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 **silvermanLaryngealComplexityOtomanguean1997**, **silvermanPhasingRecov** 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, nonmodal 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 cxan 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" (**laverVoiceQualityIndexical1968**) and racial identity (**podesvaStanceWindowLanguageRace2016**).

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 (**garellekProductionPerceptionGlottal2**
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

1

vowels (**fischer-jorgensenPhoneticAnalysisBreathy1968, espositoContrastiveBreathinessConso**

**khanPhoneticsContrastivePhonation2012, espositoDistinguishingBreathyConsonants2019**).

**espositoCrossLinguisticPatterns2020**

## 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

- **espositoCrossLinguisticPatterns2020** 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 *f0*/tone/pitch accents (rows) and voice quality on vowels (columns) and their interactions (**espositoCrossLinguisticPatterns2020**)

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

- 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 **uchiharaToneRegistrogenesisQuiavini2016** 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 **silvermanLaryngealComplexityOtomanguean1997** 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 **silvermanLaryngealComplexityO** 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

- **espositoCrossLinguisticPatterns2020** 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 silt measures which reflect the open quotient.

  – The open quotient is the proportion of the glottal cycle during which the glottis is open (**holmbergComparisonsAerodynamicElectroglottographic1995**).

- 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 **chaiH1H2Acoustic2022**, what they are actually trying to do is get at the amplitude differences in H1 as first observed by **fischer-jorgensenPhoneticA laverVoiceQualityIndexical1968**.

  – I think this is where, I and other researchers fell into a fallicy where we think that just the spectral-tilt is what matters instead of realizing what those measures where trying to accomplish which was to normalize H1 for comparison.

- **espositoCrossLinguisticPatterns2020** further discuss the role that EGG play.

- Two things that I think are relavent 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 **silvermanLaryngealComplexityOt** 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, **espositoCrossLinguisticPatterns2020** 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 *f0*/tone/pitch accents (rows) and voice quality on vowels (columns) and their interactions (**espositoCrossLinguisticPatterns2020**)

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

## 1.5.1  Type I languages

- Type I languages are those languages that lack tonal contrasts and phonation contrasts.

- According to **espositoCrossLinguisticPatterns2020** 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 *f0* are a cue to the phonation.

- Type IIa languages are those where *f0* is not a cue.

- According to **espositoCrossLinguisticPatterns2020**, these languages are quite rare and only two languages are known to exhibit this behavior:

    – Danish (**gronnumDanishStodLaryngealization2013**)

    – Gujarati (**khanPhoneticsContrastivePhonation2012**)

- Type IIb languages are those where *f0* is a cue for the different phonation types.

- These languages are sometimes called "register languages"

- These are languages such as:

    – Chanthaburi Khmer

9

- Chong

- Javanese

- Kedang

- Mon

- Suai

- Wa

- In type IIb languages the way in which *f0* is a cue is language specific.

- They describe that breathy vowels in Wa and Kedang both begin with a lower f0 than their clear/modal counterparts

- However, breathy vowels in Chanthaburi Khmer have a higher f0 than their clear counterparts.

### 1.5.3   Type III languages

- Type III languages are characterized by langugaes 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 (**kuangCovariationVoiceQuality2017**).

- 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 consistantly produced with both a specific *f0* 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 Southest Asia.

- **espositoCrossLinguisticPatterns2020** 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 **espositoVariationContrastivePhonation2010** where she describes that SAV Zapotec yllables 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 (**pickettIsthmusJuchitanZapotec2010**)

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 phona-

tion 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 con-

trasts are fused.

– In otherwords 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 *f0*/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) | | *f0* not a cue (Type IIa) | *f0* is a cue (Type IIb) | |
| **Tone contrast** | VQ not a cue (Type IIIa) | VQ is a cue (Type IIIb) | Orthogonal (Type IVa) | Fused (Type IVb) | Mixed (Type IVc) |

# 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 (**adlerAcousticsPhonationTypes2016**, **adlerDerivationVerbInitiality2018**, **foleyForbiddenCliticClusters2018**, **foleyExtendingPersonCaseConstraint2020**). According to **smith-starkAlgunasIsoglosasZapotecas2007**, 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/ (**nellisFortisLenisCajonos1980**, **jaegerInitialConsonantClusters1982**, **butlerh.Diccionario**, **avelinobecerraTopicsYalalagZapotec2004**, **longDiccionarioZapotecoSan2005**, **sonnenscheinDes**

17

Table 2.1: Vowel qualities in Santiago Laxopa Zapotec.

|      | front | central | back |
|------|-------|---------|------|
| high | i     |         | u~o  |
| mid  | e     |         |      |
| low  |       | a       |      |

The vowel /o/ is marginal in SLZ's lexicon, only appearing in a few lexical items 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 **sonnenscheinDescriptiveGrammarSan2005** 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 (**pillaiNewTestCriteria1955**, **hayFactorsInfluencingSpeech2006**, **nyczBestPracticesMeasuring2014**) and Bhattarcharyya's Affinity (**bhattacharyyaMeasureDivergen johnsonQuantifyingOverlapBhattacharyyas2015**, **warrenQualityQuantityNew2018**,
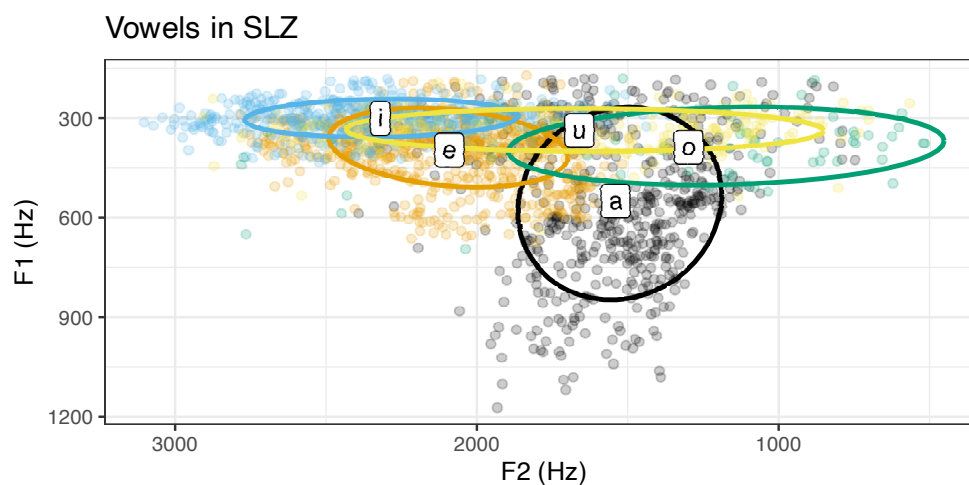
Figure 2.2: Vowel space of Santiago Laxopa Zapotec. The ellipses around each vowel mean represents 1 standard. The scale of the axes are in barks with their corresponding Hz values.

**strellufChapter3Low2018**). 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 show the process of complete and partial mergers between vowels, such as the NEAR-SQUARE vowel merger in New Zealand English (**hayFactorsInfluencingSpeech2006**). 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 over-

lap 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-garciaPhonationTyp** 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* /çạ$^L$/ 'metal; rifle; bell'

c. *cha'* /tʃaˀ$^L$/ 'cooking pot'

d. *ya'a* /çaˀa$^L$/ '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 (**nellisFortisLenisCajonos1980**, **jaegerInitialConsonantClusters1982**, **butlerh.DiccionarioZapotecoYatzachi1997**, **avelinobecerraTopicsYalalagZapotec2004**, **sonnenscheinDescriptiveGrammarSan2005**, **longDiccionarioZapotecoSan2005**).[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 Zpaotec varities, two types of creaky voice: checked and rearticulated. Checked vowels are characterized by an abrupt glottal closure which

---

[1]Breathy voice in Zapotec languages, however, is common in Central Valley Zapotecs (**munroDiCsyonaaryTee1999**, **espositoSantaAnaValle2004**, **espositoVariationContrastivePhonation2010**, **uchiharaToneRegistrogenesisQuiavini2016**, **ariza-garciaPhonationTypesTones2018**).

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]

### 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 **gerfenProductionPerceptionLaryn** 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 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 (**pikeProblemsZapotecTone1948**, **renschComparativeOtomangueanPhonology1976**, **josserandMixtecDialectHistory1983**, **silvermanLaryngealComplexityOtomanguean1997**, **beamdeazconaProblemsZapotecTone2007**,

**dicanioItunyosoTrique2010**, **dicanioCoarticulationToneGlottal2012**, **elliottChicahuaxtlaTriqui** **campbellOtomangueanHistoricalLinguistics2017a**, **campbellOtomangueanHistoricalLinguisti** **lillehaugenOtomangueanLanguages2019**, **eischensTonePhonationPhonologyPhonetics2022**). SLZ has a five-way tonal contrast which consists of three level tones (high, mid, low) and two contour tones (rising and falling).

**brinkerhoffTonalPatternsTheir2022**

## 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 (**avelinobecerraTopicsYalalagZapotec2004**). 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 (**garellekPhoneticsVoice2** and to convey paralinguistic information by "indexing the biological, psychological, and social characteristics of the speaker" (**laverVoiceQualityIndexical1968**, **podesvaStanceWindowl** Voice quality contrasts have been studied extensively by examining their correlates in the acoustic signal (**espositoCrossLinguisticPatterns2020**), resulting in a large and complex literature on acoustic correlates of phonation differences (**garellekPhoneticsVoice2019**).

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

ive amplitudes of the first harmonic and second harmonic. As established by Fischer-Jorgensen, the relative strength of the fundamental is a correlated measure of breathy voice in contrast with modal voice **fischer-jorgensenPhoneticAnalysisBreathy1968**. 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 (**garellekPhoneticsVoice2019**, **chaiH1H2Acoustic2022**).

