthermore, Garellek & Esposito

(2021) found that the prominence of the cepstral peak, a type of harmonic measurement of noise, was a better measure to distinguish voice quality contrasts in White Hmong than H1-H2 and other measures of the spectral-slope.

Chai & Garellek (2022) found that in addition to the issues mentioned above, errors in measuring H1-H2 is uncomfortably high. This is primarily due to the need to precisely measure two different harmonic amplitudes and when there are errors in calculating H1 this is turn leads to errors in calculating H2 (Arras 1998). An example of this is type of error propagation is that when there are errors in measuring the fundamental frequency, which is especially common with non-modal phonation, errors are introduced into measuring harmonics because they are based on the fundamental. Despite algorithms correcting for vowel height, a common error that occurs with calculating and measuring the fundamental frequency is when a high fundamental frequency co-occurs with a low first formant. This situation causes errors in tracking the fundamental frequency and the first formant. A final issue that can occur with measuring the harmonics is in contexts where the vowel is nasalized. Simpson (2012) shows that in these nasalized context, the first nasal pole (P0) can increase the amplitude of H2 and, when the fundamental frequency is high, H1 is instead increased.

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 & 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*.

For our study, we tested residual H1* with data from Santiago Laxopa Zapotec. This language has a complex interaction between tone and phonation types that has led traditional spectral-tilt measures not adequately to capture the differences in phonation in previous studies. 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 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 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 2020, 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 (*Santiago Laxopa* 2022) and a small number of other speakers in diaspora throughout Mexico and the United States. Similar to other Oto-manguean languages, Santiago Laxopa Zapotec is laryngeally complex, which refers to how these languages make use of contrastive tone and contrastive voice quality (Silverman 1997a,b, Blankenship 1997, 2002).

Santiago Laxopa Zapotec exhibits the standard five-vowel inventory, which is further distinguished by the use of a four-way contrast in voice quality. This variety is unique because it is a Northern Core Zapotec that has developed breathy voice in addition to the two types of laryngealization that characterize the rest of the Zapotec languages, namely checked and rearticulated (see Ariza-García (2018) for a 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).[1] 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.

These interactions between voice quality and tone present a rich environment for

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

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 on a Zoom H4n handheld recorder (16-bit, 44.5 kHz).

The wordlist consisted of 72 items repeated three times each in isolation and the carrier sentence *Shnia' X chonhe lhas* "I say X three times". 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 contained the same voice quality in each syllable. Of those 13, only five contained mixed voicing.

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 of the vowels located in the target word from the elicitation list of the carrier sentences.

These vowels were then extracted and fed into VoiceSauce for acoustic measuring (Shue, Keating & Vicenik 2009). The formants were measured using Snack (Sjölander 2004), while the fundamental frequency (*f0*) was measured using the STRAIGHT algorithm (Kawahara, Cheveigne & Patterson 1998). Spectral slope measures were corrected for formants and bandwidths (Hanson 1997, Iseli, Shue & Alwan 2007). Each vowel was measured with ten equal time intervals, resulting in 22890 data points in total.

The data was cleaned of outliers following the same steps taken by Chai & Garellek (2022) in their study. The H1*, H1*–H2*, and *f0* values were z-scored by speaker to reduce the variation between the speakers and provide a way to directly compare the different measures on the same scale. Data points with an absolute z-score value greater than 3 were considered outliers and excluded from the analyzes. 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. This is comparable to what was done in Garellek & Esposito (2021), Seyfarth & Garellek (2018), and Chai & Ye (2022).

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

After the outliers were removed, 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 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 assigned according to their position in the vowel (first, middle, and third) for statistical modeling.

3.3.3 Statistical modeling

Three linear mixed-effects regression models were fitted, one each for the z-scored H1*-H2* and residual H1*. Each model had tone and interaction between voice quality and position in the vowel as fixed effects, and vowel and interaction between speaker, word, and repetition as random intercepts.

$$Measure \sim Phonation{*}Position{+}Tone{+}(1|Speaker : Word : Repetition){+}(1|Vowel) \quad (3.1)$$

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, 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 that 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.[2]

The interaction between speaker, word, and repetition was treated as a random intercept because this allows us to take into account that each speaker said each word on the elicitation list three times. This intercept accounts for not only the intra-speaker

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

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.

3.4 Results

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 all have values lower than the modal at each of the first nine intervals. In the final interval, 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.

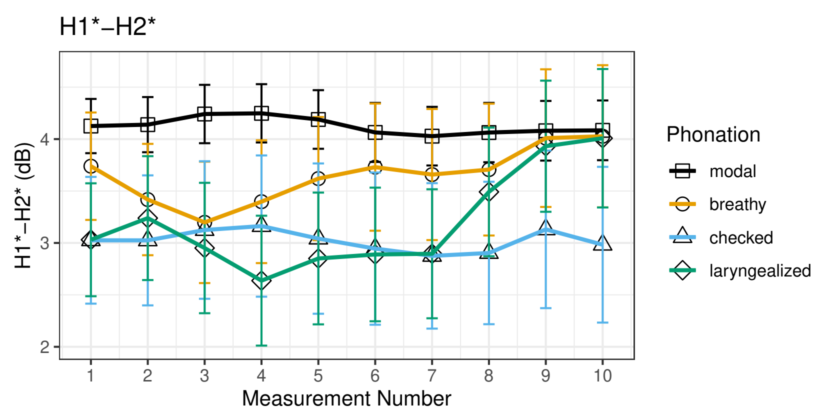

Figure 3.1: H1*-H2* across the duration of the vowel. Points represent the mean of each measure across the ten intervals. The error bars around each point represent ±1.96 Std. Error. A line was plotted over each to show how the acoustic measure functions across the ten intervals.

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.

