

Acoustic discriminability of phonation in Santiago Laxopa Zapotec*

Mykel Loren Brinkerhoff

2023-06-05

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 2 |
| 2 | Background | 3 |
| 2.1 | Psychoacoustic Model for Phonation | 5 |
| 3 | Phonation contrasts in Santiago Laxopa Zapotec | 5 |
| 3.1 | Interaction of Tone and Phonation | 8 |
| 4 | Methodology | 10 |
| 4.1 | Acoustic measurements | 10 |
| 5 | Results | 11 |
| 5.1 | Linear Discriminant Analysis | 11 |
| 5.2 | Statistical results on the output of the LDA | 13 |
| 6 | Discussion | 19 |
| 6.1 | Linear discriminant analysis discussion | 19 |
| 6.2 | Harmonic-to-noise ratio discussion | 20 |
| 6.3 | Laryngealized vowel discussion | 21 |

*I am grateful to Fe Silva-Robles and Raúl Díaz Robles for sharing their time and language expertise. I am also grateful to Grant McGuire, Jaye Padgett, Rachel Walker, Maziar Toosarvandani, Ben Eischens, Kim Tan, and Zach Horton for their help and discussions during all stages of this project. This project branched off from collaborating with Jack Duff and Maya Wax Cavallaro.

This work was supported in part by the National Science Foundation under Grant No. 2019804, the Humanities Institute at UC Santa Cruz, and the Jacobs Research Funds.

| | |
|---|-----------|
| 6.4 Laryngeal Complexity Hypothesis | 21 |
| 7 Further directions and Conclusions | 23 |
| References | 24 |

1 Introduction

Non-modal phonation is a common phenomenon in many of the world’s languages. Phonation describes how speakers alter the larynx to produce different sound qualities. Most frequently, the larynx is manipulated to produce sounds that vary from being more breathy or creaky.¹ In languages such as English, these characteristics are described as being pathological or simply the characteristic of a given speaker (e.g., Klatt & Klatt 1990). In other languages such as Gujarati, these phonation types are used phonemically where vowels can be breathy or modal (Fischer-Jørgensen 1968). This phonemic use of phonation is particularly common in the Oto-Manguean language family from southern Mexico (e.g., Suárez 1983, Campbell, Kaufman & Smith-Stark 1986, Silverman 1997, Campbell 2017a,b).

One common problem facing linguists studying phonation is determining the acoustic correlates for these different phonation types. Since Fischer-Jørgensen (1968), it has been widely assumed that certain markers in the acoustic signal define different types of phonation. The most common measurements invoked are spectral-tilt measurements and harmonics-to-noise ratios. Spectral-tilt measurements typically involve looking at the relative amplitude of different harmonics in the speech signal. Harmonics-to-noise ratios typically involve looking at how much energy there is in that same speech signal and are typically excellent indicators of periodicity.

Over the years, different authors have proposed different measurements for the phonation types in any given language. For example, Fischer-Jørgensen (1968) proposed that H1-H2, the difference in amplitude between the first and second harmonics, was the best measurement for accounting for the difference between breathy and modal vowels in Gujarati. Because phonation is acoustically multi-dimensional, there are many measurements that researchers can rely on when they research a given phonation system (Garellek 2020). This multitude of acoustic measurements presents a challenge for researchers to make cross-comparisons between and come to a common understanding of what defines a phonation type. Additionally, with many measurements, knowing which measurements are important for the production and perception of different phonation types presents a challenge. One question I address in this paper is the challenge of knowing what

¹Other types of phonation also exist but are not as frequently employed for linguistic expression (see Esling et al. 2019 for a detailed discussion on the different phonation types that exist and how the larynx produces them).

measurements are most relevant for analyzing the four phonation contrasts in Santiago Laxopa Zapotec, an understudied Northern Zapotec language from Oaxaca, Mexico.

I address this question within the framework proposed by Kreiman et al. (2014). This framework, which the authors call the PSYCHOACOUSTIC MODEL FOR PHONATION, provides researchers with a set of acoustic measures for a standardized way to investigate, compare, and theorize how phonation is produced and perceived. Further information about this framework is given in Section 2. By using this framework for investigating the phonation contrasts in Santiago Laxopa Zapotec, I contribute to the question of the efficacy of this framework in linguistic research. Following Garellek (2020), I use a linear discriminant analysis (Fisher 1936) to show that the Psychoacoustic Model of Phonation does explain the phonation contrasts in Santiago Laxopa Zapotec. This analysis shows that various harmonic-to-noise ratios and Strength-of-Excitation contribute the greatest amount of information in discriminating the phonation types.

The rest of this paper is structured as follows. In Section 2, I first present information about phonation and how linguists have attempted to capture these contrasts acoustically. Section 3 discusses the phonation contrasts in Santiago Laxopa Zapotec. A general discussion about the elicitation methods and how data for this paper has been processed for the linear discriminant analysis follows in Section 4. I present the results of the linear discriminant analysis in Section 5.1 and discuss these results in Section 6.

2 Background

In studying the phonetics of phonation contrasts, linguists have discovered multiple methods of describing and accounting for those contrasts. One of the primary ways this is accomplished is by using different acoustic markers in the signal. The most frequent type of measurement is spectral-tilt. Spectral-tilt measurements are the difference in amplitude between harmonics. One of the first studies using spectral-tilt was Fischer-Jørgensen 1968. Fischer-Jørgensen showed that the difference between the amplitude of the first and second harmonics (H1-H2) could account for the differences between breathy and modal vowels. Indeed, many early studies found these spectral-tilt measurements are particularly useful in languages with complex phonation systems such as Green Hmong (Huffman 1987, Andruski & Ratliff 2000) and Jalapa Mazatec (Silverman et al. 1995, Blankenship 1997).

Further research has continued to rely primarily on H1-H2 to distinguish different types of phonation (e.g., Huffman 1987, Klatt & Klatt 1990). Over time, various researchers began experimenting with whether or not the difference between other harmonics besides H1-H2. Over time, other researchers have proposed using other measurements besides H1-H2. These include looking at H2 minus the amplitude of the fourth harmonic (H4) or the amplitude of the harmonic

closest to the different formants, which are termed A1 through An where the number corresponds to which formant the harmonic is closest to.

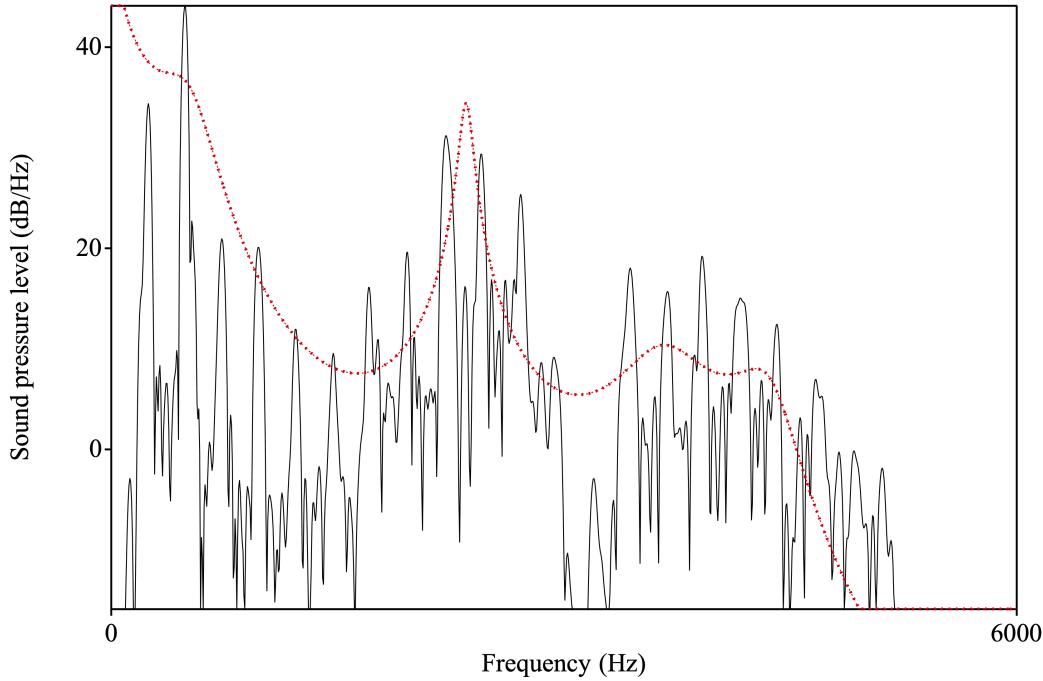


Figure 1: Spectral slice with LPC smoothed line overlaid for the vowel [e]. Each of the solid peaks represents the harmonics in the spectral slice. The leftmost black solid line peak is the first harmonic (H1), and each subsequent peak represents the next highest harmonic (H2 through Hn). The red dotted line represents an LPC smoothed line which identifies the formants by the peaks in the line.

From these studies, several patterns have emerged (see Garellek 2019 for an overview). When interpreting the results of these spectral-tilt measures, several general patterns correspond to the different phonation types. For example, vowels produced with a breathy phonation typically have a higher spectral-tilt measurement when compared to vowels produced with modal phonation, and vowels produced with creaky phonation typically have a lower spectral-tilt measurement than modal vowels.

In addition to spectral-tilt measurements, noise measurements frequently accompany phonation investigations (e.g., Garellek & Esposito 2021).