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. **sundbergObjectiveCharacterizationPhonation2022** 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-jorgensenPhoneticAnalysisB** Similarly, in a comprehensive overview of the main concerns with $H1^*-H2^*$ as a phonation type measure, **chaiH1H2Acoustic2022** 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 (**arrasIntroductionErrorPropagation1998**). 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 (**chaiH1H2Acoustic2022**). 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. **simpsonFirstSecondHarmonics2012** 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 **chaiH1H2Acoustic2022** 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 EEG data.

The contributions **chaiH1H2Acoustic2022** 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 **chaiH1H2Acoustic2022** 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 (**silvermanLaryngealComplexityOtomanguean1997**, **ariza-garciaPhonationTypesTone** 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, **espositoVariationContrastivePhonation2010** 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 **klattAnalysisSynthesisPerception199**

---

[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 **chaiH1H2Acoustic2022**. The Mandarin case is less relevant as its phonation is prosodic and tonally linked.

**arellanesarellanesDosGradosLaringizacion2010** 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 (**crowhurstInfluenceVowelLaryngealisation2016**). **barzilaiContextdependentPho** 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 (**adlerAcousticsPhonationTypes2016**, **adlerDerivationVerbInitiality2018**, **foleyForbiddenCliticClusters2018**, **foleyExtendingPersonCaseConstraint2020**, **sichelPronounsAttractionSierra2020**, **sichelFeaturalLifeNominals2020**, **brinkerhoffDownstepS**

**brinkerhoffTonalPatternsTheir2022**). It is spoken by 981 people in the municipality of Santiago Laxopa, Ixtlán, Oaxaca, Mexico and a small number of other speakers from the diaspora in Mexico and the United States.

Santiago Laxopa Zapotec exhibits a four-vowel inventory (that is, /i/, /e/, /a/, and /u/), which is further distinguished by a four-way contrast in voice quality. This variety is unique because it is a Northern Zapotec that has developed a breathy voice (/V̤/) in addition to the two types of laryngealization that characterize the rest of the Zapotec languages, namely checked and articulated. Checked vowels are defined as a modal vowel that ends with a period of creaky voice or a glottal closure (/VV̰/ or /Vˀ/ ). Rearticulated vowels are also defined as a modal vowel that also has a period of creakiness or glottal closure but aligned to the middle of a vowel (/VV̰V/ or /VˀV/). This makes an otherwise modal vowel appear as if it is interrupted by this laryngealization.[2] This difference in laryngeal timing is one of the key differences between these phonations (see **ariza-garciaPhonationTypesTones2018** 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 (**brinkerhoffTonalPatternsTheir2022**).[3]

---

[2]This is different from how rearticulated vowels are defined in other languages, where they are a modal vowel followed by an "echo vowel" which may or may not have a glottal closure intervening between the modal and "echo" vowel (**bairdPhoneticPhonologicalRealizations2011**). A fuller description of how these vowels are realized in one variety of Zapotec and the terms used are found in **avelinoAcousticElectroglottographicAnalyses2010**.

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

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.

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

## 3.3 Methods

### 3.3.1 Elicitation

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

The wordlist consisted of 72 items repeated three times in isolation and the carrier sentence *Shnia' X chonhe lhas* [ʃnːiaˀ X tʃone ɾas] "I say X three times". This sentence was chosen to minimize any effect of the phrasal position (**crowhurstInfluenceVowelLaryngealisatior** Additionally, there are no co-articulatory effects on voice quality found with this frame and the items in the wordlist. Between these 72 words, there were 11 words with

breathy voice, 9 with rearticulated voice, 10 with checked voice, and 42 with modal. Thirteen of the 72 words were disyllabic and eight of the thirteen contained the same voice quality in each syllable. Of those 13 disyllabic word, only five contained different voice qualities in each of the syllables. This resulted in 85 different syllables.[4]

The token selection for the wordlist was done in consultation with a native speaker. Similarly to **barzilaiContextdependentPhoneticEnhancement2021**, 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 |

---

[4]There is ongoing debate about the status of strong syllables in Zapotec. The majority of evidence for the presence of stress comes from non-Northern Zapotec varieties (**chavez-peonPhoneticCuesStress2008**, **mockPitchAccentStress1988**). At this time, evidence for stress in Northern Zapotec remains to be seen.

### 3.3.2 Data Processing

Each vowel of the target words in the carrier sentence condition was labeled following **garellekAcousticDiscriminabilityComplex2020** 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.

These vowels were then extracted and fed into VoiceSauce for acoustic measurement (**shueVoiceSauceProgramVoice2011**). The formants were measured using Snack (**sjolanderSnackSoundToolkit2004**), while the fundamental frequency ($f0$) was measured using the STRAIGHT algorithm (**kawaharaInstantaneousfrequencybasedPitchExtraction199**
Spectral slope measures were corrected for formants and bandwidths (**hansonGlottalCharacteristicsFe**
**iseliAgeSexVowel2007**). 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 **chaiH1H2Acoustic2022** in their study. The H1*, H1*–H2*, and $f0$ values were z-scored by speaker to reduce the variation between the speakers and provide a way to directly compare the different measures on the same scale. Data points with an absolute z-score value greater than 3 were considered outliers and excluded from the analyses. Within each vowel category, we calculated the Mahalanobis distance in the F1-F2 panel. Each data point with a Mahalanobis distance greater than 6 was considered an outlier and excluded from

the analysis. Using the Mahalanobis distance allows us to compare the data points to the mean of the F1-F2 panel for each vowel category. The larger the Mahalanobis distance is the more deviant the data point is from the mean which in turn means that the data point was improperly tracked. This is comparable to what was done in **seyfarthPlosiveVoicingAcoustics2018**, **chaiCheckedSyllablesChecked2022**, **garellekPhoneticsWhiteHmong2023**.

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

After removing the outliers, we calculated residual H1* for the remaining data points following **chaiH1H2Acoustic2022**. First, a linear mixed effects model was generated with the z-scored H1* as the response variable and the z-scored energy as the fixed effect. The uncorrelated interaction of the z-scored energy by speaker was treated as random. The energy factor resulting from this linear mixed-effects model was extracted. Finally, the z-scored H1* had the product of the z-scored energy and the energy factor subtracted from it, giving us the residual H1* measure.

The measures were then orthogonally coded according to their position in the

vowel (first, middle, and third) and for tone (H, M, L, R, F) for statistical modeling.

### 3.3.3   Statistical modeling

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

The tone and the interaction between voice quality and position in the vowel were selected as fixed effects for several reasons. The first is that five unique tones appeared in the data and it is well established that tone interacts with voice quality in different ways (see **espositoCrossLinguisticPatterns2020** and **garellekPhoneticsVoice2019** 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

---

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

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.

Additional statistical modeling was performed with two generalized additive mixed models (GAMM) using the `mgcv` package in R (**woodGeneralizedAdditiveModels2017**). 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 where 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 **garellekPhoneticsVoice2019**, it is expected that breathy vowels should have a higher spectral slope than modal vow-

---

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

els across the two measures in keeping with observations about breathy vowels in other languages. Because both checked and rearticulated vowels make use of creaky voice, they should have a lower spectral slope than modal vowels. Additionally, because there is a temporal distinction between checked and rearticulated with where the creakiness appears, this should also be captured in the two measures.

### 3.4.1   H1*−H2*

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

### 3.4.2   Residual H1*

Figure 3.2 shows the mean residual H1* values for each voice quality at each of the ten vowel intervals. In contrast to Figure 3.1, we see that breathy has a higher residual H1* measure than modal throughout the duration of the vowel, which is consistent with other observations for breathy voice (**fischer-jorgensenPhoneticAnalysisBreathy1968**). Checked and rearticulated both have lower values than the modal at each of the 10
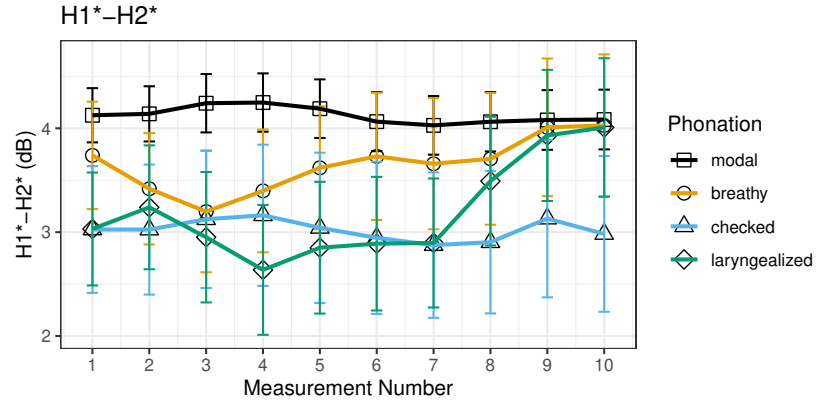
Figure 3.1: H1*−H2* across the duration of the vowel. Points represent the mean of each measure across the ten intervals. The error bars around each point represent a 95% confidence interval. A line was plotted over each to show how the acoustic measure functions across the ten intervals.

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

### 3.4.3   Model Comparison

To assess the robustness of the models, we compared the residual H1* linear mixed-effects model to the H1*−H2* linear mixed-effects model. This was done using two methods: direct comparison of the outputs of the two models in the same way as **chaiH1H2Acoustic2022** and the Akaike Information Criterion (AIC).