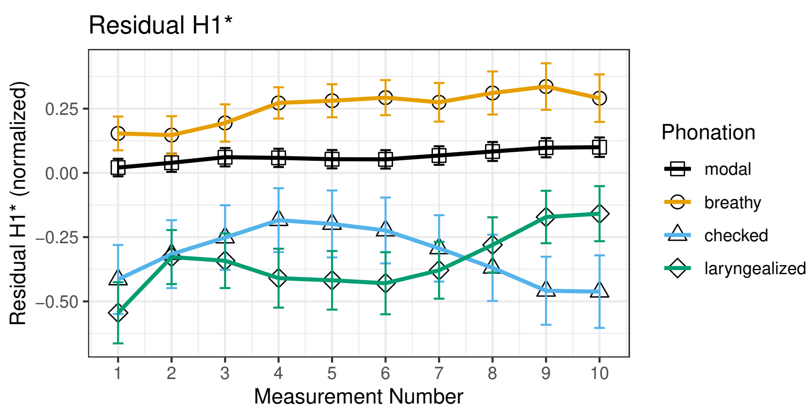

Figure 3.2: Residual H1* across the duration of the vowel. Points represent the mean of each measure across the ten intervals. The error bars around each point represent ±1.96 Std. Error. A line was plotted over each to show how the acoustic measure functions across the ten intervals.

3.4.3 Model Comparison

In order to asses the robustness of the models we compared the residual H1* linear mixed-effects model to the H1*-H2* linear mixed-effects model. This was carried out 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 3.1 shows the results of the comparison of 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 3.1: 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*	-0.03569	0.04210	-0.84781	0.39656	
	Res. H1*	0.14997	0.03175	4.72315	<0.001	***
Checked vs Modal	H1*-H2*	-0.14120	0.04050	-3.48623	<0.001	***
	Res. H1*	-0.38554	0.03061	-12.59335	<0.001	***
Rearticulated vs Modal	H1*-H2	-0.13719	0.04964	-2.76340	0.00574	**
	Res. H1*	-0.48437	0.03740	-12.95239	<0.001	***

Table 3.2 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. Even though AIC comparison is usually conducted 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 3.2: AIC for the H1*-H2* and residual H1* models.

Model	AIC	Δ AIC
H1*-H2* model	43386.99	11214.54
Residual H1* model	32172.45	0

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 the use of residual H1* in voice quality research, especially in laryngeally complex languages.

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 underdocumented 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

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' [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

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, Keating & Vicenik 2009) 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).

The STRAIGHT algorithm estimates 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 same scale.

4.2.4 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 data 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.

After removing the outliers, I 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* was the product of the z-scored energy and the energy factor subtracted from it.

Once these steps were completed, the mean of each vowel and speaker of the fifth and sixth intervals was taken. 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

Using a multidimensional scaling (MDS) analysis is a statistical method of reducing the dimensionality of a dataset to visualize the relationships between the data points (Kruskal & Wish 1978). This is especially true when many variables could contribute to the data. In the case of voice quality, this is especially true. 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 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 x voice quality interaction. This allows us to see how speakers differ in their production of the different voice qualities. This choice to focus on speaker x voice quality 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 speaker 1's modal voice, one for speaker one's checked voice, one for speaker one's rearticulated voice, and one for speaker one's 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 x voice quality interaction. Both of these interactions show us similar information. One shows us within a language, while the other shows us between languages.

The MDS analysis was conducted in R using the 'metaMDS' function in the 'vegan' package. The Manhattan distance was used to estimate the physical differences between the speaker x voice quality pairs. Because the distances are non-Euclidean, the MDS analysis was conducted using the non-metric 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. The elbow of the curve was identified as the correct number of dimensions for analysis. Figure 4.1 shows that most data is captured in a two-dimensional space. The third dimension adds more subtle information about the voice quality.

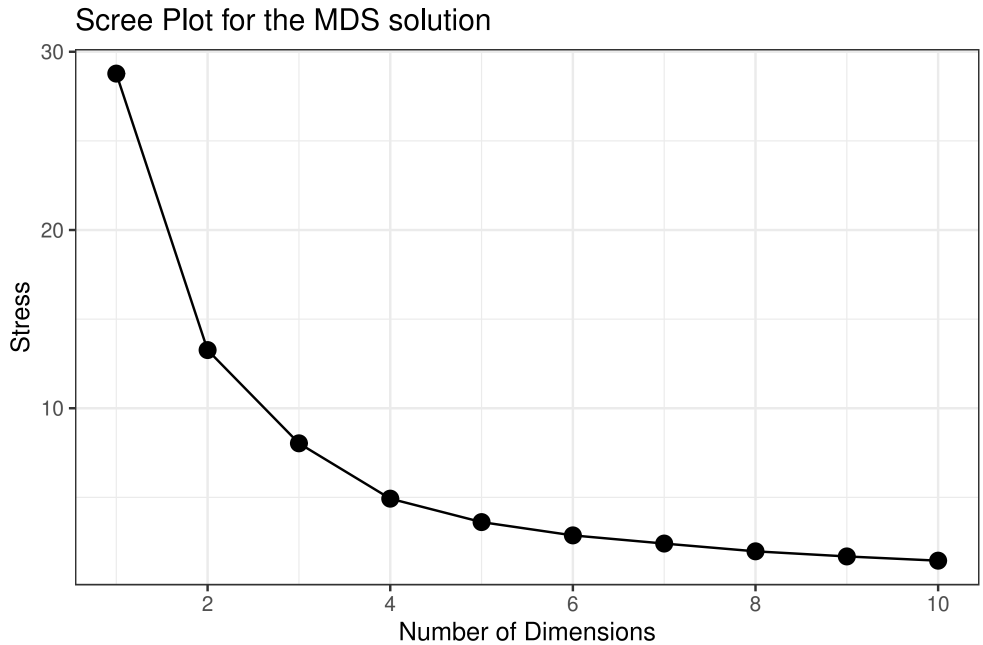

Figure 4.1: Scree plot for the MDS analysis.

4.3 Results

4.3.1 Acoustic space of voice quality

As mentioned above, the results of the MDS analysis can be represented in a two-dimensional space, as shown in Figure 4.2. In this and all subsequent plots, the breathy voice is represented by black, checked voice with orange, rearticulated voice with green, and modal voice with blue. Overall, we see that the breathy voice is located to the left of the plot, checked and rearticulated voices are tending to the right, and the modal voice is located in the center along the first dimension. The second dimension shows a modal and nonmodal split, with modal voice at the bottom of the plot and nonmodal voice at the top.