One of the most commonly used of these noise measurements is Cepstral Peak Prominence (CPP; Hillenbrand, Cleveland & Erickson 1994, Hillenbrand & Houde 1996). CPP is similar to the harmonics-to-noise ratio measure of de Krom (1993) but differs in how the ‘prominence’ of the cepstral peak is calculated. Prominence is taken as the difference in amplitude of the cepstral peak, and a regression line is used to normalize window size and overall energy. In interpreting the

measurement, a more prominent cepstral peak indicates stronger harmonics above the spectrum floor (i.e., greater periodicity in the speech signal). CPP was originally used as a diagnostic for breathy or modal voice (Blankenship 2002, Esposito 2010b) In fact, Esposito showed that CPP was the best of the eight measures she considered for distinguishing modal from breathy phonation types. Further research has shown that CPP is also a good measurement for any non-modal phonation (e.g., Andruski & Ratliff 2000, Andruski 2006, Blankenship 2002, Wayland & Jongman 2003, Avelino 2010).

2.1 Psychoacoustic Model for Phonation

The PSYCHOACOUSTIC MODEL FOR PHONATION (PMoP; Kreiman et al. 2014) is a theoretical and clinical framework that addresses the problems that linguistics and speech pathologists face when working with phonation. The PMoP was designed to assist linguists and clinicians in understanding phonation from the perspectives of production and perception. The goals of this framework are: (i) link perception to acoustics by explaining quality in terms of perceptually valid acoustic measures that combine to determine voice quality fully; (ii) link voice production to acoustics and perception by determining which changes in the physiological voice source produced perceptible changes in the acoustic signal; and (iii) iterate until the two sets of acoustic parameters align. Kreiman et al.

Table 1: Components of the psychoacoustic model of voice quality and associated parameters (from Kreiman et al. 2014).

| Model component | Parameters |
|-------------------------------------|------------------------------------|
| Harmonic source spectral slope | H1-H2 |
| | H2-H4 |
| | H4-H2k Hz |
| | H2k Hz-H5k Hz |
| Inharmonic source noise | Harmonics-to-noise ratio |
| Time-varying source characteristics | f_0 track |
| | Amplitude track |
| Vocal tract transfer function | Formant frequencies and bandwidths |
| | Spectral zeros and bandwidths |

3 Phonation contrasts in Santiago Laxopa Zapotec

Santiago Laxopa Zapotec (SLZ; *Dilla'xhunh Laxup*) is a Northern Zapotec language spoken by approximately 1000 people in the municipality of Santiago Laxopa, Ixtlán District in the Sierra

Norte of Oaxaca, Mexico (Adler & Morimoto 2016, Adler et al. 2018, Foley, Kalivoda & Toosarvandani 2018, Foley & Toosarvandani 2020). It is mutually intelligible with San Bartolomé Zoogocho Zapotec (Long & Cruz 2005, Sonnenschein 2005). SLZ has a fairly standard five-vowel inventory; see Table 2.

Table 2: Vowels inventory in Santiago Laxopa Zapotec.

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

Among Zapotecan languages, it is quite common for languages to make use of contrastive phonation (e.g., Avelino 2004, Long & Cruz 2005, Avelino 2010, López Nicolás 2016, Chávez-Péon 2010). SLZ, in addition to the five vowel qualities mentioned above, has four contrastive phonation types: modal, breathy, checked, and laryngealized. These contrasts are exemplified in the minimal quadruple in (1).

In representing the checked and laryngealized vowels, I follow the same procedure as other authors (e.g., Avelino (2010) and Uchihara (2016)) in representing the ‘glottal stop’ element as a superscript glottal stop in the IPA transcription (i.e., [a[?]]). This is primarily done as a way of standardizing the variable pronunciation of the glottal element in Zapotec, ranging from a full glottal stop (i.e., [a?]) to a creaky portion of the vowel (i.e., [a[?]]).

(1) Four-way near minimal phonation contrast

- a. *yag* /çag^L/ ‘tree; wood; almúd (unit of measurement approximately 4kg)’
- b. *yah* /ça^L/ ‘metal; rifle; bell’
- c. *yu'* /çu'^L/ ‘earth’
- d. *ya'a* /ça'a^L/ ‘market’

Breathy phonation on vowels is characterized by a raspy quality throughout the whole vowel or a portion toward the end of the vowel; see Figure 2.

On the other hand, checked vowels are characterized by an abrupt glottal closure which cuts the vowel short. This phonation is sometimes only realized as a very short period of creakiness at the end of the vowel; see Figure 3.

Laryngealized vowels are common in Zapotecan languages and have received many names. Previous descriptions have used terms such as broken, rearticulated, interrupted, and creaky to describe this phonation type (Long & Cruz 2005, Avelino 2004, 2010, Sonnenschein 2005, Adler &

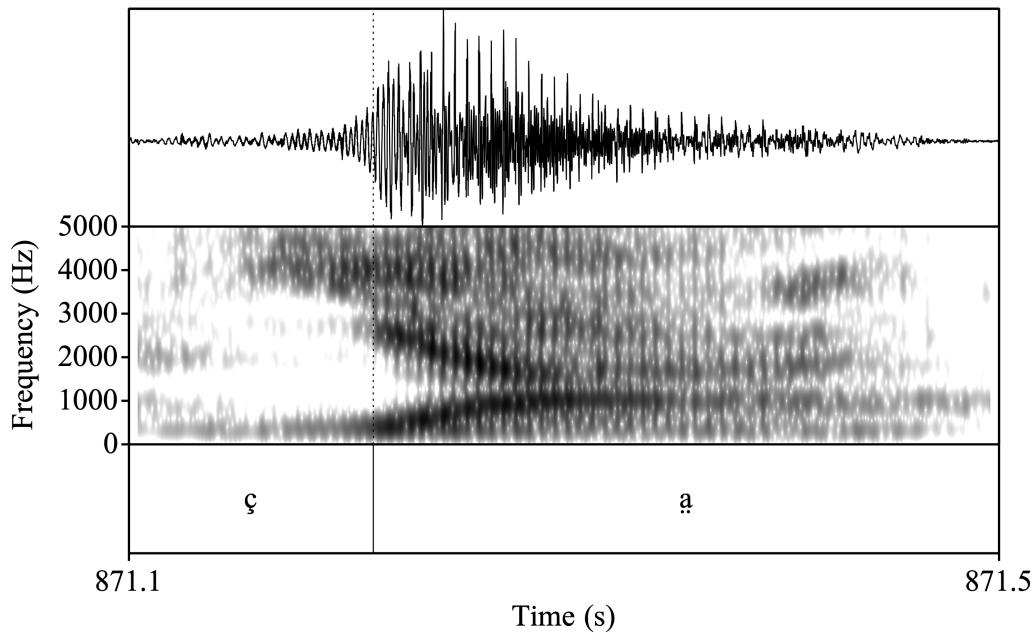


Figure 2: Breathy vowel in the word *yah* ‘metal; rifle’

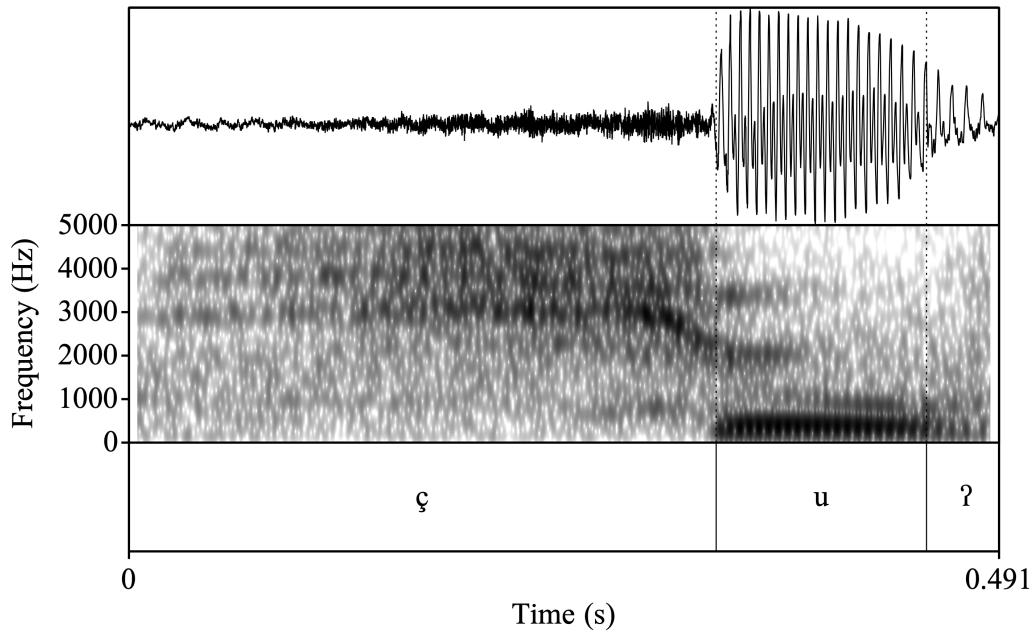


Figure 3: Checked vowel in the word *yu* ‘earth’

Morimoto 2016). To avoid confusion; I will use the term laryngealized following Avelino (2010). In addition to many different names, these vowels exhibit a wide range of allophones.

Avelino (2010) found in the closely related Yalálag Zapotec that among his consultants, there were at least four different pronunciations as seen in Table 3. In SLZ, each consulted language

Table 3: Laryngealized Vowels in Yalálag Zapotec

| | |
|----------|-------|
| /V'V/ | [V?V] |
| [VV̚] | |
| [VV̚:V̚] | |
| [VV̚V̚] | |
| | |

expert would produce this vowel differently. One consultant would re-articulate with a full glottal stop in the middle of the vowel, or they would produce creaky voice. This alternation seemed to be in free variation, but there was a greater tendency to creak in low-toned words, such as *xa'ag* [sa:g] ‘topil’², see Figure 4 for a comparison between this consultant’s pronunciation of the laryngealized vowels.