Table 3.2 compares the linear mixed effects models for H1*−H2* and residual H1*.
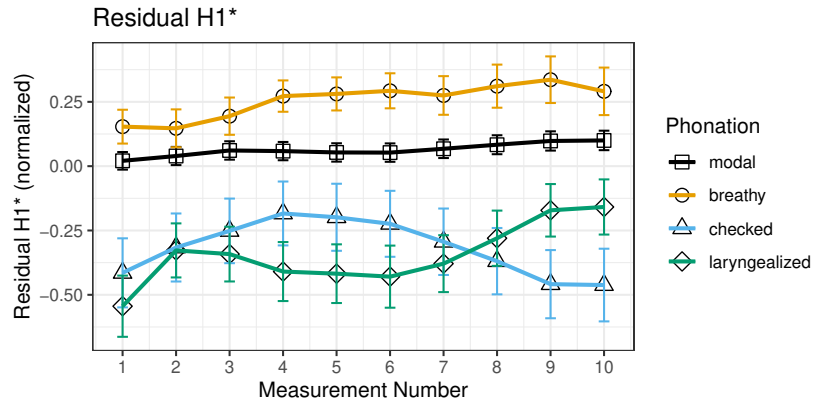
Figure 3.2: Residual H1* across the duration of the vowel. Points represent the mean of each measure across the ten intervals. The error bars around each point represent a 95% confidence interval. A line was plotted over each to show how the acoustic measure functions across the ten intervals.

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 | $\beta$ | 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 resid-

ual 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 (**burnhamMultimodelInferenceUnderstanding2004**, **burnhamAICModelSelection2011**, **burnhamModelSelectionMultimodel2004**).

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.

Do 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 show that for breathy vs. modal there was no significant difference between the two voice qualities in term 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.

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 provides a clearer distinction between the different voice qualities than the H1*−H2*
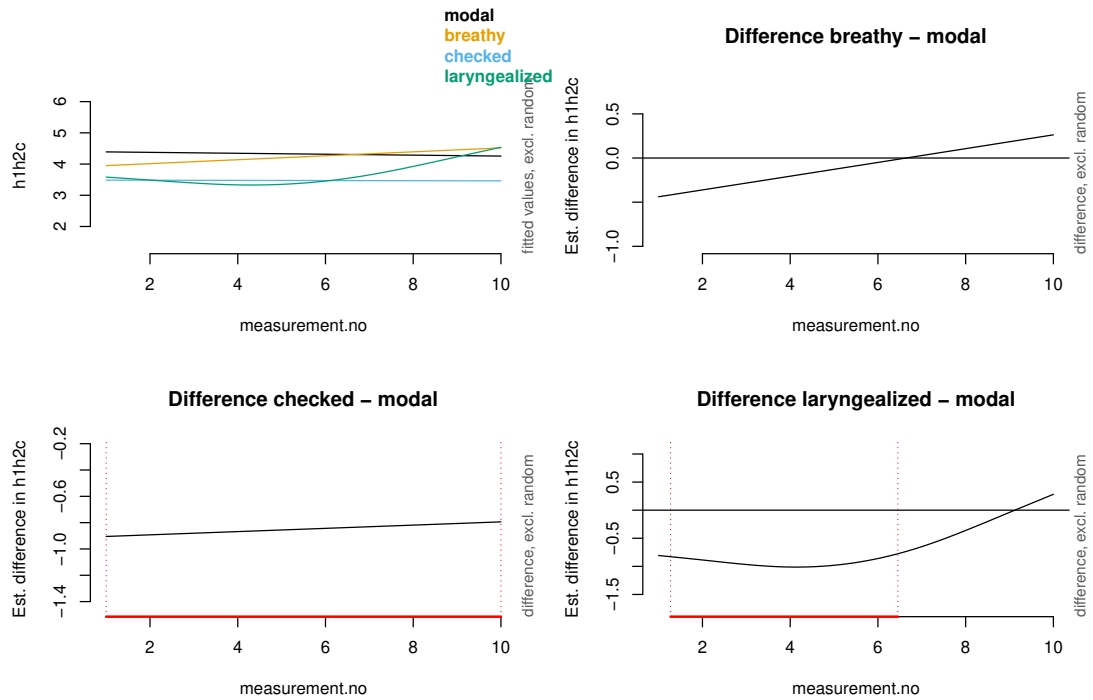
Figure 3.3: GAMM smooths and difference plots for H1*−H2* across the duration of the vowel.

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 where 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 **silvermanLaryngealComplexityOtomanguean1997**, **silvermanPhasingRecoverability1997** about the alignment of non-modal phonation in laryngeally complex languages. This

41
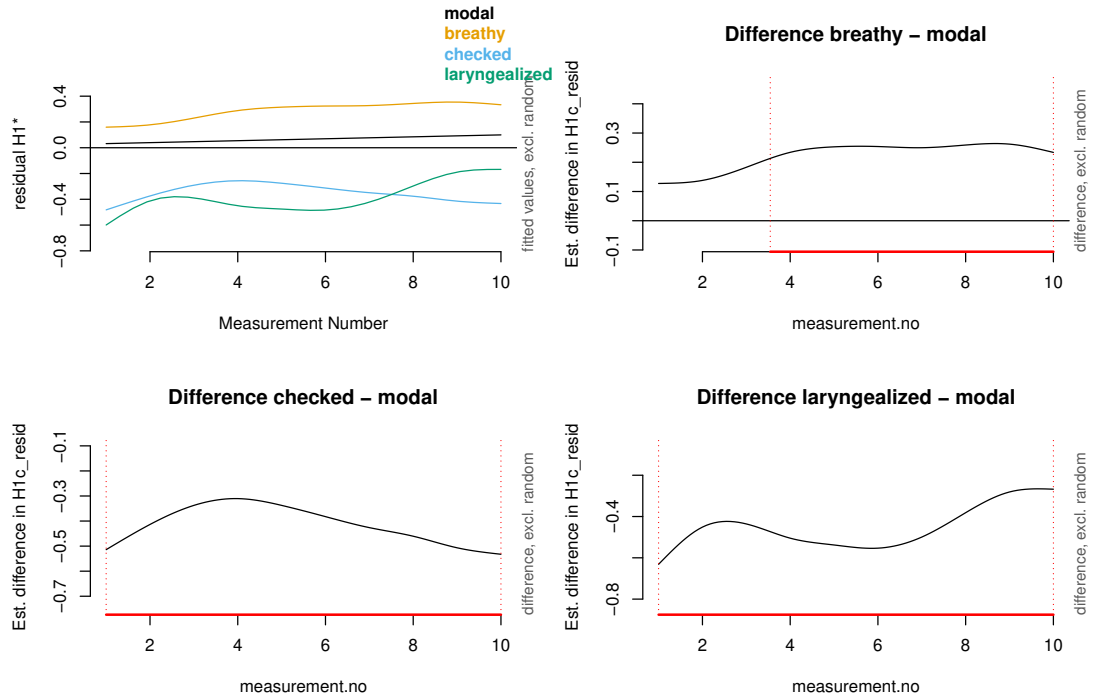
Figure 3.4: GAMM smooths and difference plots for residual H1* across the duration of the vowel.

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 **chaiH1H2Acoustic2022** 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 underdocumented language.

This study is based on the work conducted by **keatingCrosslanguageAcousticSpace2023** 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 (**ZapotecLanguageProject**). 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 (**garellekAcousticDiscriminabilityComplex2020**). The vowel was then extracted and saved as a separate file for analysis.

These vowels were fed into VoiceSauce (**shueVoiceSauceProgramVoice2011**) 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 (**kawaharaInstantaneousfrequencybasedPitchExtra** to estimate the fundamental frequency in millisecond (ms) intervals. Once the fun-

damental 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 (**sjolanderSnackSoundToolkit2004**) 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 **iseliAgeSexVowel2007**'s **iseliAgeSexVowel2007** extension of the method employed by **hansonGlottalCharacteristicsFemale1997**. 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.

### 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 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 **chaiH1H2Acoustic2022**. 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 **keatingCrosslanguageAcousticSpace2023** 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

A multidimensional scaling (MDS) analysis is a type of dimensionality reduction in order to visualize the relationships between the data points (**kruskalMultidimensionalScaling1978**). 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 **kreimanUnifiedTheoryVoice2014**, **kreimanValidatingPsychoacousticModel2021** **garellekAcousticDiscriminabilityComplex2020**, 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 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's rearticulated voice, and one for their breathy voice). This is similar to what **keatingCrosslanguageAcousticSpace2023** 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 (**oksanenVeganCommunityEcology2025**) in the R programming language (**rcoreteamLanguage** 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 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. As shown in Figure 4.1, most of the data is captured

in the first four dimensions. These four dimensions were retained for the analysis.