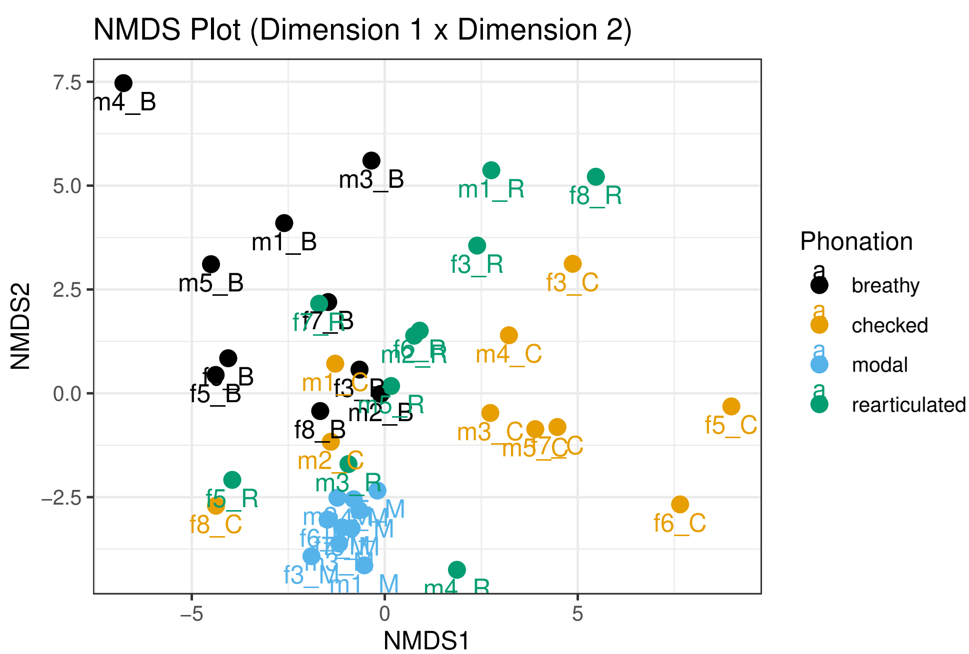

Figure 4.2: Two-dimensional MDS solution showing the first and second dimensions.

As mentioned above, the third dimension adds more information about voice quality. Adding the third dimension helps spread the groups along the first dimension, as shown in Figure 4.3. We see that breathy vowels are located at the top of the plot, and the two types of creaky voices (checked and rearticulated) are at the bottom.

When adding the third dimension to the second, we see that the breathy voices become separated from the other nonmodal voice qualities, as shown in Figure 4.4.

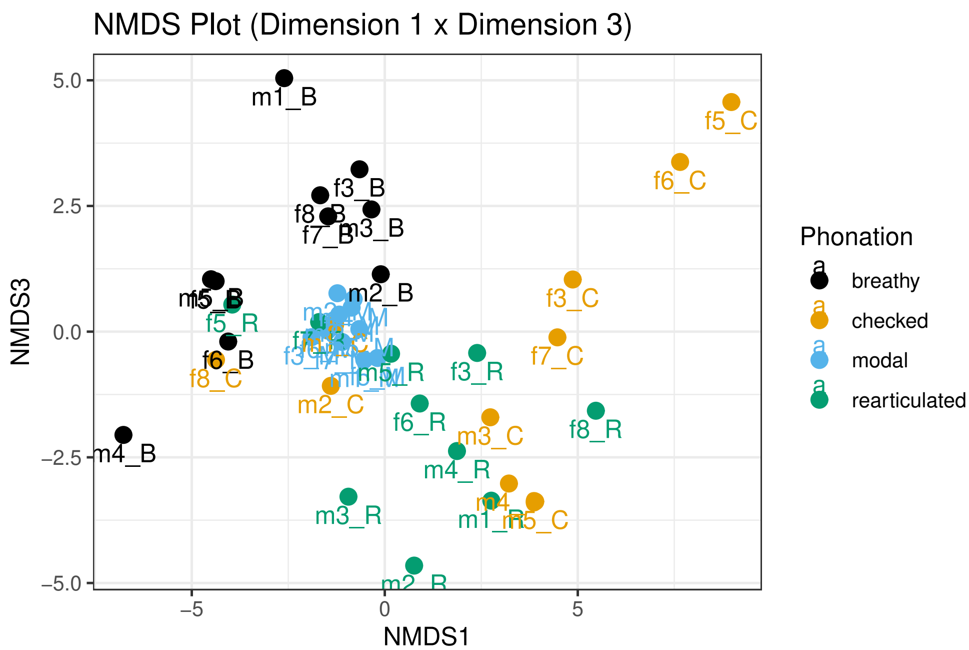

Figure 4.3: Two-dimensional MDS solution showing the first and third dimensions.

4.3.2 Acoustic correlates of voice quality

An additional step to MDS analysis involves testing which acoustic measures contribute the most weight to the different dimensions. Table 4.1 shows the results of this test. In each of the three dimensions of the MDS analysis, the acoustic measures with the highest weight are shown in bold. In the case of the first and second dimensions (D1 and D2), the acoustic measures that have weights higher than those of other parameters are in boldface (weights > 4.0). In the case of the third dimension (D3), the acoustic measures that have weights higher than those of other parameters are in boldface (weights > 3.0).