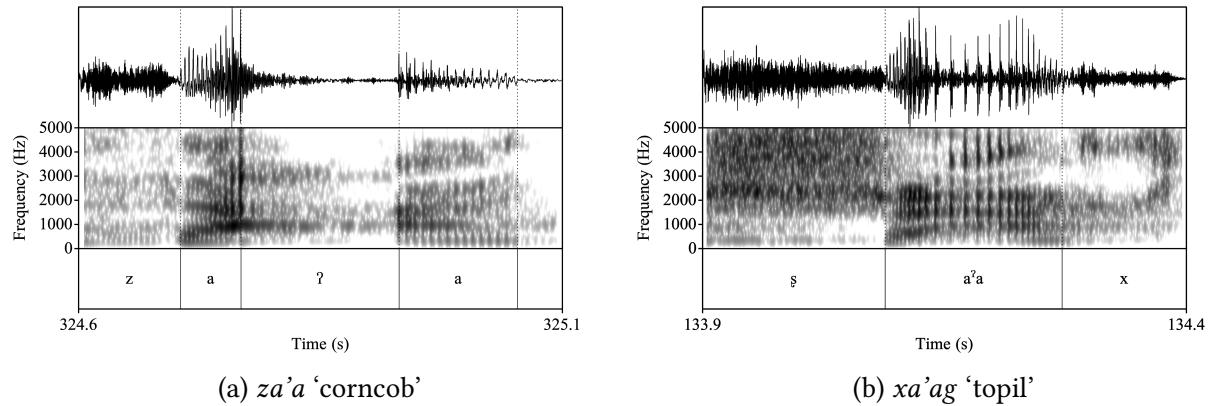


Figure 4: Comparison of FSR’s laryngealized vowels in *za'a* ‘corncob’ and *xa'ag* ‘topil’

The other consultant only produces creaky voice for these vowels regardless of the tone of the word. During one of the elicitation sessions, we conducted a perceptual check that these were, in fact, the same vowels, and both consultants reliably identified the words. They produced laryngealized vowels according to their own idiosyncrasies. However, a more detailed perception study is beyond the scope of this paper.

3.1 Interaction of Tone and Phonation

Most previous work on the interaction of tone has been focused on the languages of East and Southeast Asia (e.g., Masica 1976, Thurgood 2002, Yip 2002, Enfield 2005, Michaud 2012, Brunelle &

²A *topil* is a type of government office in traditional Oaxacan communities somewhat akin to a sheriff.

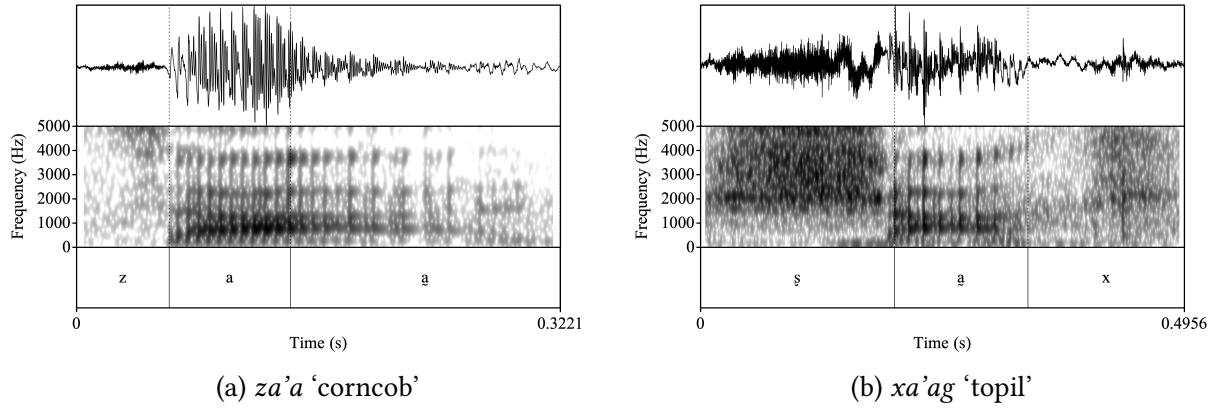


Figure 5: Comparison of RD's laryngealized vowels in *za'a* 'corncob' and *xa'ag* 'topil'

Kirby 2016). What has been found in these descriptions is that certain tones and phonations are codependent. For example, Smalley (1976) and Ratliff (1992) both describe White Hmong's -g tone as being a mid-low tone with breathy phonation and Mandarin's tone 3 is often associated with creaky phonation (Hockett 1947). Brunelle (2009) found that creaky phonation plays an important role in producing certain tones. Additionally, work on S'gaw Karen has found that two tones are only differentiated by some form of non-modal phonation (Boehm p.c.).

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

Chávez-Péón (2010) describes the tone and phonation interactions in San Lucas Quiaviní Zapotec (SLQZ), a central valley variety of Zapotec. The distribution of tone and phonation is found in Table 4. We see that in SLQZ, both low- and falling-tones have the full range of possible combinations. However, we see gaps in the high-tone for breathy and rising tones that can only occur with modal phonation.

Table 4: SLQZ tone and phonation interactions (Chávez-Péón 2010).

| | Modal | Breathy | Creaky | Interrupted |
|---------|-------|---------|--------|-------------|
| High | ✓ | – | ✓ | ✓ |
| Low | ✓ | ✓ | ✓ | ✓ |
| Falling | ✓ | ✓ | ✓ | ✓ |
| Rising | ✓ | – | – | – |

Based on elicitation data collected from 2020-2022, SLZ has a more expansive distribution of

tone and phonation when compared to SLQZ but seems to be very similar to other Northern Zapotec varieties (e.g., Avelino 2004). The distribution of SLZ tonal and phonation interactions are given in Table 5 for the number of analyzed tokens.

Table 5: SLZ tone and phonation interactions.

| | Modal | Breathy | Checked | Laryngealized |
|----------|--------------|----------------|----------------|----------------------|
| High | ✓ | – | ✓ | ✓ |
| Mid | ✓ | ✓ | ✓ | ✓ |
| Low | ✓ | ✓ | ✓ | ✓ |
| High-Low | ✓ | ✓ | ✓ | ✓ |
| Mid-High | ✓ | ✓ | – | ✓ |

One of the striking things in this is the lack of high tone with breathy phonation. This gap is interesting because of the long-time association of high pitch with breathiness (a good overview of this association and other phonation types—is found in Esling et al. 2019). This breathy phonation and high tone gap are common across the Zapotecan languages (Campbell p.c.). Regarding breathy phonation in SLQZ, Uchihara (2016) offers some convincing evidence that the phonation originated in syllables with low tone and then spread to other tones via analogy. If this is also true for SLZ, then we should be able to find similar distributions that Uchihara observes when we compare SLZ to its closest relative, San Bartolomé Zoogocho Zapotec, which lacks a breathy phonation. Such an investigation would be important in understanding how breathy vowels originated in the Zapotecan family, but it is beyond the scope of this paper.

4 Methodology

18 native language speakers of SLZ participated in this study (seven male). In this study, I present data from 6 speakers (three male). All speakers live in Santiago Laxopa Cruz, Ixtlán, Oaxaca, Mexico. The data was collected using a Zoom H4nPro handheld recorder (44.1kHz, 16-bit). Participants were recorded saying approximately 76 words in isolation, and the carrier sentence *shnia’ X chonhe lhas* [ʃnia’ X tʃone ras] ‘I say X three times’. This phrase was repeated three times.

4.1 Acoustic measurements

The vowels from the target words in carrier sentences were annotated in Praat (Boersma & Weenink 2021). Following Garellek (2020), the vowel onset was set to the second glottal pulse after the beginning of the vowel, and the vowel offset was set to the last glottal pulse before the

drop in amplitude of the following consonant or change in formant amplitude when the following consonant was a sonorant (e.g., nasal).

VoiceSauce (Shue, Keating & Vicenik 2009) was used to compute the different acoustic measurements across each vowel. Each vowel was resampled at 16kHz. The fundamental frequency (f_0) was calculated using STRAIGHT (Kawahara, Chevaigne & Patterson 1998). Formants and bandwidths were calculated using SNACK (Sjölander 2004).

A total of 1785 vowels were initially considered for analysis. Outliers were removed based on f_0 and vowel formants distance. 28 vowels whose f_0 had a Mahalanobis distance greater than 6 were removed as outliers.³ An additional 85 vowels were removed as outliers because their F1 and F2 had a standard deviation greater than 3. This resulted in a total of 1672 vowels being used for this analysis.

Because phonation in SLZ involves dynamic changes (i.e., glottalization occurring in the middle or end of the vowel for laryngealized and checked vowels), each acoustic parameter had three different associated measurements. These three measures are the average over the entire vowel, the change in the parameter from the beginning to the middle of the vowel, and the change in the parameter from the middle to the vowel's end. These last two measurements will be called "delta" measurements. This method is not unusual and was used by Garellek (2020) for the phonation contrasts in !Xóõ.

5 Results

5.1 Linear Discriminant Analysis

A Linear Discriminant Analysis (LDA: Fisher 1936) is a type of dimension reduction technique in statistics. What this means in practice is that an LDA takes multiple variables as the input to the model and reduces those multiple variables into a few new variables called linear discriminant functions (LD). These LDs are then used to identify which original variables affect the data most. For this study, an LDA was conducted in R (R Core Team 2021) using the MASS package from Venables, Ripley & Venables (2002).