Scree Plot for the MDS solution with Duration added
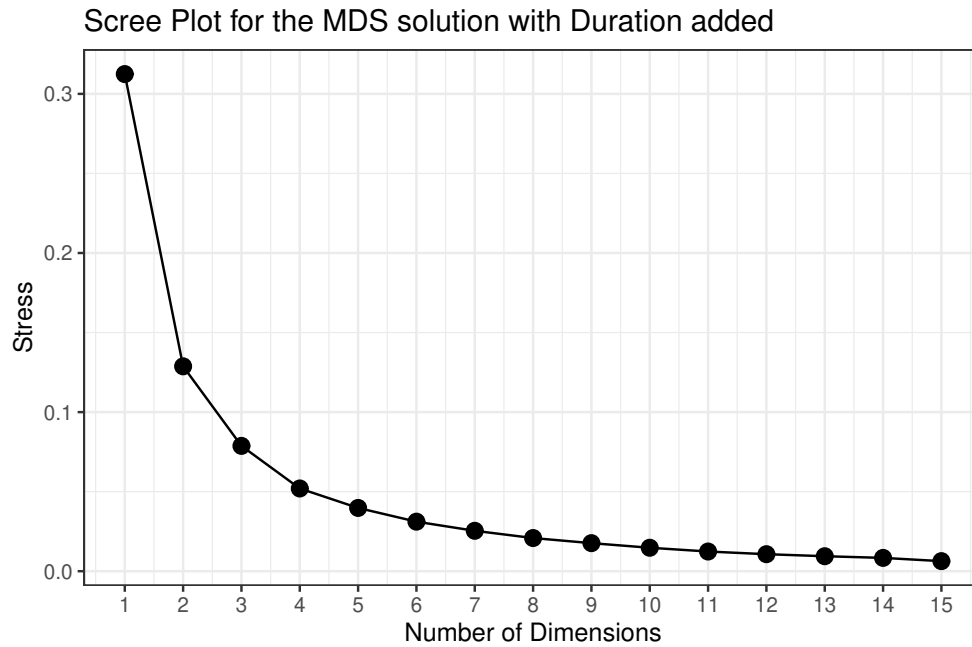
Figure 4.1: Scree plot showing the stress for each dimension for the MDS analysis.

## 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

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

voice with black. In each of the plots, modal voice is generally more densely packed than the nonmodal voice qualities. This is likely due to the fact that modal voice represents approximately 60% of the data, while the nonmodal 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 from breathy, modal, and creaky voices. The second dimension separates the modal voice, bottom of the plot, from the nonmodal voice qualities, top of the plot.

Figure 4.3 compares the first dimension to the third dimension. In this plot, We observe that 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 the 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 the the fourth dimension
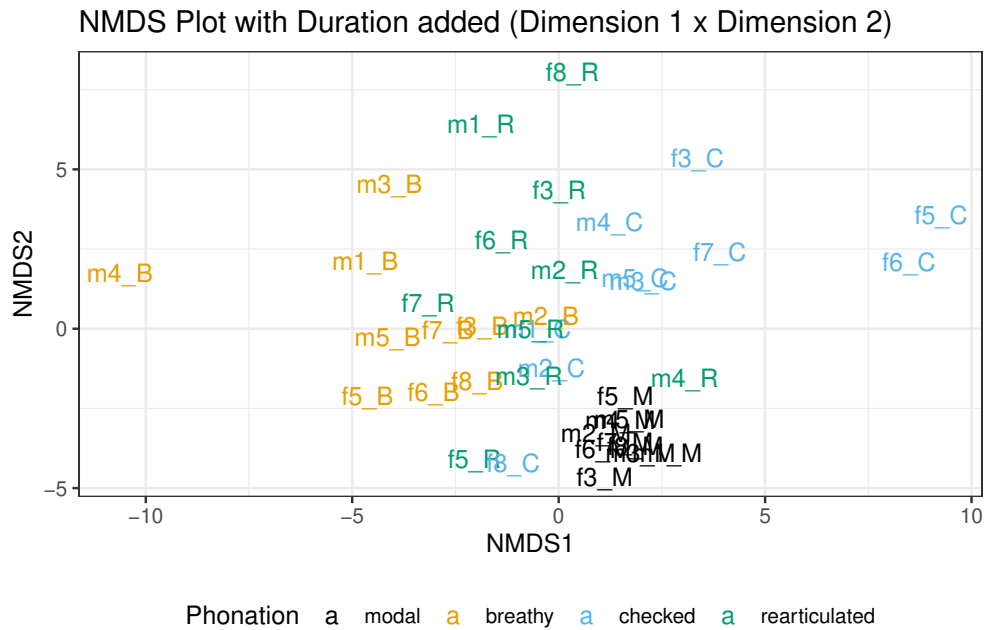
Figure 4.2: Two-dimensional MDS solution showing the first and second dimensions.

moves the modal voice from the bottom-center of the plot to almost the exact center of the plot.

Figure 4.5 shows the second dimension plotted against the third dimension. This plot is essentially the same as Figure 4.3, with 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 nonmodal voice are separated into two different clusters, with modal voice located at 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 modal voice, but no discernible pattern is observed for the other phonations.
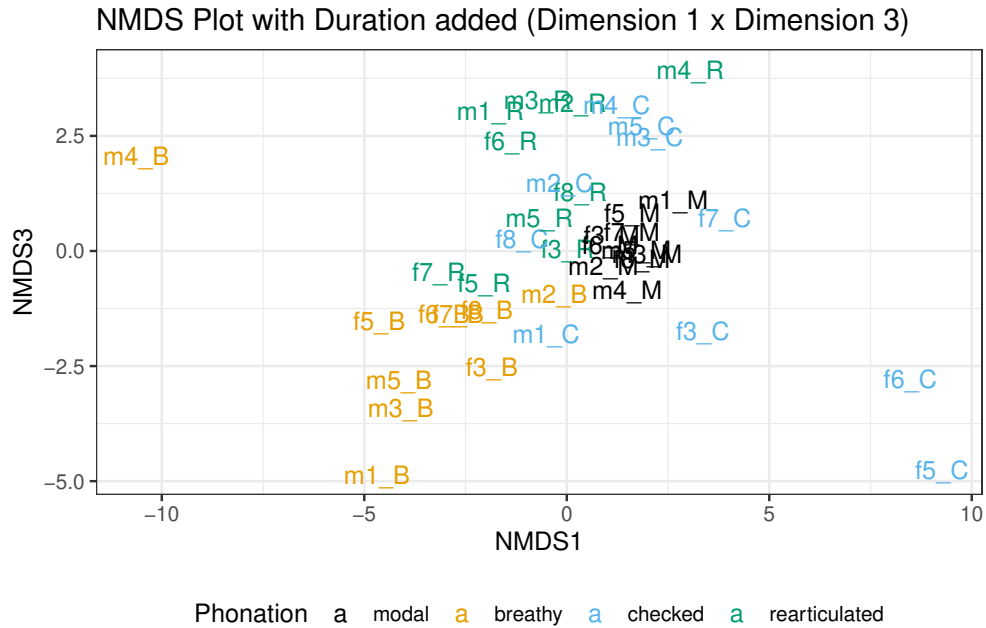
Figure 4.3: Two-dimensional MDS solution showing the first and third dimensions.

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, while rearticulated voice is located more to the right of the plot. Again, we observe that the modal voice is located at the center of the plot.

#### 4.3.1.1 Interim summary

The plots of the MDS analysis shows us 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

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

modal, and finally creaky voice. This is similar to what **keatingCrosslanguageAcousticSpace2023** found in their study that the first dimension captures this continuum. This continuum is very similar to the open-quotient model of voice quality put forward by **gordonPhonationTypesCrosslinguistic2001**, as illustrated in Figure 4.8. Because of this similarly to the open-quotient model, the first and third dimensions are likely capturing the open-quotient of the glottis or something similar.

From the plots involving the second dimension, we see that this dimension separates the modals from the nonmodals. This is similar to what, we observe with the various harmonics-to-noise ratios and cepstral peak prominence (CPP) which are all

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

measures of the the amount of noise present in the signal across various bandwidths (**dekromCepstrumBasedTechniqueDetermining1993**, **hillenbrandAcousticCorrelatesBreathy1** **blankenshipTimingNonmodalPhonation2002**, **ferrerriesgoWhatMakesCepstral2020**).
In addition to the amount of noise it could also be capturing the strength of the vocal fold vibration, similar to what **garellekVoicingGlottalConsonants2021** found in their study of glottal consonants and non-modal phonation. In their study, they found that modal voice had the highest strength of vocal fold vibration (as measured by the Strength of Excitation) while the nonmodals had a lower strength of vocal fold vibration. This suggests that the second dimension is capturing the amount of aperiodic noise in the signal, the strength of vocal fold vibration, or both.
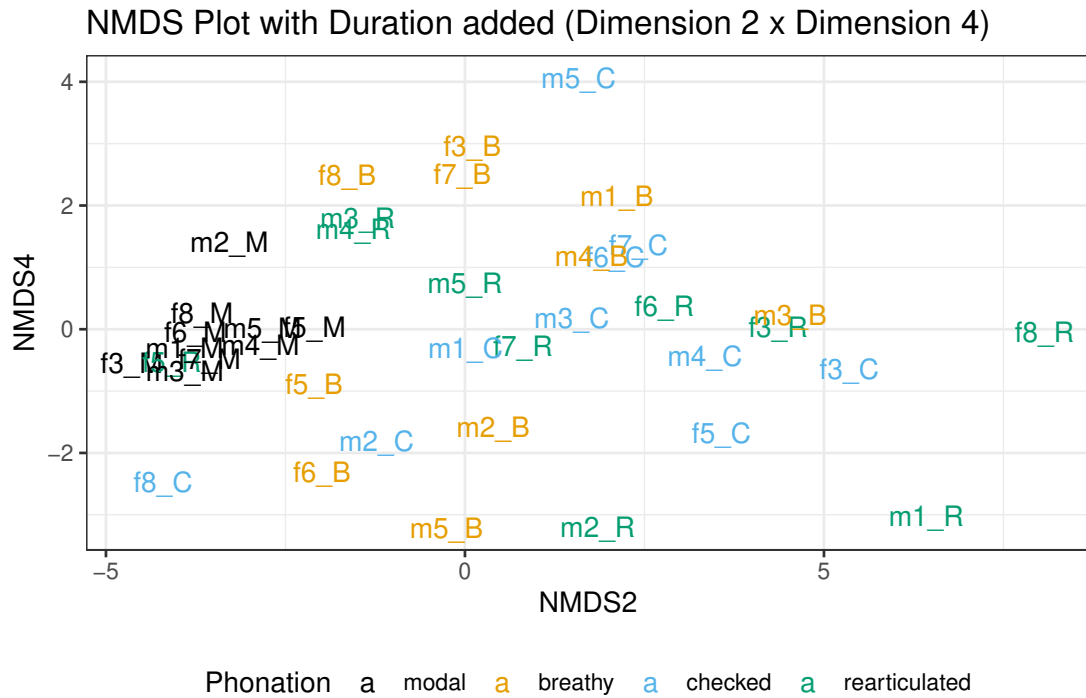
Figure 4.6: Two-dimensional MDS solution showing the second and fourth dimensions.