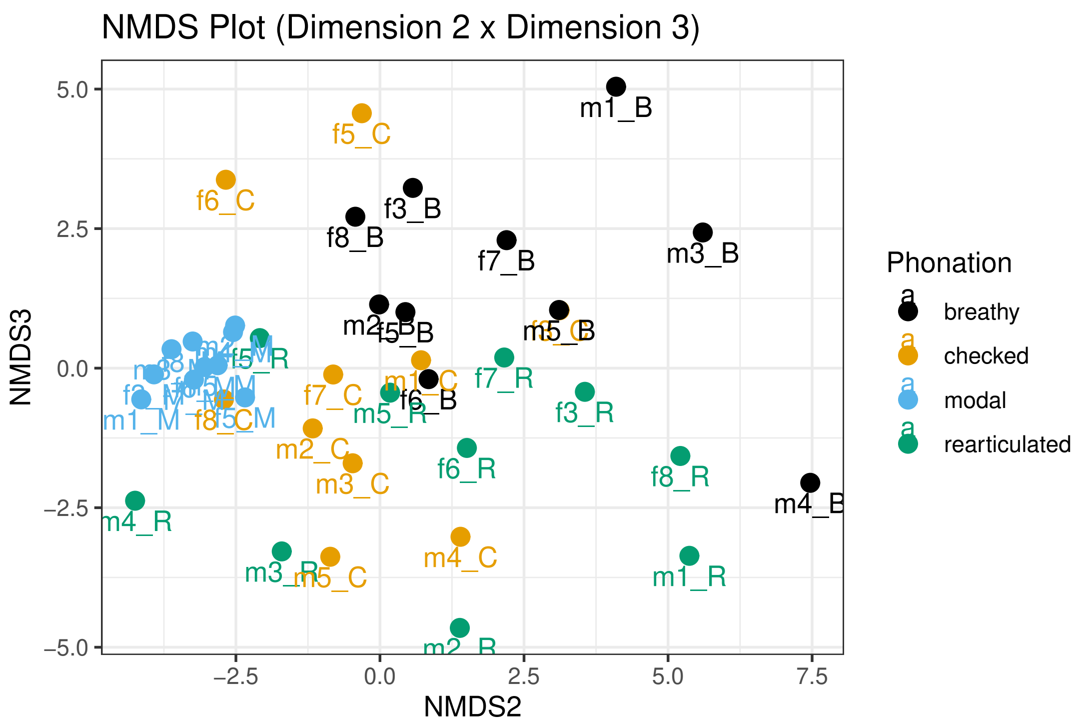

Figure 4.4: Two-dimensional MDS solution showing the second and third dimensions.

We see that for D1, the acoustic measures that have the highest weight on the first dimension are the amplitudes for the first three formants (i.e., A1*, A2*, A3*) and HNR < 500 Hz (i.e., a harmonics-to-noise ratio for everything from 0 to 500 Hz). For D2, the acoustic measures with the highest weight are H1*-A1*, H1*-A2* (i.e., spectral-slope measures), and the amplitudes of the first two formants. For D3, we see that HNR < 1500 HZ, HNR < 2500 Hz, HNR < 3500 Hz, and Residual H1* and H2* have the highest weights.

Table 4.1: Weight of each acoustic measure along each of the three dimensions indicated by the MDS solution (D1, D2, D3). Parameters that have weights higher than other parameters are in bold (weights > 4.0 for D1 and D2, and weights > 3.0 for D3).

Acoustic Measure	D1	D2	D3
H1*-H2*	1.03	1.01	0.39
H2*-H4	1.15	3.98	2.13
H1*-A1*	2.22	**5.15**	1.84
H1*-A2*	2.93	**4.66**	1.00
H1*-A3*	2.37	3.24	0.90
H4*-H2k*	1.47	0.31	1.59
H2k*-H5k*	3.73	0.73	0.84
residual H1*	1.75	0.97	**4.24**
H2*	1.76	0.94	**4.09**
H4*	0.79	**4.28**	0.10
A1*	**4.96**	**5.48**	0.17
A2*	**5.30**	**4.90**	1.38
A3*	**4.54**	2.91	1.11
CPP	**4.08**	0.10	1.68
HNR < 500 Hz	**5.66**	1.47	1.81
HNR < 1500 Hz	3.95	2.68	**3.08**
HNR < 2500 Hz	3.15	1.63	**3.42**
HNR < 3500 Hz	2.86	0.55	**3.19**
Strength of Excitation	2.09	0.78	0.36
SHR	2.39	0.50	0.47
Energy	2.22	3.91	0.64

4.4 Discussion

The results of the MDS analysis show that the acoustic space in which SLZ's voice quality occupies is similar to other languages. Similar to what Keating et al. (2023) found in their study, the first dimension appears to roughly be similar to the open quotient of the glottis as proposed by Gordon & Ladefoged (2001). In this model, voice quality is seen as the result of the glottis being more open or closed during phonation. 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 from Gordon & Ladefoged (2001) is shown in Figure 4.5.

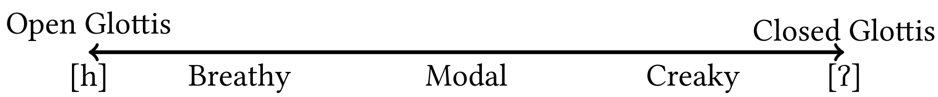

Figure 4.5: A diagram showing the relationship between breathy, modal, and creaky phonation types. Based on Gordon & Ladefoged (2001).

As mentioned above, the measures that contribute the most to this first dimension are the amplitudes of the first three formants, CPP, and Harmonics-to-Noise Ratio < 500 Hz. Interestingly, even though this dimension is similar to the open-quotient model put forward by Gordon & Ladefoged (2001), we do not observe measures traditionally associated with the open-quotient (i.e., spectral-slope). Instead of seeing traditional spectral-slope measures, we find the three formant amplitudes used to normalize the amplitude of the fundamental like in the measures H1*-A1*, H1*-A2*, and H1*-A3*. This suggests that the first dimension is more about the formants' amplitude than the signal's spectral-slope. This is combined with CPP and HNR < 500 Hz, which measures the harmonics-to-noise ratio for the first 500 Hz of the signal. This suggests that the first dimension is also concerned with the amount of noise in the signal.

