Voice Quality and Laryngeal Complexity in Santiago Laxopa Zapotec

Mykel Loren Brinkerhoff

University of California, Santa Cruz

6 June 2025

Outline

1. Introduction
   - Overview of my dissertation
   - What is Voice Quality?
   - Santiago Laxopa Zapotec
2. Previous research
   - Measuring Voice Quality
   - Modeling Voice Quality
   - Laryngeal Complexity
3. My Results
   - Data and Methods
   - Acoustic Landscape
   - Random Forests show important measures
   - Laryngeal Complexity in SLZ
4. Summary and conclusions

Research Overview

Overview:

This presentation explores what characterizes voice quality in a Santiago Laxopa Zapotec (SLZ), how is it is structured acoustically, and how this structure helps explain SLZ's laryngeal complexity.

Research Overview

Research Questions:

- How is phonation's acoustic space structured in a single language?
- Which measures are important for capturing phonation contrasts?
- How do these measures help explain SLZ's laryngeal complexity?

Research Overview

Answers to Research Questions:

- The acoustic space is three-dimensional and the dimensions are comparable to Keating et al.'s (2023) findings.
  - Dimensions are correlated with glottal-airflow (D1/D3) and nonmodal-to-modal (D2) continua.
- Only a handful of measures are important for capturing phonation contrasts.
- Measures show that SLZ's laryngeal complexity is weakly articulated and shows phasing.

What is Voice Quality?

- Broad sense = long-term characteristics of an individual's voice (Abercrombie 1967, Laver 1980).
- Narrow sense = how larynx affects phonetic characteristics of speech sounds (e.g., Esling et al. 2019).
- Sometimes used interchangeably with phonation
  - Barzilai & Riestenberg 2021 use *voice quality* for phonetic description and *phonation* for phonological contrasts.

How is voice quality used?

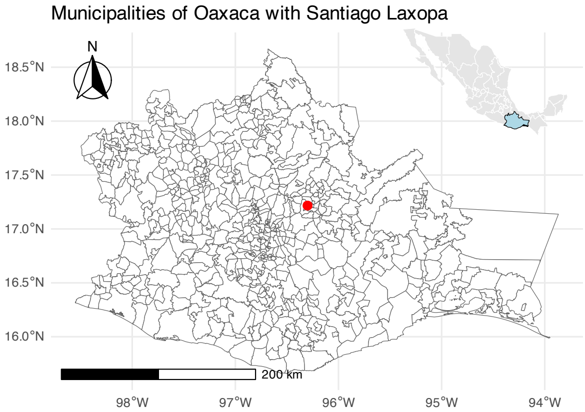
- Paralinguistic information by "indexing the biological, psychological, and social characteristics of the speaker" (e.g., Laver 1968, Podesva 2016)
- Phonological contrasts (e.g., Esposito & Khan 2020).

Why study voice quality?

- Voice quality plays a crucial role in speech communication.
- It is used in many languages to create phonological contrasts.
- It is a rich area of research that can help us understand speech production and perception.

What is Santiago Laxopa Zapotec

- Santiago Laxopa Zapotec (SLZ; *Dille'xhunh Laxup*) is a Sierra Norte variety of Zapotec (Oto-Manguean).
- Spoken by c. 1,000 speakers in Santiago Laxopa and in diaspora.



Why study SLZ?

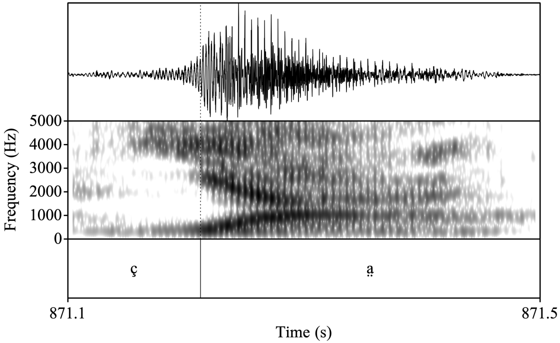
- SLZ is a relatively understudied language, so it provides an opportunity to contribute to our linguistic understaning.
- SLZ has a complex laryngeal system that is not well understood.
  - SLZ has a four-way phonation contrast.
  - SLZ has a five-way tonal contrast.
  - These contrasts are orthogonal, meaning that they can occur independently of each other.