The fourth dimension is less clear in 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 is possibly capturing something about modal voice.

The rest of this chapter will focus on the how the different acoustic measures contribute to the different dimensions of the MDS analysis. This will be followed by a general discussion of the both the MDS analysis dimensions and the acoustic measures that are correlated with these dimensions. This discussion will focus on how the

**NMDS Plot with Duration added (Dimension 3 x Dimension 4)**

Phonation  a  modal  a  breathy  a  checked  a  rearticulated
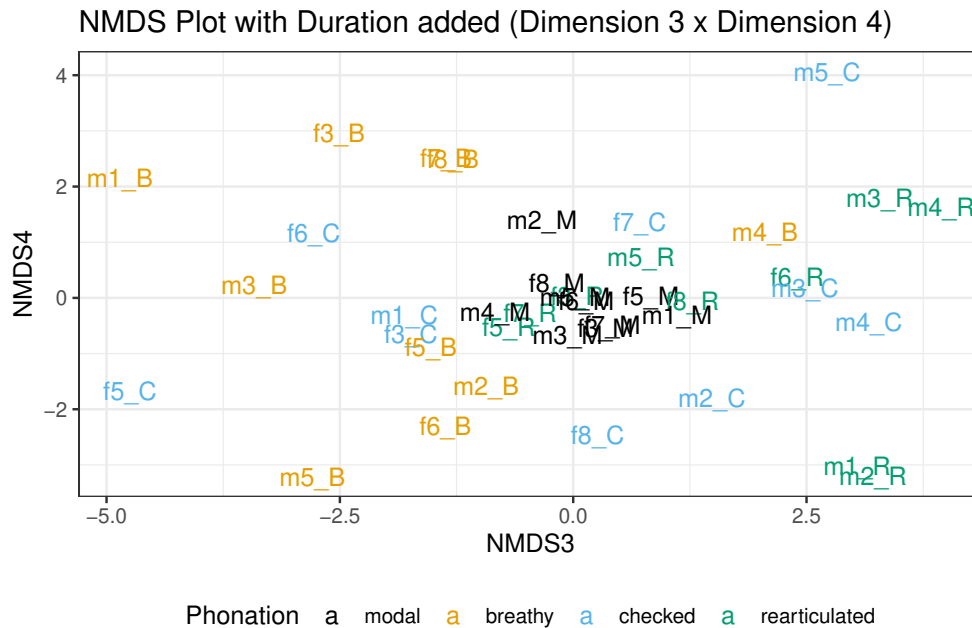
Figure 4.7: Two-dimensional MDS solution showing the third and fourth dimensions.



Open Glottis                                                    Closed Glottis

[h]          Breathy          Modal          Creaky          [ʔ]

Figure 4.8: A diagram showing the relationship between breathy, modal, and creaky phonation types from **gordonPhonationTypesCrosslinguistic2001**.

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

58

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 (**kruskalMultidimensionalScaling1978**, **hastieElementsStatisticalLearning2009**). 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 bolded. The choice to bold the four largest correlations was arbitrary and was done to make it easier to discuss. Even though only the four largest correlations are bolded, there are several correlations that share similar correlation values. These cases will be discussed as the are 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

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 |

**garellekPhoneticsVoice2019**, **garellekTheoreticalAchievementsPhonetics2022** for an overview of these measures). Because these measures all capture the same information and they are all highly correlated with the first dimension, we can conclude that the first dimension is primarily concerned with the spectral slope of the signal (i.e., the open-quotient of the glottis).

The first dimension is also positively correlated with the amplitude of the first three formants (i.e., 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 nonmodal). This suggests that the second dimension is concerned with the amount of noise in the signal.

Additionally, 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" (**mittalStudyEffectsVocal2014**). his suggests that the second dimension is primarily concerned with the amount of noise in the signal.

## 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 **keatingCrosslanguageAcousticSpace20** found in their study, the first dimension appears to roughly be similar to the open quotient of the glottis as proposed by **gordonPhonationTypesCrosslinguistic2001**. 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 **gordonPhonationTypesCrosslinguistic2001** is shown in Figure 4.9.

Open Glottis                                            Closed Glottis

[h]        Breathy        Modal        Creaky        [ʔ]
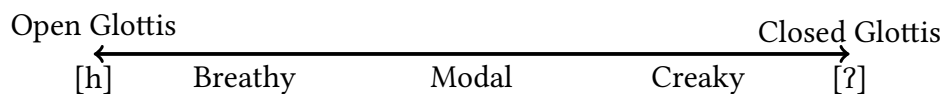
Figure 4.9: A diagram showing the relationship between breathy, modal, and creaky phonation types. Based on **gordonPhonationTypesCrosslinguistic2001**.

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 **gordonPhonationTypesCrosslinguistic2001**, 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 (**holmbergComparisonsAerodynamicElectroglottographic1995**, **kreimanMeasures**, **garellekModelingVoiceSource2016**, **garellekPhoneticsVoice2019**, **chaiH1H2Acoustic2022**).

The third dimension adds more information on nonmodal voice quality. Figure 4.3 and Figure 4.5, this third dimension separates the breathy voice from the other nonmodal 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 (**chaiH1H2Acoustic2022**, **brinkerhoffResidualH1Measure2**

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 deter-

mine which measures are the most important is by performing a bootstrap aggregating

version of a classification and regression tree analysis (**breimanClassificationRegressionTrees1986**,

**breimanBaggingPredictors1996**).

# 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 land-scape 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 **hastieElementsStatisticalLearning20** **jamesIntroductionStatisticalLearning2021** 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. 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 (**hastieElementsStatisticalLearning2009**, **jamesIntroductionStatisticalLearning2021**).

## 5.3 Decision trees in linguistics

The use of decision trees in linguistics is not new. One of the first uses was done by **tagliamonteModelsForestsTrees2012**, 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 (**keatingCrosslanguageAcousticSpace2023**).

In their study, **keatingCrosslanguageAcousticSpace2023** 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 **keatingCrosslanguageAcousticSpace2023**'s

(**keatingCrosslanguageAcousticSpace2023**) 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    What is bagging

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
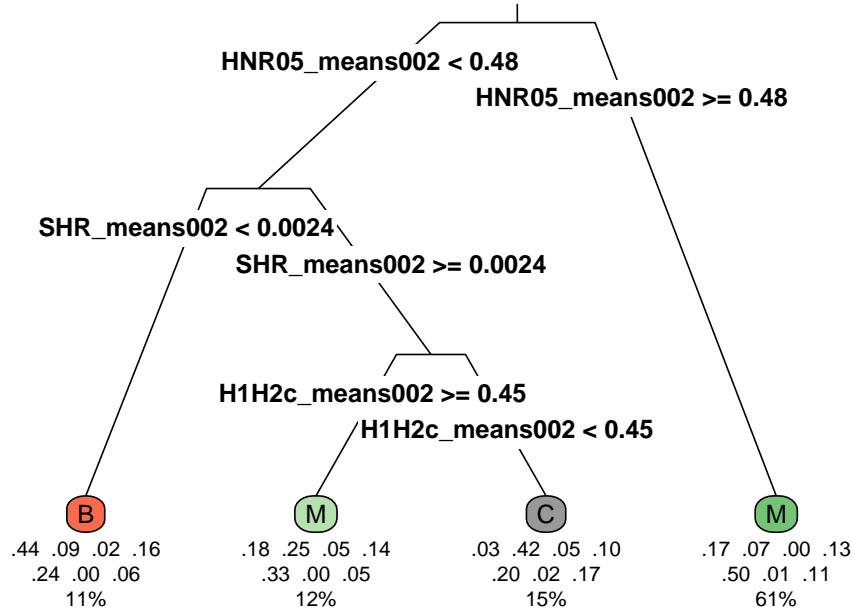
Figure 5.1: Classification tree of phonation categories from **keatingCrosslanguageAcousticSpace2023**. Abbreviations used in this figure are: HNR05_means002: harmonics-to-noise ratio over the frequency range from 0 Hz to 500 Hz for the middle third of each vowel; SHR_means002: subharmonic-to-harmonic 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.

is that decision trees do not have the same predictive accuracy as other regression or

classification models (see **hastieElementsStatisticalLearning2009** for discussion).

One way to overcome these disadvantages, is to make use of a technique called

bootstrap aggregating, or bagging (**breimanBaggingPredictors1996**). 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, 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.

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

### 5.5.1  Data processing

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 me 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 indicated that the data point was improperly tracked. This is comparable to what was done in **seyfarthPlosiveVoicingAcoustics2018**, **chaiCheckedSyllablesChecked2022**, **garellekPhoneticsV**

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 three was excluded. Additionally, residual H1* was calculated and included in the analysis.

Once these steps were completed, the mean of each vowel and speaker for the fourth through seventh intervals was taken. This is similar to what **keatingCrosslanguageAcousticSpa** 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. 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 was positive for the MDS analysis.