The second dimension divides the space into modal versus nonmodal voice quality. The acoustic measures that contribute the most weight to this dimension are the spectral-slope measures H1*-A1* and H1*-A2* and the harmonic amplitudes of H4*, A1*, and A2*. This suggests that the second dimension is more about the spectral-slope of the signal than about the amount of noise in the signal. This is interesting given that traditional spectral-slope measures are associated with the open-quotient model of voice quality (Holmberg et al. 1995, Kreiman, Gerratt & Antoñanzas-Barroso 2007, Garellek et al. 2016, Garellek 2019, Chai & Garellek 2022).

The third dimension adds more information on nonmodal voice quality. Figure 4.3 and Figure 4.4, this third dimension separates the breathy voice from the other non-modal phonation types. The measures contributing the most to this dimension are the harmonics-to-noise ratio for the first 1500 Hz, 2500 Hz, and 3500 Hz. In addition, the residual H1* and H2* have the highest weights, which is interesting given that residual H1* has been argued to be a more robust measure of the spectral-slope of the signal than traditional spectral-slope measures (Chai & Garellek 2022, Brinkerhoff & McGuire 2024). Furthermore, as discussed in Chapter 3, residual H1* represents the voice quality in SLZ better than H1*-H2* and H1*-A3*. This dimension is characterized by the harmonics-to-noise ratios for the first 1500 Hz, 2500 Hz, and 3500 Hz. This suggests that the third dimension is more about the signal's spectral quality than about the formants' amplitude. This is combined with residual H1* and H2*, which are measures of the spectral-slope of the signal.

4.5 Conclusion

Although the discussion has predominately been about the weights 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 weight does not mean that it does not contribute to the acoustic space, but it is still important to understand the acoustic space in SLZ. 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.

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

Chapter 5

Trees reveal the importance of measures in SLZ

5.1 Introduction

The MDS analysis presented in Chapter 4 helps us to understand the acoustic landscape of SLZ. This helps reveal the multidimensional nature of voice quality and how all the acoustic measures work together to produce that acoustic landscape. The chapter also discussed how certain measures contributed more weight than other acoustic measures to each of the different dimensions. However, it does not tell us what measures are more important in separating the voice qualities from from one another. This is where decision trees can be most helpful.

5.2 What are Decision Trees

Decision trees are a statistical tool that helps to reveal which variables 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 variables (see Hastie, Tibshirani & Friedman 2009, James et al. 2021 for explanations on the statistics and how to conduct this 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. Additionally, 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 were 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 most important in making the split in the acoustic space for voice quality (Keating et al. 2023).

In their study, Keating et al. (2023) 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

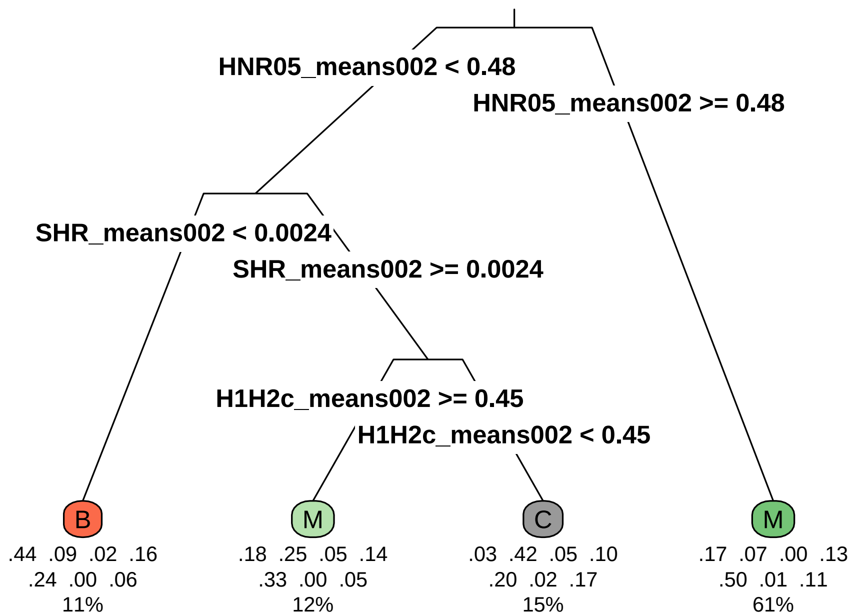

Figure 5.1: Classification tree of phonation categories from Keating et al. (2023). Abbreviations used in this figure are: HNR05_means002: harmonics-to-noise ratio over the frequency range from 0 Hz to 500 Hz for the middle third of each vowel; SHR_means002: subharmonic-to-harmonic ratio for the middle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

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, however, that 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 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 Bagging 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).

One way to overcome these disadvantages, is to make use of a technique called bootstrap aggregating, or bagging (Breiman 1996). This means that instead of growing a single tree on 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. 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 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 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.

In many instances of bagging trees, 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 where 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.

In this chapter, I will use bagging trees to understand the importance of the different acoustic measures in making the splits in the acoustic space of SLZ. This will help to understand which measures are most important in separating the different voice qualities from one another.

5.5 Bagging Trees in SLZ

Using the same data as in Chapter 4, I will use bagging trees to understand the importance of the different acoustic measures in making the splits in the acoustic space of SLZ. In order to determine the number of trees that are needed to stabilize the prediction, I compared models with different numbers of trees. The results of this comparison are shown in Figure 5.2.