In the case of SLZ phonation, each of the different acoustic measurements from Kreiman et al.'s (2014) were treated as the variables that the LDA used to determine the new LDs. For the statistical model to not be unduly swayed by large numerical differences in the raw numbers, it is standard that all variables are normalized. In this instance, all acoustic measurements proposed

³The Mahalanobis distance is defined as the distance between a point x and a distribution with mean μ and covariance matrix Σ . It is calculated using the following formula: $D(x) = \sqrt{(x - \mu)^T \Sigma^{-1} (x - \mu)}$. It is often used for outlier detection, and points with a significantly large Mahalanobis distance from the distribution are considered potential outliers.

by the PMoP were standardized by speaker using a z-score transformation except for the formants and bandwidths, which were standardized by speaker and vowel quality. Strength-of-Excitation (SoE) was standardized by speaker following Garellek et al. (2021) where each SoE value was: first log-normalized, then the maximum and minimum log(SoE) values were calculated for each speaker. Finally, the log(SoE) values were subtracted from the minimum value and divided by the difference between the maximum–minimum values.

The LDA model was trained on 70% of the total data and then tested on the remaining 30%. The number of LDs produced by the model is always the number of categories you try to predict minus one. Because SLZ has four phonation types, the LDA produced three LDs which accounted for 100% of the variation in the data. The first linear discriminant (LD1) accounted for 53.24% of the variation, while the second (LD2) and third (LD3) linear discriminant functions accounted for 30.73% and 16.03%, respectively. The confusion matrix from the LDA is shown in Table 6; overall, the accuracy of the model’s classification of the test set was 70.45%, much higher than chance at 25%.

Table 6: Confusion matrix from the linear discriminate analysis for the testing set.

| Actual → Predicted ↓ | modal | breathy | checked | laryngealized |
|-------------------------|-------|---------|---------|---------------|
| modal | 294 | 28 | 25 | 24 |
| breathy | 16 | 32 | 9 | 15 |
| checked | 13 | 5 | 22 | 3 |
| laryngealized | 4 | 6 | 3 | 12 |

Modal vowels represent 64% of the training data, breathy vowels represent 15.2%, checked vowels represent 10.8%, and laryngealized vowels represent 10%. Modal vowels were the most accurately discriminated (89.9%), whereas laryngealized vowels were the least accurately discriminated (22.2%). All phonation types are most confusable with modal ones. Additionally, laryngealized vowels were frequently confused with breathy vowels.

Within each LD, the different acoustic measures are assigned a coefficient indicating the measure has strength on the LD. When evaluating which measures had the greatest strength you take the absolute value of the coefficient and look for the largest coefficients. This evaluation produced the results reported in Table 7 (only the three strongest measures are shown for each LD).

LD1 discriminates modal vowels from the other phonation types and is most strongly associated with HNR <2500 Hz, HNR <3500 Hz, and SoE; see Figure 6.

LD2 discriminates breathy and checked vowels from the other phonation types and is most strongly associated with HNR <2500 Hz and SoE. HNR <500 Hz is also strongly associated with

Table 7: Acoustic measures with the highest absolute coefficient values for the three linear discriminant functions.

| LD1 | LD2 | LD3 |
|--------------|--------------|-----------------------------|
| HNR <2500 Hz | HNR <2500 Hz | HNR <2500 Hz |
| HNR <3500 Hz | HNR <500 Hz | HNR <3500 Hz |
| SoE | SoE | Δ HNR <2500 Hz (end) |

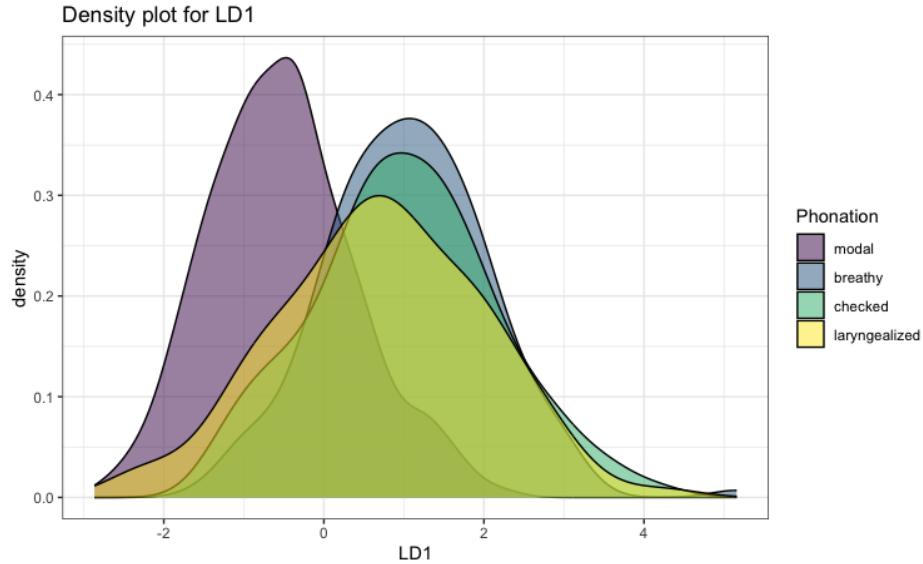


Figure 6: Density plot showing LD1 grouped by SLZ’s phonation types.

LD2; see Figure 7.

LD3 discriminates laryngealized vowels from the other phonation types and is also highly associated with the same measures as LD1 except for SoE, which is instead replaced by Δ HNR <2500 Hz (end), which represents the difference in HNR <2500 Hz from the middle to the end of the vowel; see Figure 8.

5.2 Statistical results on the output of the LDA

The five resulting acoustic measures from the LDA were then analyzed for statistical significance through a mixed-effects linear regression using R’s lme4 package (Bates et al. 2015). The acoustic measures were treated as the dependent variable in each linear regression model, with phonation as the predictor. The interaction between speaker and phonation was treated as a random variable, as was each of the token words. When evaluating the mixed-effects linear regression models, the modal vowel was also treated as the intercept of the model. This means the regression

³DV ~Phonation + (Phonation|Speaker) + (1|Word)

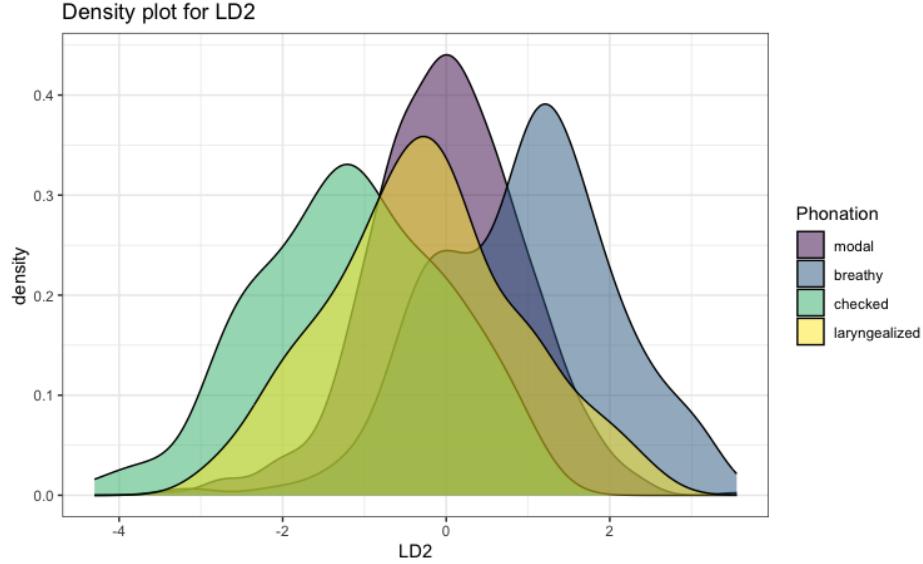


Figure 7: Density plot for LD2 grouped by SLZ's phonation types.

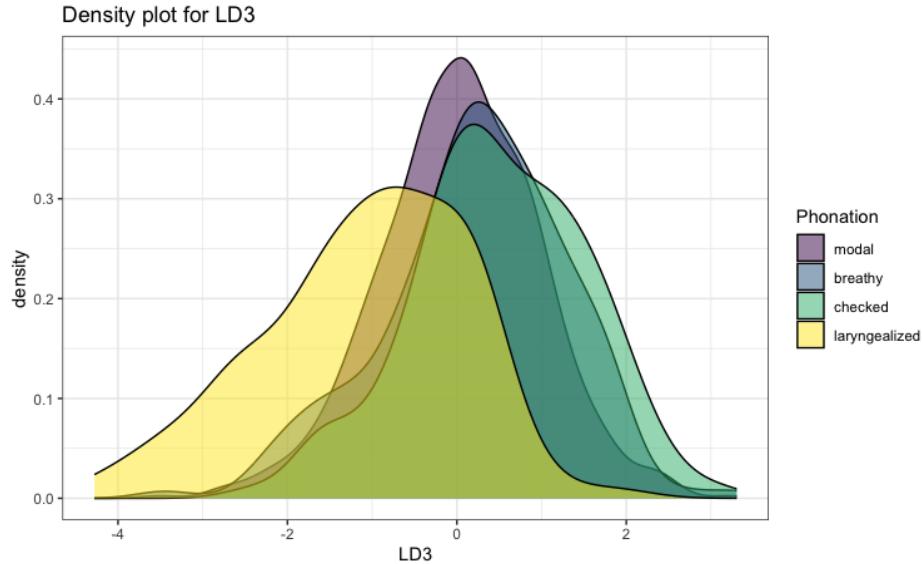


Figure 8: Density plot for LD3 grouped by SLZ's phonation types.

model results for each non-modal phonation type are made to the modal vowel.

The first linear discriminant function was most strongly affected by HNR<2500 Hz, HNR<3500 Hz, and Strength-of-Excitation. The two HNR measures are associated with measures of periodicity but in different bandwidths. HNR<2500 Hz is concerned with measuring periodicity below 2500 Hz. Because non-modal phonation is associated with aperiodicity, we expect the values for breathy, checked, and laryngealized vowels to be lower than those for modal vowels. As we see in Figure 9, the values for breathy, checked, and laryngealized vowels are all lower than the modal

vowels.