Using this data, I will use bagging trees to understand the importance of the dif-

ferent 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. The x-axis shows the number of trees that were used in the model and the y-axis shows the out-of-bag error. The out-of-bag error is a measure of how well the model predicts the data that was not used to train the model. The lower the out-of-bag error, the better the model is at predicting unseen data.
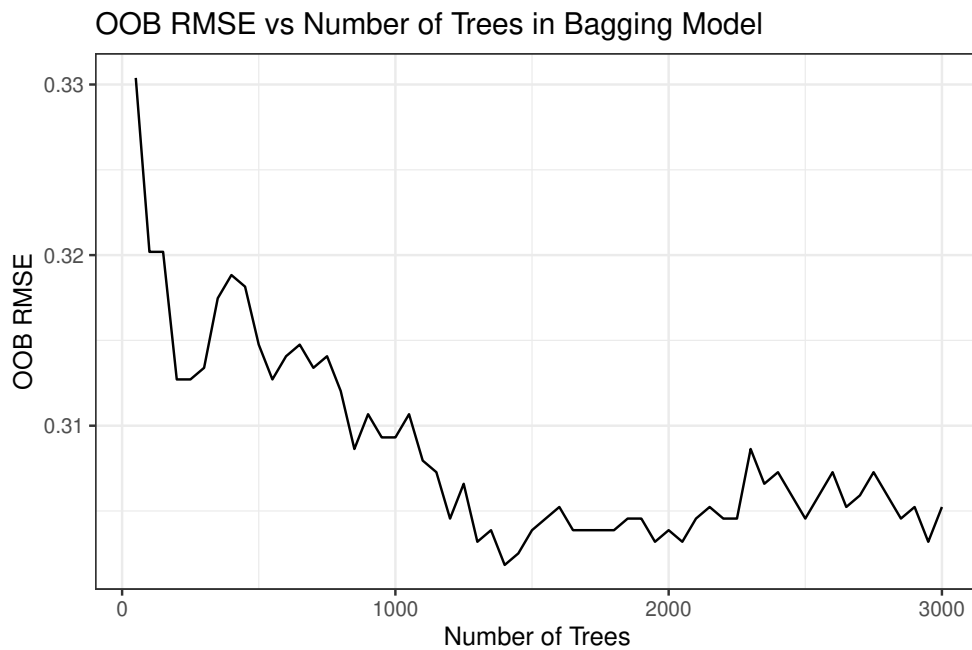


Figure 5.2: Plot showing the out-of-bag error as a function of the number of trees ran in the bagging models.

From Figure 5.2, it is clear that the out-of-bag error stabilizes at approximately 1400 trees after which the out-of-bag error increases and decreases while hovering around an out-of-bag error rate of 0.305. This means that the number of trees that are required

for the the bagging model to be the most stable is 1400 trees. In order to further insure

the stability of the model, I also ran a 10-fold cross-validation on the bagging model

with 1400 trees. The results of this model are shown in Figure 5.3. In interpreting

the variable importance plot we arrange the variables by their importance in making

the splits in the data. The x-axis shows the mean decrease in Gini index, which is a

measure of impurity, for each predictor and the y-axis shows the different predictors

that were used in the model. Typically, the top 3-5 predictors are the most important

in making the splits in the data (**boehmkeHandsOnMachineLearning2019**).
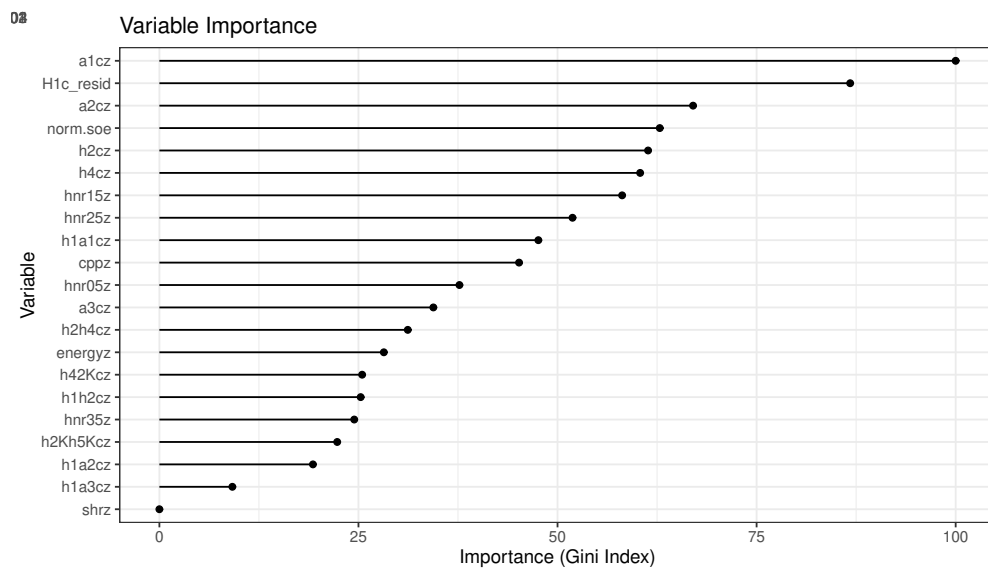


Figure 5.3: Variable importance plot showing the importance of the different acoustic
measures in the bagging model based on the total amount that the Gini index is de-
creased by splits over a given predictor, averaged over all 1400 trees.

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

## 5.6   Discussion of the bagging results

It is interesting to not that in both the MDS analysis and the bagging trees, the same three variables are found to play a role in the acoustic space of SLZ. These two analyses complement each other and help us to better understand what measures to focus are attention on. From Chapter 4, we saw that the acoustic measure A1* was heavily weighted for both dimensions 1 and 2. It is not surprising then that this measure would rank so high in the variable importance.

The other two measures that the bagging analysis found to be of most importance are the measures residual H1* and HNR 1500 Hz. These two measures were also important for the analysis in Chapter 4. However, they were only heavily weighted for dimension 3.

The rest of this section will discuss each of the measures that were found to be important in the bagging analysis and how they relate to the different voice qualities in SLZ and other linguistic phenomena.

### 5.6.1    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-jorgensenPhoneticAnalysisBreathy1968** 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 5.4, 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 (**brunelleRegisterEasternCham2005**, **brunelleDialectExperiencePerceptual2012 brunelleTonePhonationSoutheast2016**).

As described in **brunelleTonePhonationSoutheast2016**, many of the languages

A1*
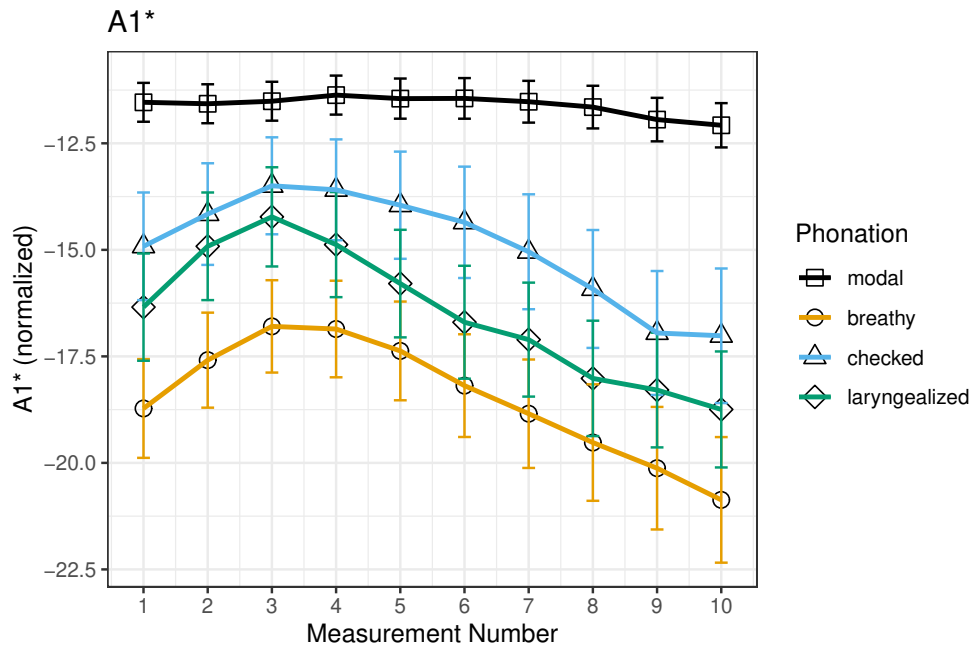


Figure 5.4: Plot showing the distribution of A1* across the different voice qualities in SLZ.

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" (**brunelleTonePhonationSoutheast2016**), and was first used by **hendersonMainFeaturesCambodian1952** to describe the categorical contrasts found in Khmer. The characteristics that define the higher and lower registers are found in Table 5.1.

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 (**lottoEffectVoiceQuality1997**). However, these studies do not

Table 5.1: Possible phonetic correlates of register. From **brunelleTonePhonationSoutheast2016**.

| 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., **chenAcousticCorrelatesEnglish1997**, **delvauxPerceptionContrasteNasalite2009**, **macmillanIntegralityNasalizationF11999**, **pruthiAcousticParametersAutomatic2004**, **schwartzAcousticsNormalNasal1968**, **stevensAcousticPhonetics2000**, **stylerAcousticalPerceptu**, **stylerAcousticalFeaturesVowel2017**). 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 5.4.

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* (**matisoffRhinoglottophiliaMysteriousConnection1975**, **ohalaPhoneticExplanationsNasal1975**, **ohalaPhoneticsNasalPhonology1993**, **bennettMayanPhonology2016**). In **blevinsEvolutionPona** **matisoffRhinoglottophiliaMysteriousConnection1975**, this association is attributed to the acoustic and perceptual similarities between nasalization and breathy voice. In

one study, **garellekBreathyVoiceNasality2016** 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 the contrast of breathy voice. Essentially, the reverse of what was reported by **garellekBreathyVoiceNasality2016**. 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.