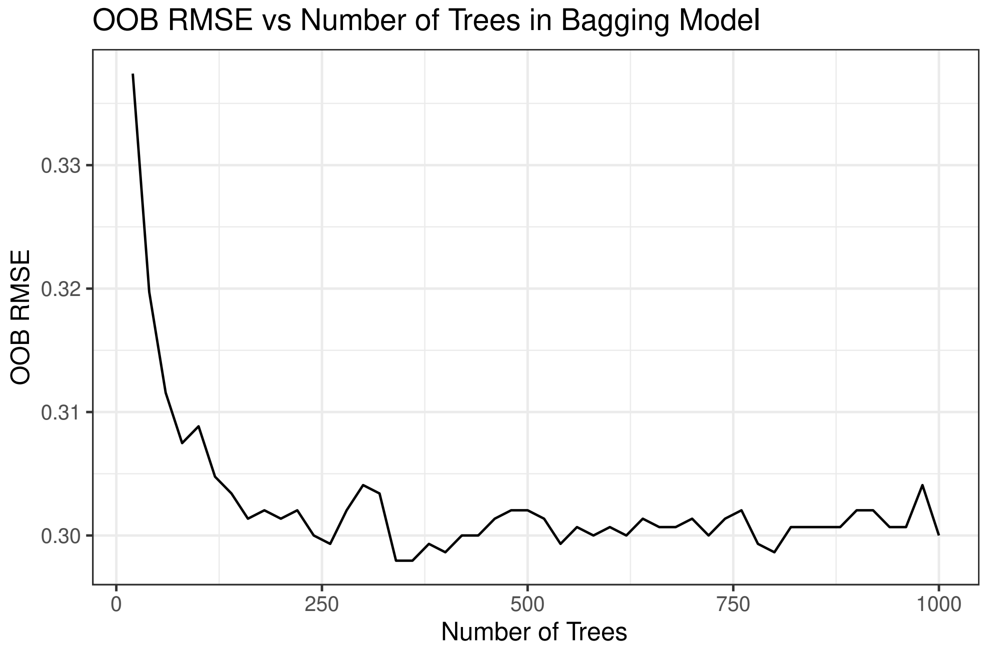

Figure 5.2: Plot showing the out-of-bag error as a function of the number of trees ran in the bagging models.

From Figure 5.2, it is clear that the out-of-bag error stabilizes at approximately 400 trees. This means that the number of trees that are required for the the bagging model to be the most stable is also 400 trees. In order to further insure the stability of the model, I also ran a 10-fold cross-validation on the bagging model with 400 trees. The results of this model are shown in Figure 5.3.

From Figure 5.3, the variables with the largest mean decrease in Gini index are A1* (the amplitude of the harmonic closest to the first formant), residual H1*, and the harmonics-to-noise ratio over the frequency range from 0 Hz to 1500 Hz (HNR 1500 Hz). This means that these three variables are the most important in making the splits in the acoustic space of SLZ. This is consistent with the results of the MDS analysis

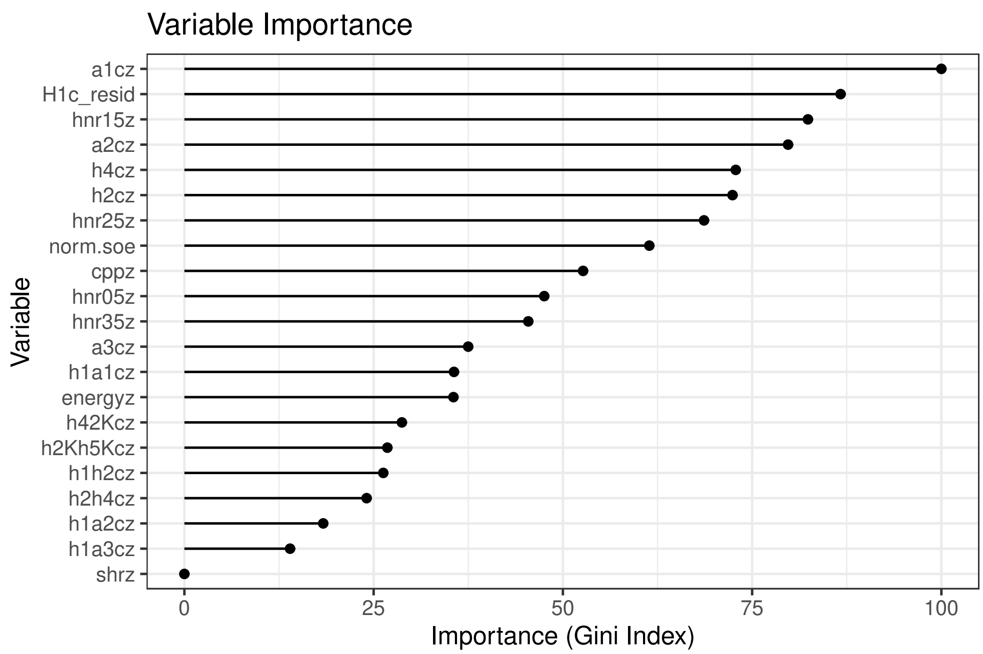

Figure 5.3: Variable importance plot showing the importance of the different acoustic measures in the bagging model based on the total amount that the Gini index is decreased by splits over a given predictor, averaged over all 400 trees.

in Chapter 4, where these three variables were some of the variables that contributed the most to the different dimensions of the acoustic space.

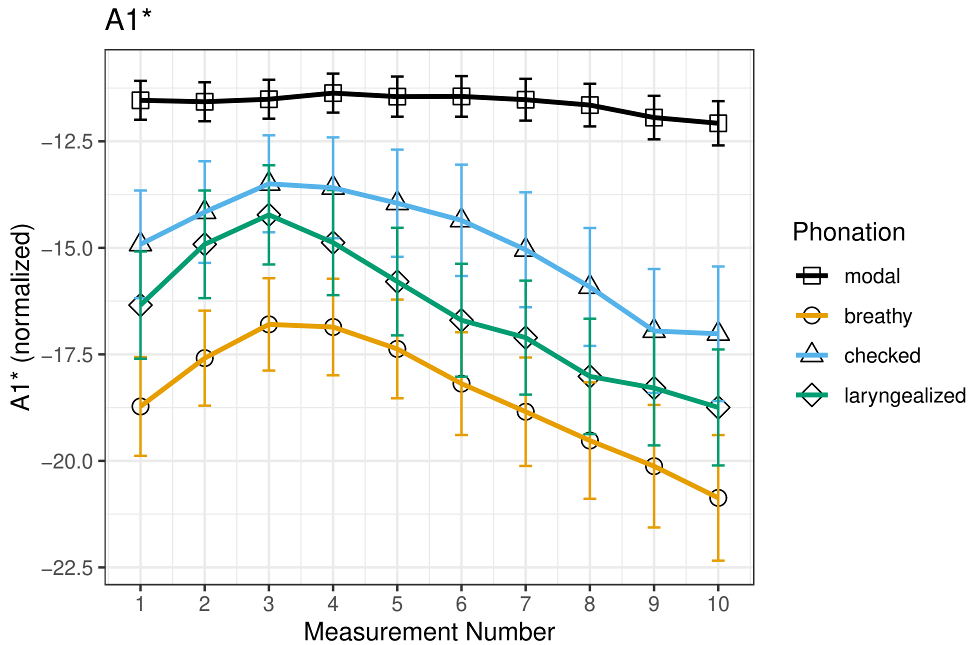

Figure 5.4: Plot showing the distribution of A1* across the different voice qualities in SLZ.