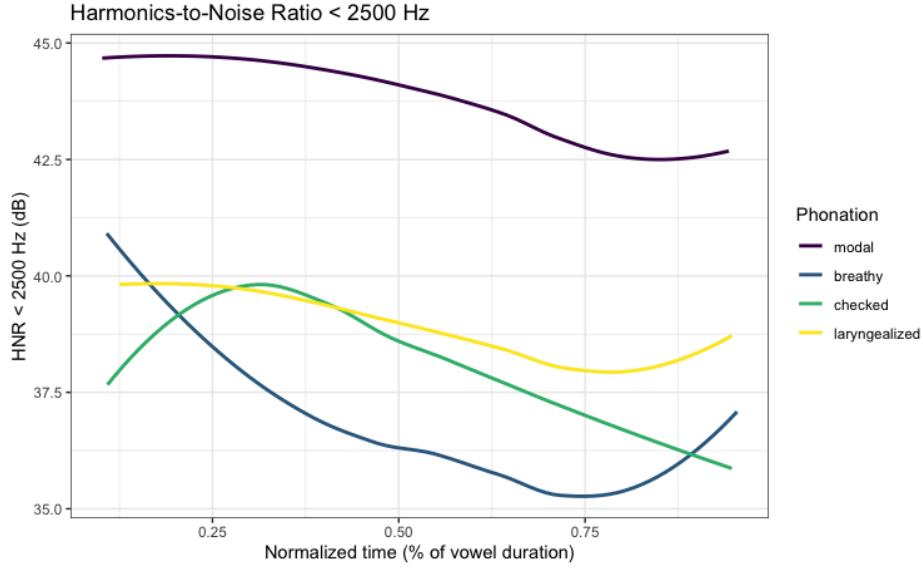


Figure 9: Line graph showing the change of HNR<2500 Hz across the vowels grouped by SLZ's phonation types.

When we turn to the mixed-effects linear regression model for HNR<2500 Hz, we observe that all three values will change negatively to the modal's values for every step of change in modal vowels. However, only the changes in breathy and checked vowels are significant; see Table 8.

Table 8: Results of the mixed-effects linear regression analysis for HNR < 2500 Hz.

| | Estimate | Std. Error | df | t value | p-value | |
|---------------|----------|------------|----------|---------|---------|----|
| Modal | 0.1347 | 0.0558 | 151.5648 | 2.414 | 0.01696 | * |
| Breathy | -0.4914 | 0.1530 | 9.9072 | -3.213 | 0.00939 | ** |
| Checked | -0.4669 | 0.1769 | 6.6601 | -2.640 | 0.03499 | * |
| Laryngealized | -0.1654 | 0.1810 | 17.8292 | -0.914 | 0.37293 | |

When we turn to the acoustic measure HNR<3500 Hz, which measures periodicity below 3500 Hz, we see that the three non-modal phonation types all show lower values than modal vowels; see Figure 10. Similar to HNR<2500 Hz, these lower HNR values indicate that these phonation types show aperiodicity in their production.

In contrast to the mixed-effect linear regression model for HNR<2500 Hz, the mixed-effects linear regression for HNR<3500 Hz shows that the differences we observe between the non-modal phonation types and the modal vowels are not statistically significant; see Table 9.

The last acoustic measure strongly associated with the first linear discriminant function was Strength-of-Excitation, a measure of the intensity of voicing regardless of the noise in the sig-

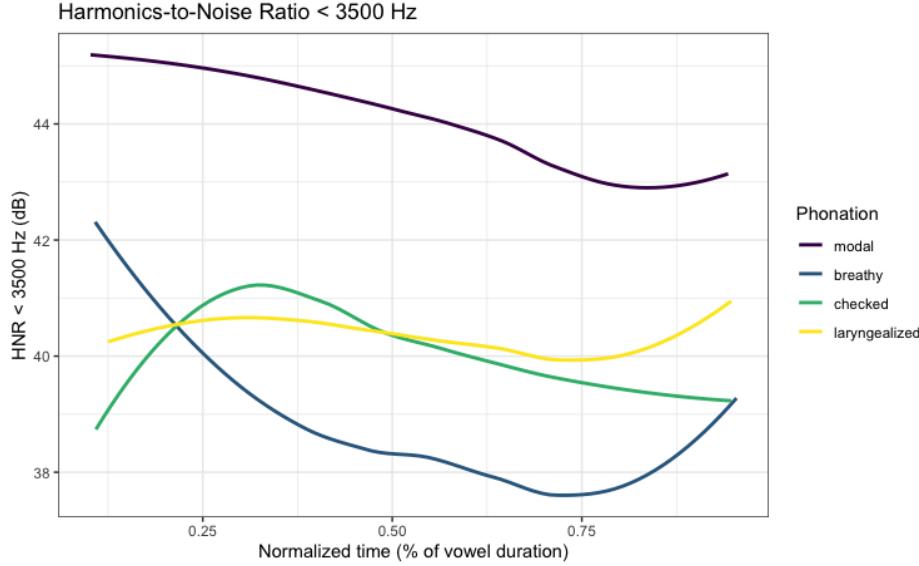


Figure 10: Line graph showing the change of $\text{HNR} < 3500 \text{ Hz}$ across the vowels grouped by SLZ's phonation types.

Table 9: Results of the mixed-effects linear regression analysis for $\text{HNR} < 3500 \text{ Hz}$.

| | Estimate | Std. Error | df | t value | p-value |
|---------------|----------|------------|-----------|---------|----------|
| Modal | 0.10018 | 0.05403 | 150.80549 | 1.854 | 0.0657 . |
| Breathy | -0.33967 | 0.15350 | 9.77806 | -2.213 | 0.0519 . |
| Checked | -0.37542 | 0.17313 | 6.62514 | -2.168 | 0.0690 . |
| Laryngealized | -0.09672 | 0.16999 | 20.16718 | -0.569 | 0.5757 |

nal. This SoE measure means that the closer the value is to zero, the less intense the voicing is and indicates a constriction closure to the glottis (Mittal, Yegnanarayana & Bhaskararao 2014, Garellek et al. 2021). When we consider the same type of graph showing the SoE values for each of the different phonation types across the vowel, we see that modal vowels have the highest SoE values across the length of the vowel, which is consistent with it lacking a glottal gesture; see Figure 11. However, what is interesting to see is that checked vowels begin at roughly the same height as modal vowels for SoE. Still, as the length of the vowel increases, the checked vowels progressively have lower and lower SoE values until becoming the lowest at the end of the vowel. Both breathy and laryngealized also exhibit a lower SoE value than modal phonation, which is consistent with having glottal gestures.

When evaluating this measure with a mixed-effects linear regression model, we observe that the changes in SoE across all non-modal phonation types are statistically significant; see Table 10.

The second linear discriminant function was similarly associated with $\text{HNR} < 2500 \text{ Hz}$ and SoE like they were for the first linear discriminant function. Instead of $\text{HNR} < 3500 \text{ Hz}$, a different

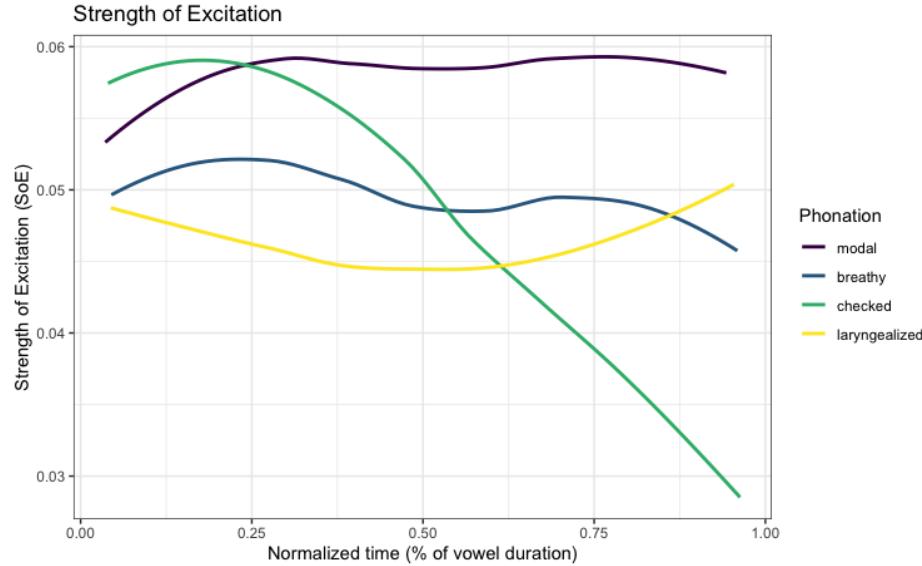


Figure 11: Line graph showing the change of Strength-of-Excitation across the vowels grouped by SLZ's phonation types.

Table 10: Results of the mixed-effects linear regression analysis for SoE.

| | Estimate | Std. Error | df | t value | p-value | |
|---------------|----------|------------|----------|---------|----------|-----|
| Modal | 0.65006 | 0.04328 | 5.33922 | 15.018 | 1.43e-05 | *** |
| Breathy | -0.11532 | 0.03057 | 7.41179 | -3.772 | 0.006265 | ** |
| Checked | -0.11516 | 0.02230 | 9.38943 | -5.165 | 0.000517 | *** |
| Laryngealized | -0.08573 | 0.03299 | 11.87923 | -2.599 | 0.023424 | * |

HNR measure strongly affected this second linear discriminant function. HNR<500 Hz measure periodicity below 500 Hz. Figure 12 shows that, like the other HNR measures, all three non-modal phonation types show a lower HNR value consistent with them being aperiodic. However, when the mixed-effects linear regression model was run on the acoustic measure, the only checked phonation reached significance; see Table 11.