## 5.6.2   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 the vowel.

In Figure 5.5, 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; **hastieGeneralizedAdditiveModels1986**, **woodGeneralizedAdditiveModels201** **soskuthyGeneralisedAdditiveMixed2017**, **wielingAnalyzingDynamicPhonetic2018**) shows that the temporal behavior of the two types of laryngealization is also correct. Checked voice is associated with a lower score toward the end of the vowel and rearticulated voice is associated with a lower score in the middle of the vowel. Additionally, the results of the GAMM show that for the first three time points there is no significant difference between breathy and modal voice. Suggesting that breathy voice is
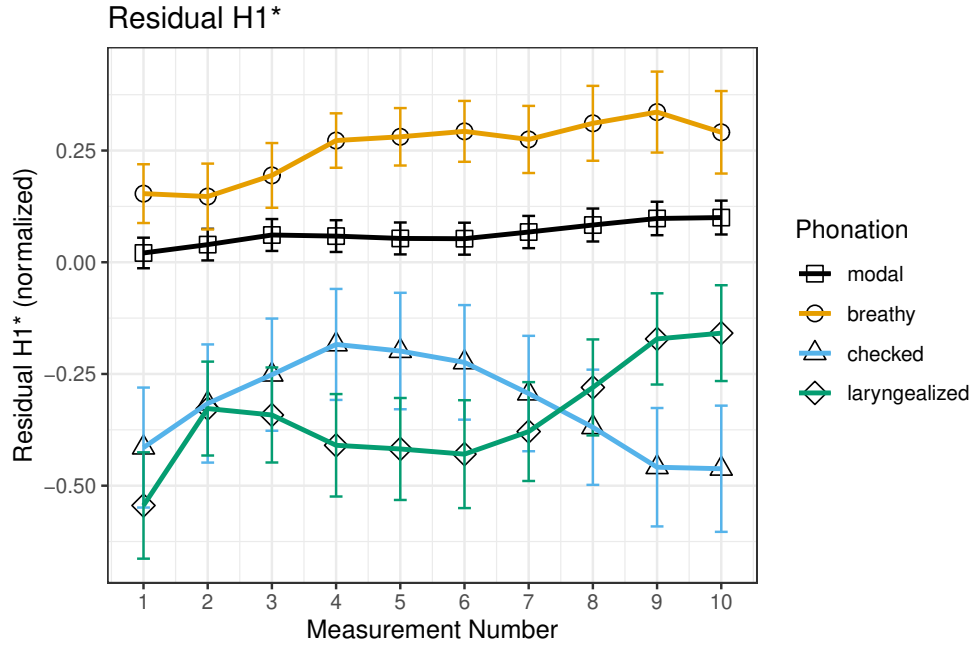
Figure 5.5: Plot showing the distribution of residual H1* across the different voice qualities in SLZ.

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.

### 5.6.3 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.,

**hillenbrandAcousticCorrelatesBreathy1996**, **blankenshipTimeCourseBreathiness1997**,
**kentVoiceQualityMeasurement1999**).

As seen in Figure 5.6, 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.
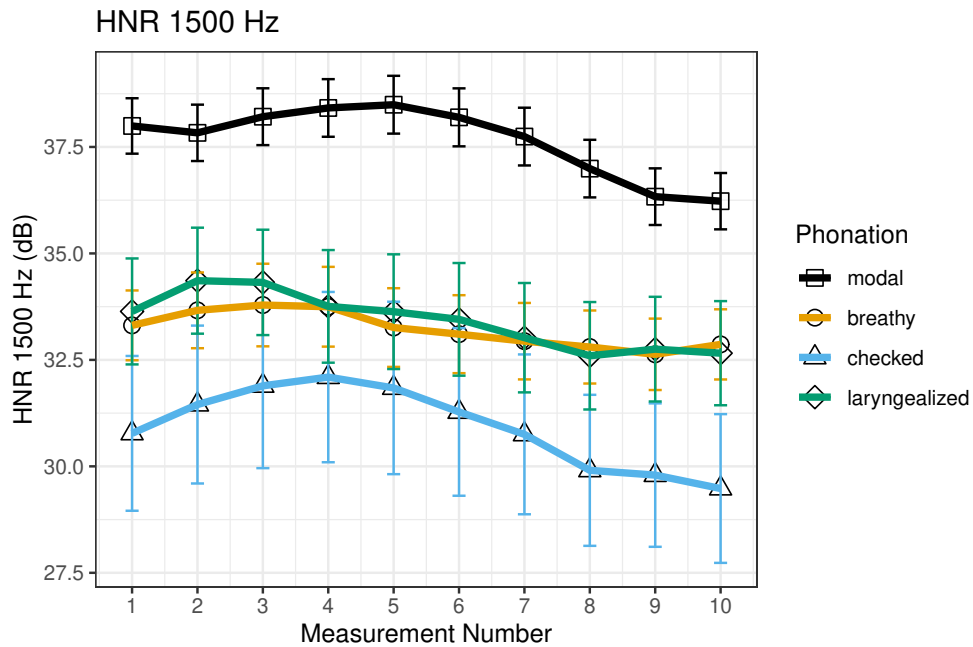


Figure 5.6: Plot showing the distribution of HNR < 1500 Hz across the different voice
qualities in SLZ.

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 in-
stances 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 HNR < 1500 Hz of the non-modal phonations.

This measure will be very important in Chapter 6, where this measure will be used to test the predictions of the laryngeal complexity hypothesis.

## 5.7   Conclusion

In conclusion, the results of the bagging analysis show that the acoustic measures A1*, residual H1*, and HNR < 1500 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 syllablic nucleus (**silvermanLaryngealComplexityOtomanguean1997**, **silvermanPhasingRecoverability1997**). This use of both tone and phonation on the same nucleus is one of the defining characteristics of the Oto-Manguean languages (**silvermanLaryngealComplexityOtomanguean1997**). 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 lan-

guages of Mpi and Tamang (**silvermanLaryngealComplexityOtomanguean1997**,
**silvermanPhasingRecoverability1997**) and the Mayan language Yucatec Mayan
(**frazierPhoneticsYucatecMaya2013**).

### 6.1.1   Phasing and recoverability

According to **silvermanLaryngealComplexityOtomanguean1997**, **silvermanPhasingRecover**
one of the defining aspects of laryngeal complexity is the concept of phasing and re-
coverability. 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 laryn-
geally 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 por-
tions of the vowel that can be analyzed separately and that need to be analyzed tempo-
rally rather than spectrally (**silvermanLaryngealComplexityOtomanguean1997**).

For example, in the Oto-Manguean language Jalapa Mazatec, breathiness or creak-
iness is realized only on the first portion of the vowel either as full laryngeal consonant
or as a laryngeal feature on the vowel (**silvermanLaryngealComplexityOtomanguean1997**).
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

86

or creakiness is phased with respect to the tone.

**silvermanLaryngealComplexityOtomanguean1997** 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

**silvermanLaryngealComplexityOtomanguean1997** 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 (**ritsmaFrequenciesDominantPerception1967**, **remezIntonationSinusoidalSentences1993**). 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 (**silvermanPhasingRecoverability** Creaky voice on the other had obscures the pulse periods due to its aperiodic and unstable glottal vibration (**ladefogedSoundsWorldLanguages1996**). This is what was observed in Mazatec where the harmonic structure is gone and the pulses are indiscernible (**kirkQuantifyingAcousticProperties1993**). Additionally, the perception

of pitch is rendered indiscernible when the pulse periods are varied by 10% or more (**rosenbergPitchDiscriminationJittered1966**).

These observations lead **silvermanLaryngealComplexityOtomanguean1997** 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 **silvermanLaryngealComplexityOtomanguean1997**'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 (**lindblomEconomySpeechGestures1983**). According to **lindblomEconomySpeechGestures1983**, 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 (**ohalaPhoneticExplanationsNasal19 chenAcousticCorrelatesEnglish1997**, **stylerAcousticalPerceptualFeatures2015**).

According to **silvermanLaryngealComplexityOtomanguean1997**'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, **ohalaProductionTone11978** 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 where held open or closed determined the type of phonation that was produced (**ladefogedSoundsWorldLanguages1996**). This aspect of phonation was shown in Figure 4.8 and repeated here as Figure 6.1.

For **silvermanLaryngealComplexityOtomanguean1997**'s **silvermanLaryngealComplexityOt**

Open Glottis ——————————————————————— Closed Glottis
[h]        Breathy         Modal         Creaky      [ʔ]

Figure 6.1: A diagram showing the relationship between breathy, modal, and creaky phonation types. Based on **gordonPhonationTypesCrosslinguistic2001**.

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., **eslingVoiceQualityLaryngeal2019**). 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 **silvermanLaryngealComplexityOtomanguean1997**'s **silvermanLaryngealC** 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 laryngeal-

ization interrupts a vowel and appears in the middle.

According to **silvermanLaryngealComplexityOtomanguean1997**, 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 **silvermanLaryngealComplexityOtomanguean1997** provides data from three Oto-Manguean languages: Jalapa Mazatec, Comaltepec Chinantec, and Copala Trique. These languages are shown in Table 6.1.