5.6 Discussion

5.6.1 Importance of A1*

5.6.2 Importance of Residual H1*

5.6.3 Importance of HNR 1500 Hz

5.7 Conclusion

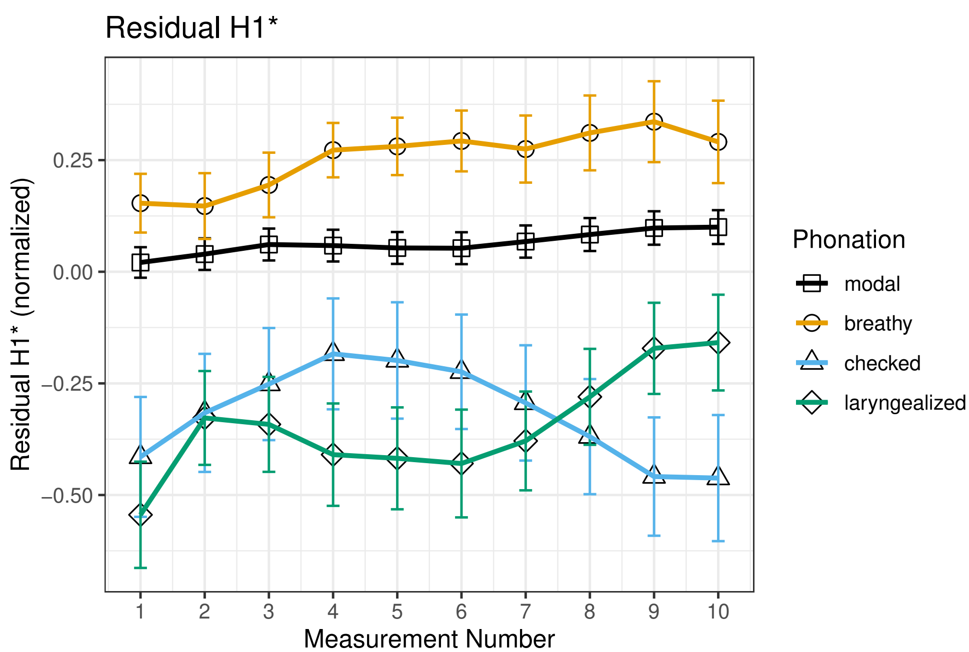

Figure 5.5: Plot showing the distribution of residual H1* across the different voice qualities in SLZ.

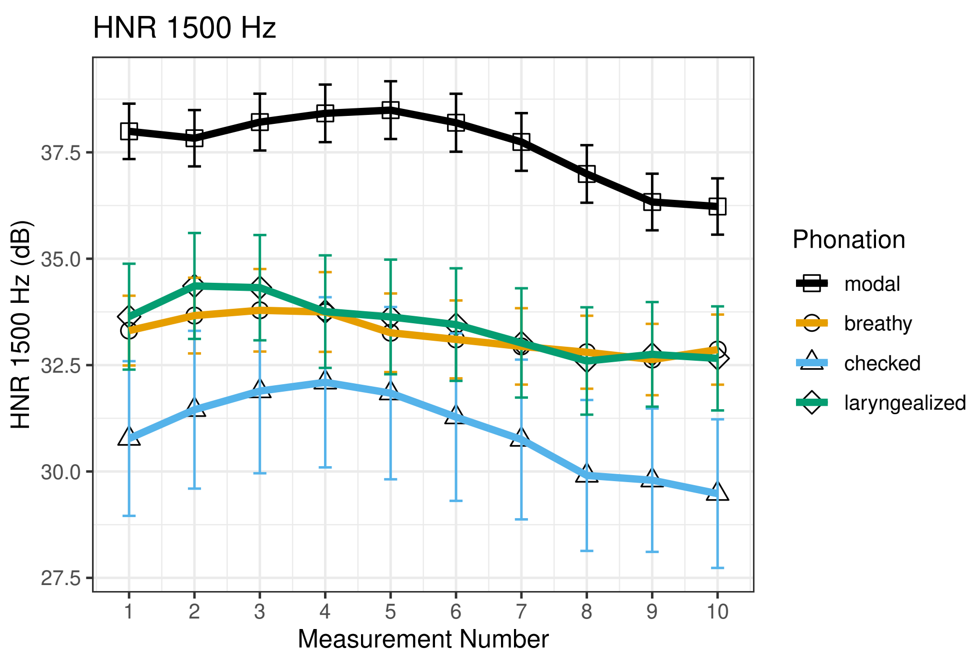

Figure 5.6: Plot showing the distribution of HNR 1500 Hz across the different voice qualities in SLZ.

Chapter 6

Testing the laryngeal complexity hypothesis

6.1 Laryngeal Complexity

Some other research was once performed.

6.2

Figure 6.1: A first figure.

Figure 6.2: A second figure.