Phonation in SLZ

- SLZ has a four-way phonation contrast:
  - Modal ([a])
  - Breathy ([a̤])
  - Checked ([a͡ʔ] or [a͡a̰])
  - Rearticulated ([a͡ʔa], [a͡a̰a], or [a̰])

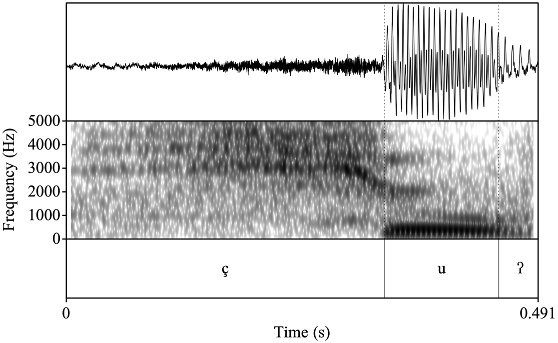
Breathy Phonation in SLZ

- Characterized by aperiodicity and noise in the harmonic structure.
- Longer duration than modal voice
- Lower fundamental frequency ($f0$)



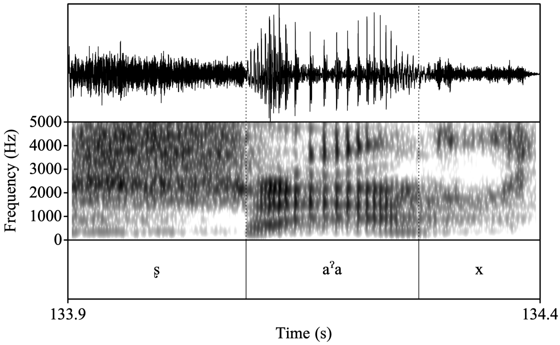
Checked Phonation in SLZ

- Glottal occlusion at the end of vowel or creaky voice in second half vowel.
- Decrease in voicing amplitude during second half.



Rearticulated Phonation in SLZ

- Glottal occlusion or creaky voice in the middle of the vowel.
- Longer duration than modal voice
- Decrease in voicing amplitude during first half.



Tone in SLZ

- SLZ shows a five-way tonal contrast on monosyllabic nouns (Brinkerhoff, Duff & Wax Cavallaro 2022).

High	*xha*	[z̧a˦]	‘clothing.poss’
Mid	*lhill*	[ɾiʒ˧]	‘house.poss’
Low	*yu’*	[ç̦uˀ˨]	‘earth’
Rising	*yu’u*	[ç̦uˀu˧˦]	‘lime (Sp. *cal*)’
Falling	*yu’u*	[ç̦uˀu˦˨]	‘house’

Interaction of Tone and Phonation in SLZ

- Tone and phonation are orthogonal in SLZ.

	Modal	**Breathy**	**Checked**	**Rearticulated**
High	✓	–	✓	✓
Mid	✓	✓	✓	✓
Low	✓	✓	✓	✓
Rising	✓	✓	–	✓
Falling	✓	✓	✓	✓

Measuring voice quality

- Long been established that phonation has correlates in the acoustic signal (e.g., Fischer-Jørgensen 1968, Klatt & Klatt 1990).
- Fischer-Jørgensen (1968) found that a strengthened fundamental (i.e., H1) correlated with breathy vowels in Gujarati.
  - Proposed H1—H2 amplitude as a measure to normalize for overall intensity differences
- Subsequent research has shown that H1—H2 is useful for classifying voice quality contrasts.
- However, it is not without its problems (see Chai & Garellek 2022 for an overview).

Measuring voice quality

- H2 and other normalizations (A1, A3) are really attempts to understand the relative strength of the fundamental (H1) to higher harmonic energy
- H1 and H2 are affected by subglottal pressure differently (Sundberg 2022)
  - This undermines the original reasoning behind H1—H2
- Researchers sometimes find that H1—H2 does not distinguish phonation types (e.g., Esposito 2010, 2012, Garellek & Esposito 2023).
- Error in measuring H1—H2 is uncomfortably high (e.g., Chai & Garellek 2022).

Measuring voice quality

- Many measures have been proposed to capture voice quality.
- Measures can be broadly categorized into several categories (Gordon & Ladefoged 2001):
  - Periodicity (e.g., HNR or CPP)
  - Energy
  - Spectral tilt (e.g., H1*−H2*, residual H1*)
  - Pitch (e.g., $f0$)
  - Duration
- Linguists have used combinations of these measures to model phonation (e.g., Blankenship 2002, Brunelle & Kirby 2016, Esposito 2012).