Table 6.1: Implicational hierarchy of laryngeal complexity. The symbols h and ʔ represent laryngealization. The symbol V represents where the modal vowel is located in relation to the laryngealization. Modified from **silvermanLaryngealComplexityOtomanguean1997**.

| Language | Prevocalic | Postvocalic | Interrupted |
|---|---|---|---|
| Jalapa Mazatec | hV˥, ʔV˥ | — | — |
| Comaltepec Chinantec | hV˥, ʔV˥ | Vh˥, Vʔ˥ | — |
| Copala Trique | hV˥, ʔV˥ | Vh˥, Vʔ˥ | VhV˥, VʔV˥ |

For many descriptions of languages with laryngeal complexity, the implicational hierarchy seems to hold. This is certainly the case in the other Trique languages (**dicanioPhoneticsPhonologySan2008**, **dicanioItunyosoTrique2010**, **dicanioCoarticulationTone**, **dicanioPhoneticsFortisLenis2012**, **dicanioCueWeightPerception2014**, **dicanioGlottalTogglingIt**, **elliottChicahuaxtlaTriqui2016**, **hollenbachPhonologyMorphologyTone1984**). However, as mentioned by **frazierPhoneticsYucatecMaya2013**, 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 fea-

ture on the vowel. Indeed, **silvermanLaryngealComplexityOtomanguean1997**,

**silvermanPhasingRecoverability1997** 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 (**dicanioPhoneticsPhonologySan2008**,

**dicanioItunyosoTrique2010**, **dicanioCoarticulationToneGlottal2012**, **dicanioPhoneticsFortisLe**

**dicanioCueWeightPerception2014**, **dicanioGlottalTogglingItunyoso2020**, **elliottChicahuaxtlaTr**

**hollenbachPhonologyMorphologyTone1984**), but in Jalapa Mazatec, the laryn-

gealization is realized as a laryngeal feature on the vowel (**kirkQuantifyingAcousticProperties1993**,

**garellekAcousticConsequencesPhonation2011**).

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

(**cortesSanSebastianMonte2023**, **eischensTonePhonationPhonologyPhonetics2022**,

**gerfenPhonologyPhoneticsCoatzospan1999**, **gerfenProductionPerceptionLaryngealized2005**).

Additionally, in many of this languages, the larngealization can appear either in the

middle of the vowel, what **silvermanLaryngealComplexityOtomanguean1997** calls

interrupted, or at the end of the vowel (**cortesSanSebastianMonte2023**, **eischensTonePhonationPho**

This is directly in contrast to the implicational hierarchy which states that if a lan-

guage 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-garciaPhonationTypesTones2018** for a typology of phonation in Zapotec languages). For example, **avelinobecerraTopicsYalalagZapotec2004**, **avelinoAcousticElectroglottographicAnalyses2010** 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 (**espositoSantaAnaValle2004**, **espositoAcousticElectroglottographicStudy2012**). 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, **frazierPhoneticsYucatecMaya2013** describes the phonetic properties of tone and voice quality in Yucatec Mayan. In this study, **frazierPhoneticsYucatecMay**

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 vowels[1] or as a vowel with creaky voice. **frazierPhoneticsYucatecMaya2013** 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-jorgensenPhoneticAnalysisStod1989**, **gronnumDanishStodLaryngealization2013**, **penaStodTimingDomain2022**, **penaProductionPerceptionStod2024**). 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 $f0$ (**fischer-jorgensenPhoneticAnalysisStod1989**, **gronnumDanishStodLaryngealization2013**). **penaStodTimingDomain2022**, **penaProductionPer**

---

[1] This is very similar to rearticulated vowels in SLZ, where a vowel has a period of glottalization in the middle of the vowel that is often realized as a glottal stop. This is different from how **bairdPhoneticPhonologicalRealizations2011** describes "broken" or "rearticulated" vowels in K'ichee' another Mayan language.

showed that even though Danish is not traditionally classified as a laryngeally complex language it still shows evidence of phasing between the primary and secondary cues to stød, with the primary cue of phonation being produced in the second half of the syllablic rhyme and the secondary cue of pitch being produced in the first half of the syllablic rhyme.

In contrast to the descriptive studies, instrumental studies have focused on the acoustic properties of laryngeal complexity, primarily how laryngealization affects $f0$ and the harmonic structure of the vowel. For example, **garellekAcousticConsequencesPhonation201** found that in Jalapa Mazatec, the laryngealization causes $f0$ perturbation in the first portion of the vowel. **garellekAcousticConsequencesPhonation2011** conclude that this is evidence for the phasing between laryngealization and tone as predicted by **silvermanLaryngealComplexityOtomanguean1997**'s **silvermanLaryngealComplexityOtomang** theory of laryngeal complexity. Additionally, this same phenomenon of $f0$ 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 (**dicanioCoarticulationT** For these previous studies the main piece of evidence for laryngeal complexity has been the perturbation of the $f0$ 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 $f0$ to determine if there is phasing. For example **wellerInteractionsToneGlottalization2023**,

**wellerLexicalToneVowel2023**, **wellerVoiceQualityTone2024** have investigated laryngeal complexity in San Sebastián del Monte Mixtec using a combination of $f0$ 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 $f0$ perturbations in the signal, these previous studies have missed the opportunity to look at the full range of acoustic properties. **wellerInteractionsToneGlottalization2023**, **wellerLexicalToneVowel2023**, **wellerVoiceQualityTone2024** have done an excellent job by showing that by looking at SoE, in addition to $f0$, 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 (**dekromCepstrumBasedTechniqueDetermining1993**, **ferrerriesgoWhatMakesCepstral2020**, **garellekPhoneticsVoice2019**). It is well

understood that aperiodicity is one of the defining characteristics of non-modal phona-
tion (**ladefogedSoundsWorldLanguages1996**). 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 de-
termine if there is a clear phasing between modal and non-modal phonation in SLZ us-
ing a generalized additive mixed model (GAMM) analysis (**hastieGeneralizedAdditiveModels1986**,
**woodGeneralizedAdditiveModels2017**, **soskuthyGeneralisedAdditiveMixed2017**,
**wielingAnalyzingDynamicPhonetic2018**).

## 6.3 Analysis of laryngeal complexity

## 6.4 Results

## 6.5 Discussion

**humbertConsonantTypesVowel1978**

## 6.6 Conclusion

# Chapter 7

# Modeling laryngeal complexity

## 7.1 The Laryngeal Articulator Model

eslingThereAreNo2005, eslingVoiceQualityLaryngeal2019, moisikPhonologicalPotentialsL

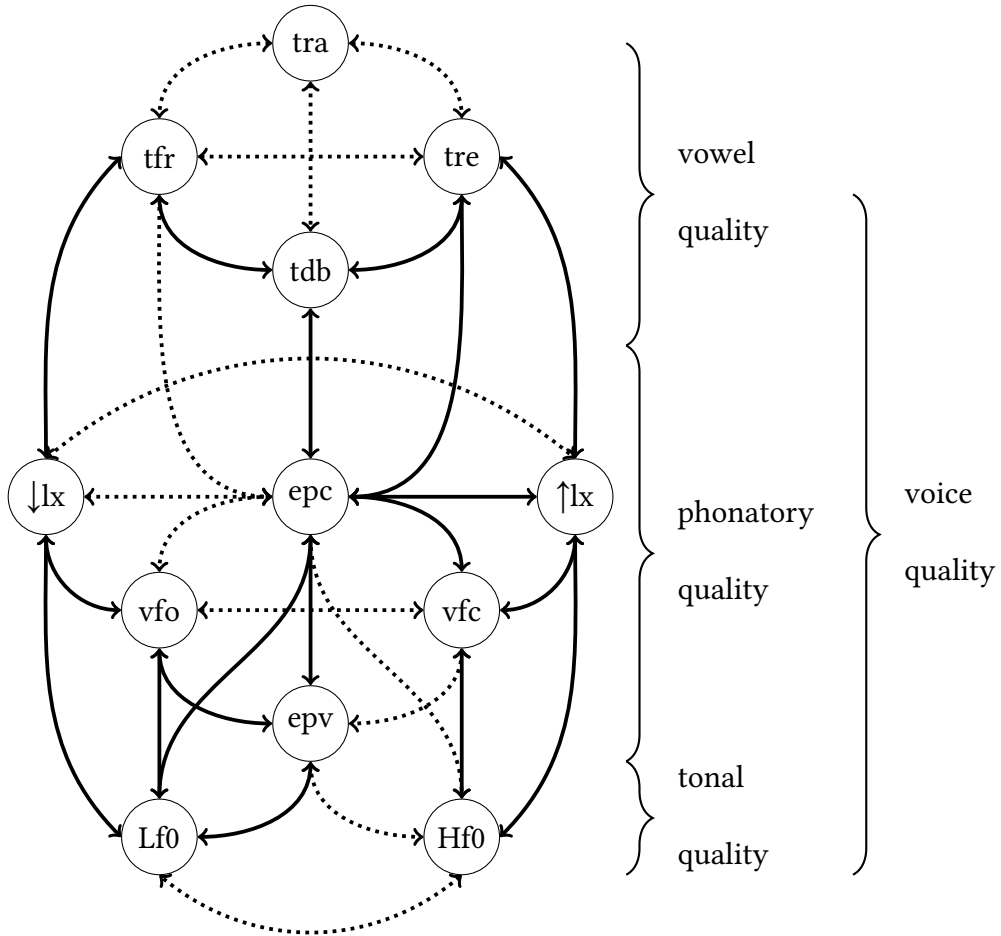moisikMultimodalImagingGlottal2015, moisikModelingBiomechanicalInfluence2014

Figure 7.1: The Laryngeal Articulator Model from **eslingVoiceQualityLaryngeal2019**. 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.

## 7.2 Modeling laryngeal complexity

## 7.3 Alternative accounts

### 7.3.1 Articulatory Phonology account

### 7.3.2 Q-theory account

## 7.4 Conclusion

# Chapter 8

# Conclusion

# Appendix A

# Some Ancillary Stuff

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