Table 11: Results of the mixed-effects linear regression analysis for HNR < 500 Hz.

| | Estimate | Std. Error | df | t value | p-value | |
|---------------|----------|------------|---------|---------|---------|----|
| Modal | 0.14980 | 0.07553 | 9.64369 | 1.983 | 0.07652 | . |
| Breathy | -0.05906 | 0.20426 | 6.38671 | -0.289 | 0.78162 | |
| Checked | -0.70904 | 0.16811 | 6.77873 | -4.218 | 0.00424 | ** |
| Laryngealized | -0.29020 | 0.23947 | 7.58130 | -1.212 | 0.26197 | |

The third linear discriminant function only introduced one new acoustic measure. Δ HNR<2500

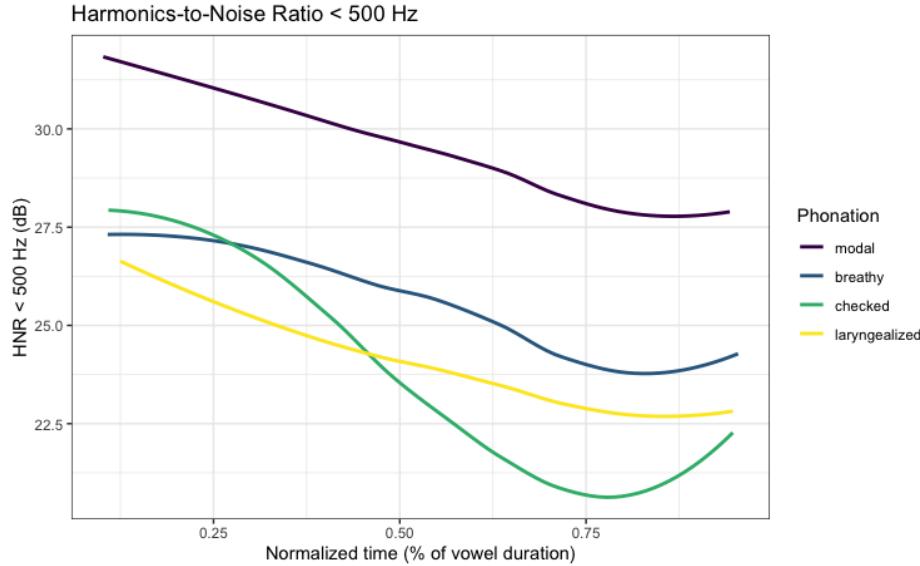


Figure 12: Line graph showing the change of HNR < 500 Hz across the vowels grouped by SLZ's phonation types.

Hz (end) measures the difference between the middle and end of the vowel for HNR<2500 Hz. As seen in Figure 9, the HNR measure changes most drastically from the middle of the vowel to the end of the vowel for checked phonation. The other phonation types also change but do not change as quickly. The difference can be seen in Figure 9 when we compare the values in the middle portion of the graph to those at the end for each phonation type. This magnitude of change is reflected in the linear regression for Δ HNR<2500 Hz (end), with checked phonation being the only one that showed actual statistical significance; see Table 12.

Table 12: Results of the mixed-effects linear regression analysis for Δ HNR<2500 Hz (end).

| | Estimate | Std. Error | df | t value | p-value |
|---------------|----------|------------|----------|---------|------------|
| Modal | -0.02397 | 0.06574 | 32.69794 | -0.365 | 0.71774 |
| Breathy | 0.12279 | 0.14446 | 11.41768 | 0.850 | 0.41280 |
| Checked | 0.32531 | 0.10589 | 41.57481 | 3.072 | 0.00374 ** |
| Laryngealized | -0.12616 | 0.27805 | 8.21579 | -0.454 | 0.66176 |

6 Discussion

6.1 Linear discriminant analysis discussion

The results of the linear discriminant analysis showed that the acoustic measures used in the PMoP are sufficient to account for the different phonation types in SLZ. The linear discriminant analysis performed well above chance at 70.45%. The linear discriminant analysis further showed that each of the four could be discriminated by the three discriminant functions it produced. However, as mentioned above in Section 5.1, the linear discriminant analysis performed better at classifying some phonation types over others. The model could classify modal vowels best at almost 90% accuracy; breathy vowels at 45%; checked at 37%; and laryngealized at 22%. The confusion matrix in Table 6 shows that all nonmodal phonation types were predominantly confused with modal vowels. This is not too unsurprising because of the highly variable nature in which these nonmodal phonation types are produced.

The linear discriminant analysis additionally showed that the acoustic measures that contributed the most to the three linear discriminant functions were harmonic-to-noise ratios along three different bandwidths and strength-of-excitation. Even though these acoustic measures had the greatest association, it does not mean that these are the only measures necessary to understand the phonation contrasts. Each linear discriminant function comprises a linear combination of all the acoustic measures and their associated coefficient. In more phonological terms, we can imagine these linear discriminant functions similar to the formula phonologists use to derive each candidate's harmonic score within Harmonic Grammar (Smolensky & Legendre 2006). In Harmonic Grammar, each constraint contributes weight to the formula, with some constraints contributing a greater weight and others a lower weight; a linear discriminant function assigns a numerical coefficient to each acoustic measure. Just because one acoustic measure has a lower coefficient does not mean an acoustic measure is not important to how speakers produce or perceive the phonation contrasts in Zapotec. A combination of several low-weighted acoustic measures may produce enough “weight” to overcome the “weight” of these heavier harmonic-to-noise ratios, similar to gang effects in Harmonic Grammar.

Instead of using the absolute value of the coefficients in the linear discriminant functions to determine which acoustic measures are to consider, another method is to take the mean predictive value of the three linear discriminant functions and determine the correlation values for each acoustic measure within each linear discriminant function. This is the method used by Garellek (2020) for determining which acoustic measures to use in his analysis for !Xóõ.

Both methods have their merits, but using the coefficients of the acoustic measures instead of their correlations is in line with more general practices used in the statistical literature. Different acoustic measures have the strongest correlation when using this other method with this

study's linear discriminant analysis. The three highest acoustic measures according to absolute correlation score for each linear discriminant function are reported in Table 13. Evaluating these measures will be left for further analysis.

Table 13: Acoustic measures with the highest absolute correlations for the three linear discriminant functions.

| LD1 | LD2 | LD3 |
|---------------|--------------------------|--|
| SoE | f_0 | ΔSoE (end) |
| f_0 | ΔSoE (end) | ΔF3 (end) |
| HNR < 1500 Hz | B1 | $\Delta\text{HNR} < 1500 \text{ Hz}$ (beg) |

6.2 Harmonic-to-noise ratio discussion

As previously mentioned, the acoustic measures that contributed the most “weight” to the three linear discriminant functions were predominantly harmonic-to-noise ratios across three different bandwidths. Out of the three harmonic-to-noise ratios that the linear discriminant functions called out, all were highly significant for at least one non-modal phonation, except for HNR<3500 Hz. This acoustic measure never reached significance; however, the p -value for the breathy and checked non-modal phonations hovered very close to the alpha value of 0.05. HNR<3500 Hz was not as informative because it subsumes the narrower bandwidths of <500 Hz and <2500 Hz with its wider bandwidth of <3500 Hz. It is also important to remember that under the Psychoacoustic Model of Phonation (PMoP; Kreiman et al. 2014), a single acoustic measure is not more important than another, but the acoustic measures associated with the harmonic spectral slopes, the inharmonic source noise, the time-varying source characteristics, and the vocal tract transfer functions combined together tell a complete story of how phonation is produced in a language.

When relying on the coefficient values, the “delta” score contributed little to the linear discriminant functions. This result is surprising, considering how the different phonation types behave dynamically, especially when distinguishing between laryngealized and checked vowels. One reason for this could be the same sample size reported in this paper of six speakers. It is very likely that the data from the remaining 12 speakers could improve the model and show the importance of these “delta” scores. However, when we consider the correlation of the mean predictive value of the linear discriminant function and the acoustic measures instead of the “weights”, as shown in Table 13 above, these “delta” scores make up the three highest correlations for the third linear discriminant function and the second highest correlation in the second linear discriminant function. This suggests that depending on evaluating by correlation of the linear discriminant functions and their component variables or evaluating the raw “weight” of the acoustic measures

could produce different outcomes. Accounting for why this is beyond the scope of this paper but will be left for further research.

6.3 Laryngealized vowel discussion

When we focus on how well the linear discriminant analysis performed at classifying the different acoustic measures, we notice that the model did very poorly for laryngealized vowels, which the linear discriminant analysis performed the worst on at 22% chance, which is below chance. As mentioned above in Section 3, laryngealized vowels show the greatest variability in their production inter- and intra-speaker. For some productions of this vowel, speakers would produce a vowel with model phonation throughout its entirety. The only indication that the vowel was non-model was a small period of two to three glottal pulses that had a lower amplitude than the rest of the vowel, very similar to what Gerfen & Baker (2005) found for laryngealized vowels in Mixtec. At other times, the production was with a tense voice rather than a canonical creaky voice.

Another aspect of laryngealized voice that the linear discriminant analysis revealed is that

6.4 Laryngeal Complexity Hypothesis

Another question has to do with the predictions from the Laryngeal Complexity Hypothesis. The LARYNGEAL COMPLEXITY HYPOTHESIS (LCH) has its origins in work by Silverman (1997) and Blankenship (1997, 2002). The basic premise of the LCH is that in languages that have both tone and phonation, there needs to be an ordering between the two laryngeal gestures for tone and phonation for optimal perceptibility. This premise comes from the understanding that the same mechanism responsible for tone is also responsible for the production of phonation: the vocal folds and glottis. The rate at which vocal folds vibrate is responsible for changing the fundamental frequency, perceived as pitch, which is grammaticalized as tone in lexical tone languages. The faster the vocal folds vibrate, the higher the pitch. And the slower the vocal folds vibrate, the lower the pitch.