Chapter 7

Modeling laryngeal complexity

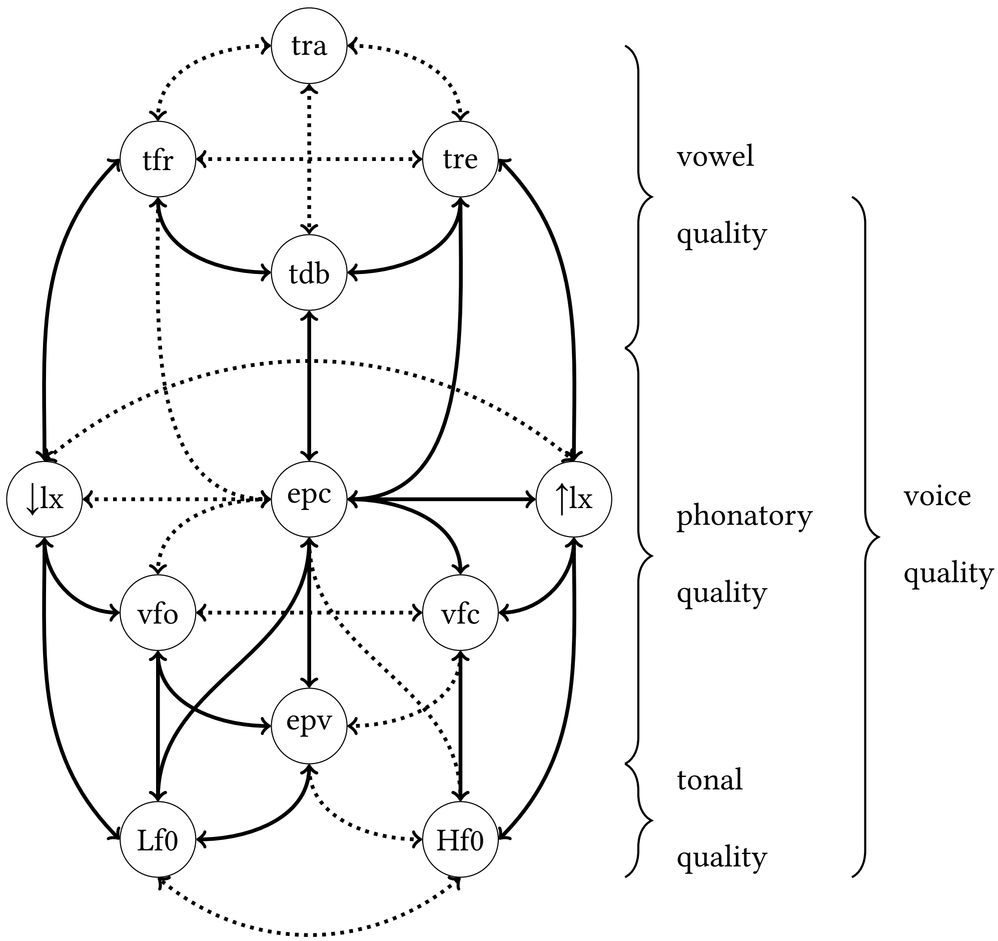

Figure 7.1: The Laryngeal Articulator Model from Esling et al. (2019). This model shows the interactions between the laryngeal articulators (labeled circles). Syngeristic interactions are shown with solid lines, while anti-syngeristic interactions are shown with dotted lines.

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.

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 O. 1998. *An Introduction to Error Propagation: Derivation, Meaning and Examples of Equation CY= FX CX FXT*. 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.

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.

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.

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

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

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

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

Brinkerhoff, Mykel Loren & Grant McGuire. 2024. On residual H1* as a measure of voice quality. In Taehong Cho, Sahyang Kim, Jeff Holiday & Sang-Im Lee-Kim (eds.), *Proceedings of the 19th Conference on Laboratory Phonology*, 39–40. Seoul, Korea: Hanyang Institute for Phonetics and Cognitive Sciences of Language5.

Buder, Eugene H. 1999. Acoustic analysis of voice quality: A tabulation of algorithms 1902-1990. In Raymond D. Kent & Martin J. Ball (eds.), *Voice quality measurement*, 119–244. San Diego: Singular Pub. Group.

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. 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 & 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). 47. https://doi.org/10.3390/languages7010047.

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

DiCanio, Christian T. 2012. 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.

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., 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. & 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.

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.

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. 2020. Extending the Person-Case Constraint to Gender: Agreement, Locality, and the Syntax of Pronouns. *Linguistic Inquiry*. 1–40. https://doi.org/10.1162/ling_a_00395.

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 & Christina M. Esposito. 2021. Phonetics of White Hmong vowel and tonal contrasts. *Journal of the International Phonetic Association*. 1–20. https://doi.org/10.1017/S0025100321000104.

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 & 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.

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 & 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.

Holmberg, Eva B., Robert E. Hillman, Joseph S. Perkell, Peter C. Guiod & Susan L. Goldman. 1995. Comparisons Among Aerodynamic, Electroglottographic, and Acoustic Spectral Measures of Female Voice. *Journal of Speech, Language, and Hearing Research* 38(6). 1212–1223. https://doi.org/10.1044/jshr.3806.1212.

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

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

James, Gareth, Daniela Witten, Trevor Hastie & Robert Tibshirani. 2021. *An introduction to statistical learning: with applications in R*. Second edition (Springer Texts in

Statistics). New York: Springer. 1 p. https://doi.org/10.1007/978-1-0716-1418-1.

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.

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

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

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

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

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.

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.

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

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.

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

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

Podesva, Robert J. 2016. Stance as a Window into the Language-Race Connection: Evidence from African American and White Speakers in Washington, DC. In H. Samy Alim, John R. Rickford & Arnetha F. Ball (eds.), *Raciolinguistics: How Language Shapes Our Ideas About Race*, 203–220. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190625696.003.0012.

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

Santiago Laxopa. 2022. *Santiago Laxopa: Economy, Employment, Equity, Quality of Life, Education, Health and Public Safety*. Data México. https://datamexico.org/en/profile/geo/santiago-laxopa (25 April, 2022).

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. 2009. VOICESAUCE: A program for voice analysis. *The Journal of the Acoustical Society of America* 126(4). 2221. https://doi.org/10.1121/1.3248865.

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 *Clasidicación de las lenguas indígenas de México: Memorias del III Coloquio Internacional de Lingüística Mauricio Swadesh*, 69–133. Mexico City: UNAM: Instituto de Investigaciones Antropológicas.

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

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.

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.

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.

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.