The vocal folds are also the primary articulator for phonation. Ladefoged (1971) and Gordon & Ladefoged (2001) treated phonation as a by-product of how open or closed the glottis is during vocal fold vibration. This is schematized by Figure 13. The more open the glottis is, the breathier the sound to the point where the sound becomes the sound [h]. The more closed the glottis is, the creakier the sound is to the point where the sound becomes [?].

The LCH assumes phonation and tone are produced at the vocal folds and glottis. Because these same organs are responsible for these two phenomena, a mismatch exists in trying to have both simultaneously. It is assumed by the LCH that because of this issue, there needs to be

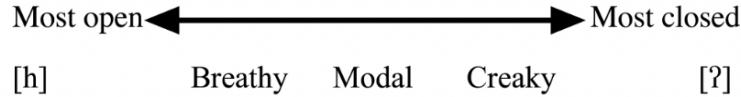


Figure 13: One-dimensional model for phonation. Based on Ladefoged (1971) and Gordon & Ladefoged (2001)

strict ordering in the glottal gestures. This means the tonal gesture must be produced before or after the phonation gesture. The reason for this is that if the gestures overlap, there will be a perturbation of the tone, and the listeners will not be able to differentiate the tone reliably. The LCH assumes that there is a close link between production and perception. This assumption places the responsibility on making sure the acoustic cues are the most perceptually salient on the speaker. The speaker is responsible for producing both tone and phonation so that the listener can differentiate the different cues for both tone and phonation. These assumptions can best be represented by Figure 14.

In Figure 14, which is taken from DiCanio (2012), the cue for tone is represented by the Pitch Target and the Glottal Gesture represents the gestures needed to produce phonation. When the Pitch Target is not co-articulated with the Glottal Gestures, there is the greatest perceptual recoverability for the listener. The LCH argues for this as the tones are the most recoverable on modal vowels. This modal portion is then ordered or phased relative to the non-modal portion. This ordering is the speaker's responsibility to accommodate the listeners' perceptibility. If, however, the Pitch Target and the Glottal Gesture were overlapping, as represented by the lower half in Figure 14, then the cues for pitch and phonation would be overlapping and making it perceptually more difficult for

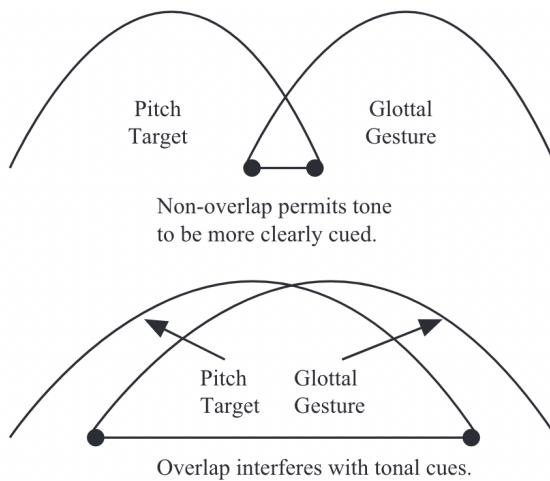


Figure 14: Representation taken from DiCanio (2012).

Some work has been done investigating this in other languages, most notably DiCanio's (2012) investigation into glottals in Itunyoso Trique, which is also an Oto-Manguean language. DiCanio in his study found that when the magnitude of coarticulation for glottal consonants occurs on the vowels, there is a strong correlation between the magnitude of overlap and the amount of perturbation in the f0 signal. If the degree of overlap was minor, then the acoustic signal had little to no perturbation. These results were found by consulting the spectral tilt of the vowels with the f0 measures and performing a generalized linear mixed effects model with the speaker treated as a random effect.

Another study on Jalapa Mazatec (Garellek & Keating 2011) also investigated the interaction of tone and phonation. Jalapa Mazatec is a language with both contrastive tone and phonation, and Garellek & Keating (2011) validated the claims made by the LCH, in that tone and phonation seemed to be ordered with each other when it comes to at least one of the phonation types.

This paper describes the tonal and phonation inventories in Santiago Laxopa Zapotec, an ideal language for testing the viability of the LCH because of its use of both contrastive tone and contrastive phonation.

Another question this study raises is the status of laryngealized vowels and their realizations. We observed from just two speakers that these vowels are highly variable. Assuming that Avelino's (2010) observations on this vowel in the closely related Yalálag Zapotec hold in SLZ, there is considerably more variation that exists within and between speakers. Suppose all of these varying realizations of a phonological category are evident. What does this mean for this category's status and the cues necessary for speakers to produce and realize this category?

7 Further directions and Conclusions

In conclusion, this paper has briefly introduced the tonal and phonation systems of Santiago Laxopa Zapotec, an understudied variety of Sierra Norte Zapotec. This system is essential for exploring the acoustic cues used in different languages to differentiate each phonation type.

Contrary to acoustic work on other Zapotec varieties by Esposito (2010b), which showed that female speakers' phonation contrasts were best characterized by H1-H2 and male speakers' contrasts by H1-A3, SLZ suggests that both male and female speakers' breathy voice is best described by H1-A3 and checked and laryngealized voice are characterized by a difference in the timing of spectral-tilt measurements and CPP values depending on the speaker.

Because of these phonation type differences in behavior, one could, in theory, speak to how SLZ compensates for using the larynx for both tone and phonation. This paper has not explicitly spoken on this issue, and the Laryngeal Complexity Hypothesis presented by Silverman (1997) and Blankenship (1997, 2002). The LCH states that when a language has tone and phonation,

these different aspects are ordered or phased concerning one another. This phasing allows for the most significant perceptibility in the acoustic signal. This allows the listener to interpret the acoustic signals for tone and phonation adequately.

Overall, there seems to be some information from the spectral-tilt and CPP analysis that speaks to the question of the LCH. If the reader recalls, CPP was the most important cue for differentiating checked and laryngealized vowels from modal and breathy vowels for FSR. The key is the lower CPP value, which corresponds to greater aperiodicity. It is worth repeating that we did see a difference in the timing of the aperiodicity between checked and laryngealized vowels. In laryngealized vowels, this dip in CPP occurs in the middle third of the vowel, and for checked vowels, this dip occurs in the final third. However, the fact that for breathy vowels, the most noticeable cue is the higher H1-A3 which happens throughout the entire vowel, might suggest that breathiness does not behave in the same way as the other phonation types. The behavior from checked and laryngealized vowels seems to support the LCH. Unfortunately, no firm conclusions can be made about the LCH, and further investigation is required to determine the validity of the LCH in SLZ.

This ability of speakers to interpret the acoustic signals in conjunction with the acoustic signals for tone is of great interest. It would benefit from a perception experiment to determine what the speakers are using to differentiate the different phonation types. This is especially true for laryngealized vowels, which have such varied pronunciations.

This study will benefit from further analysis and data collection. Now that the world is safer regarding COVID-19, collecting data from more speakers is essential to corroborate the data and analysis from FSR and RD.

References

- Adler, Jeff, Steven Foley, Jed Pizarro-Guevara, Kelsey Sasaki & Maziar Toosarvandani. 2018. The derivation of verb initiality in Santiago Laxopa Zapotec. In Jason Merchant, Line Mikkelsen, Deniz Rudin & Kelsey Sasaki (eds.), *A reasonable way to proceed: Essays in honor of Jim McCloskey*, 31–49. Santa Cruz, Berkeley, Chicago: University of California.
- Adler, Jeffrey M. & Maho Morimoto. 2016. Acoustics of phonation types and tones in Santiago Laxopa Zapotec. *The Journal of the Acoustical Society of America* 140(4). 3109–3109. <https://doi.org/10.1121/1.4969713>.
- Andruski, Jean E. 2006. Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics* 34(3). 388–404. <https://doi.org/10.1016/j.wocn.2005.07.001>.

- Andruski, Jean E. & Martha Ratliff. 2000. Phonation types in production of phonological tone: the case of Green Mong. *Journal of the International Phonetic Association* 30(1-2). 37–61. <https://doi.org/10.1017/S0025100300006654>.
- Avelino, Heriberto. 2004. *Topics in Yalálag Zapotec, with particular reference to its phonetic structures*. Los Angeles, CA: University of California, Los Angeles dissertation. 315 pp.
- Avelino, Heriberto. 2010. Acoustic and Electroglossographic Analyses of Nonpathological, Non-modal Phonation. *Journal of Voice* 24(3). 270–280. <https://doi.org/10.1016/j.jvoice.2008.10.002>.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1). 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Blankenship, Barbara. 1997. *The time course of breathiness and laryngealization in vowels*. Los Angeles, CA: University of California, Los Angeles dissertation.
- Blankenship, Barbara. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30(2). 163–191. <https://doi.org/10.1006/jpho.2001.0155>.
- Boersma, Paul & David Weenink. 2021. *Praat: Doing Phonetics by Computer [Computer Program]*. <http://www.praat.org/> (3 December, 2021).
- Brunelle, Marc. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37(1). 79–96. <https://doi.org/10.1016/j.wocn.2008.09.003>.
- Brunelle, Marc & James Kirby. 2016. Tone and Phonation in Southeast Asian Languages. *Language and Linguistics Compass* 10(4). 191–207. <https://doi.org/10.1111/lnc3.12182>.
- Campbell, Eric W. 2011. Zenzontepet Chatino aspect morphology and Zapotecan verb classes. *International Journal of American Linguistics* 77(2). 219–246.
- Campbell, Eric W. 2017a. Otomanguean historical linguistics: Exploring the subgroups. *Language and Linguistics Compass* 11(7). e12244. <https://doi.org/10.1111/lnc3.12244>.
- Campbell, Eric W. 2017b. Otomanguean historical linguistics: Past, present, and prospects for the future. *Language and Linguistics Compass* 11(4). e12240. <https://doi.org/10.1111/lnc3.12240>.
- Campbell, Lyle, Terrence Kaufman & Thomas Smith-Stark. 1986. Meso-America as a linguistic area. *Language* 62(3). 530–570.
- Chávez-Péón, Mario E. 2010. *The interaction of metrical structure, tone, and phonation types in Quiavini Zapotec*. University of British Columbia dissertation. <https://doi.org/10.14288/1.0071253>.
- DiCanio, Christian T. 2012. Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics* 40(1). 162–176. <https://doi.org/10.1016/j.wocn.2011.10.006>.

- Enfield, Nick J. 2005. Areal linguistics and mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.
- Esling, John H., Scott R. Moisik, Allison Benner & Lise Crevier-Buchman. 2019. *Voice Quality: The Laryngeal Articulator Model*. 1st edn. (Cambridge Studies in Linguistics 162). Cambridge University Press. <https://doi.org/10.1017/9781108696555>.
- Esposito, Christina M. 2010a. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 38(2). 306–316. <https://doi.org/10.1016/j.wocn.2010.02.002>.
- Esposito, Christina M. 2010b. Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association* 40(2). 181–198. <https://doi.org/10.1017/S0025100310000046>.
- Fischer-Jørgensen, Eli. 1968. Phonetic Analysis of Breathy (Murmured) Vowels in Gujarati. *Annual Report of the Institute of Phonetics University of Copenhagen* 2. 35–85. <https://doi.org/10.7146/aripuc.v2i.130674>.
- Fisher, R. A. 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* 7(2). 179–188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>.
- Foley, Steven, Nick Kalivoda & Maziar Toosarvandani. 2018. Forbidden clitic clusters in Zapotec. In Daniel Edmiston, Marina Ermolaeva, Emre Hakgüder, Jackie Lai, Kathryn Montemurro, Brandon Rhodes, Amara Sankhagowit & Michael Tabatowski (eds.), *Proceedings of the Fifty-third Annual Meeting of the Chicago Linguistic Society*, 87–102.
- Foley, Steven & Maziar Toosarvandani. 2020. Extending the Person-Case Constraint to Gender: Agreement, Locality, and the Syntax of Pronouns. *Linguistic Inquiry*. 1–40. https://doi.org/10.1162/ling_a_00395.
- Garellek, Marc. 2019. The phonetics of voice. In William F. Katz & Peter F. Assmann (eds.), *The Routledge handbook of phonetics* (Routledge Handbooks in Linguistics), 75–106. Abingdon, Oxon ; New York, NY: Routledge.
- Garellek, Marc. 2020. Acoustic Discriminability of the Complex Phonation System in !Xóõ. *Phonetica* 77(2). 131–160. <https://doi.org/10.1159/000494301>.
- Garellek, Marc, Yuan Chai, Yaqian Huang & Maxine Van Doren. 2021. Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association*. 1–28. <https://doi.org/10.1017/S0025100321000116>.
- Garellek, Marc & Christina M. Esposito. 2021. Phonetics of White Hmong vowel and tonal contrasts. *Journal of the International Phonetic Association*. 1–20. <https://doi.org/10.1017/S0025100321000104>.

- Garellek, Marc & Patricia Keating. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41(2). 185–205. <https://doi.org/10.1017/S0025100311000193>.
- Gerfen, Chip & Kirk Baker. 2005. The production and perception of laryngealized vowels in Coatzozapan Mixtec. *Journal of Phonetics* 33(3). 311–334. <https://doi.org/10.1016/j.wocn.2004.11.002>.
- Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29(4). 383–406. <https://doi.org/10.1006/jpho.2001.0147>.
- Hillenbrand, James, Ronald A. Cleveland & Robert L. Erickson. 1994. Acoustic Correlates of Breathy Vocal Quality. *Journal of Speech, Language, and Hearing Research* 37(4). 769–778. <https://doi.org/10.1044/jshr.3704.769>.
- Hillenbrand, James & Robert A. Houde. 1996. Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech. *Journal of Speech, Language, and Hearing Research* 39(2). 311–321. <https://doi.org/10.1044/jshr.3902.311>.
- Hockett, Charles F. 1947. Peiping Phonology. *Journal of the American Oriental Society* 67(4). 253–267. <https://doi.org/10.2307/596062>.
- Huffman, Marie K. 1987. Measures of phonation type in Hmong. *The Journal of the Acoustical Society of America* 81(2). 495–504. <https://doi.org/10.1121/1.394915>.
- Kawahara, Hideki, Alain De Cheveigne & Roy D. Patterson. 1998. An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite. In *5th International Conference on Spoken Language Processing (ICSLP 1998)*, paper 0659–. ISCA. <https://doi.org/10.21437/ICSLP.1998-555>.
- Klatt, Dennis H. & Laura C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 87(2). 820–857. <https://doi.org/10.1121/1.398894>.
- Kreiman, Jody, Bruce R. Gerratt, Marc Garellek, Robin Samlan & Zhaoyan Zhang. 2014. Toward a unified theory of voice production and perception. *Loquens* 1(1). e009. <https://doi.org/10.3989/loquens.2014.009>.
- de Krom, Guus. 1993. A Cepstrum-Based Technique for Determining a Harmonics-to-Noise Ratio in Speech Signals. *Journal of Speech, Language, and Hearing Research* 36(2). 254–266. <https://doi.org/10.1044/jshr.3602.254>.
- Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago: University of Chicago.
- Long, Rebecca & Sofronio Cruz. 2005. *Diccionario zapoteco de San Bartolomé Zoogocho, Oaxaca*. 2nd edition (Serie de vocabularios y diccionarios indígenas "Mariano Silva y Aceves" no. 38). Coyoacán, D.F. [Mexico]: Instituto Lingüístico de Verano. 531 pp.

- López Nicolás, Oscar. 2016. *Estudios de la fonología y gramática del zapoteco de Zóochiña*. Mexico City, Mexico: CIESAS Doctoral thesis.
- Masica, Colin P. 1976. *Defining a linguistic area: South Asia*. Chicago: The University of Chicago Press.
- Michaud, Alexis. 2012. The Complex Tones of East/Southeast Asian Languages: Current Challenges for Typology and Modelling. 8.
- Mittal, Vinay Kumar, B. Yegnanarayana & Peri Bhaskararao. 2014. Study of the effects of vocal tract constriction on glottal vibration. *The Journal of the Acoustical Society of America* 136(4). 1932–1941. <https://doi.org/10.1121/1.4894789>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. manual. R Foundation for Statistical Computing. Vienna, Austria.
- Ratliff, Martha Susan. 1992. *Meaningful tone: a study of tonal morphology in compounds, form classes and expressive phrases in White Hmong* (Special Report no. 27, 1992). DeKalb, Ill: Northern Illinois University, Center for Southeast Asian Studies. 269 pp.
- Shue, Yen-Liang, Patricia Keating & Chad Vicenik. 2009. VOICESAUCE: A program for voice analysis. *The Journal of the Acoustical Society of America* 126(4). 2221. <https://doi.org/10.1121/1.3248865>.
- Silverman, Daniel. 1997. Laryngeal complexity in Otomanguean vowels. *Phonology* 14(2). 235–261. <https://doi.org/10.1017/S0952675797003412>.
- Silverman, Daniel, Barbara Blankenship, Paul Kirk & Peter Ladefoged. 1995. Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics* 37(1).
- Sjölander, Kåre. 2004. *Snack Sound Toolkit*. Stockholm, Sweden: KTH.
- Smalley, William A. 1976. The problems of consonants and tone: Hmong (Meo, Miao). In W. A. Smalley (ed.), *Phonemes and orthography: Language planning in ten minority languages of Thailand*, vol. 43 (Pacific Linguistics, Series C), 85–123. Canberra: Australian National University.
- Smolensky, Paul & Géraldine Legendre. 2006. *The harmonic mind: from neural computation to optimality-theoretic grammar*. Cambridge, Mass: MIT Press. 2 pp.
- Sonnenschein, Aaron H. 2005. *A descriptive grammar of San Bartolomé Zoogocho Zapotec*. Munich: Lincom Europa.
- Suárez, Jorge. 1983. *The Mesoamerican Indian languages*. Cambridge: Cambridge University Press.
- Thurgood, Graham. 2002. Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica* 19(2). 333–363. <https://doi.org/10.1075/dia.19.2.04thu>.
- Uchihara, Hiroto. 2016. Tone and registrogenesis in Quiaviní Zapotec. *Diachronica* 33(2). 220–254. <https://doi.org/10.1075/dia.33.2.03uch>.
- Venables, W. N., Brian D. Ripley & W. N. Venables. 2002. *Modern applied statistics with S*. 4th ed (Statistics and Computing). New York: Springer. 495 pp.

- Villard, Stéphanie. 2015. *The Phonology and Morphology of Zacatepec Eastern Chatino*. Austin, TX: University of Texas at Austin dissertation.
- Wayland, Ratree & Allard Jongman. 2003. Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics* 31(2). 181–201. [https://doi.org/10.1016/S0095-4470\(02\)00086-4](https://doi.org/10.1016/S0095-4470(02)00086-4).
- Yip, Moira Jean Winsland. 2002. *Tone* (Cambridge Textbooks in Linguistics). Cambridge ; New York: Cambridge University Press. 341 pp.