New measures of voice quality

- New measures are being developed all the time.
- Chai & Garellek's (2022) residual H1* is a new measure of spectral tilt that is more robust than traditional spectral tilt measures.
  - Does this by removing the the effects of Energy from the H1* measure.
  - This was the original goal of Fischer-Jørgensen (1968)
- Brinkerhoff & McGuire (2025) found that residual H1* is a better measure of spectral tilt than H1*−H2* for Santiago Laxopa Zapotec's four-way phonation contrast.

Too many measures

- There are many measures of voice quality, but it is not clear which ones are the most useful.
- It also makes it difficult to compare results across studies.
- VoiceSauce makes this a common issue (Shue et al. 2011).

Too many measures

- Kreiman et al. (2014, 2021) tackle this by proposing a psychoacoustic model of voice quality.
- Found certain measures are perceptually more important than others.

Model Component	**Parameters**
Harmonic source spectral shape	H1–H2 H2–H4 H4–2 kHz 2 kHz–5 kHz
Inharmonic source excitation	Spectrally-shaped noise-to-harmonics ratio
Time-varying source characteristics	$f0$ mean and standard deviation (or $f0$ track) Amplitude mean and standard deviation (or amplitude track)
Vocal tract transfer function	Formant frequencies/bandwidths Spectral zeroes/bandwidths

Modeling voice quality

- Early models proposed that voice quality is one dimensional and represents glottal airflow (Ladefoged 1971, Ladefoged & Maddieson 1996).

Open Glottis ⟷ Closed Glottis

[h] Breathy Modal Creaky [ʔ]

UC SANTA CRUZ

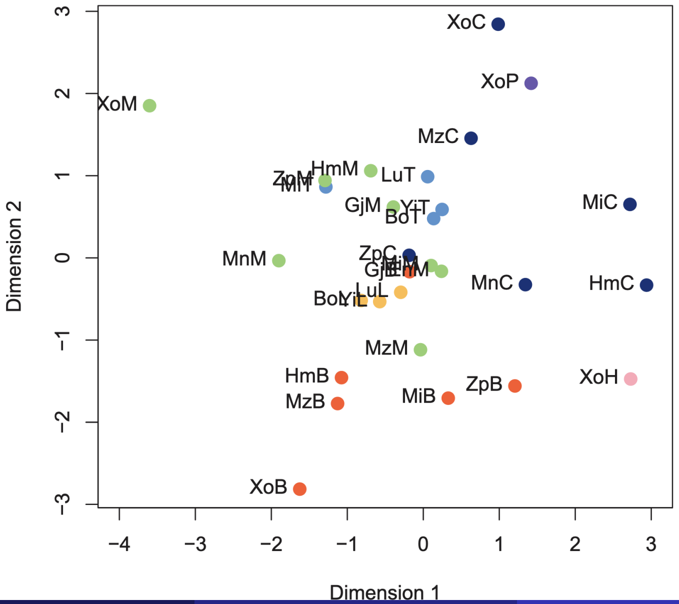
Voice quality's multidimensionality

- More recent work has shown that voice quality is not one-dimensional, but minimally five-dimensional (e.g., Garellek et al. 2016, Kreiman et al. 2021).
    - Especially in the case of individual speaker differences.
- Garellek et al. (2013) has argued that dimensionality might not be as complex for capturing phonation contrasts.

Keating et al. (2023)

- Explored phonation's cross-linguistic acoustic space.
- Found a two-dimensional space for phonation across 11 languages.
    1. First dimension = nonmodal-to-modal continuum.
    2. Second dimension = glottal-airflow continuum.
- Found languages with more contrasts used more of the acoustic space than languages with fewer contrasts.
- Found correlations between dimensions and acoustic measures.
    1. First dimension = periodicity and energy measures.
    2. Second dimension = spectral tilt and periodicity measures.

Keating et al. (2023)



What is Laryngeal Complexity

- *Laryngeal complexity* describes languages that use both contrastive tone and phonation (Silverman 1997a,b).
- Very common in the Oto-Manguean language family, but is also found in other languages.

Phonation's phasing

-

Implicational hierarchy of patterns

- Silverman (1997a) proposed a hierarchy of laryngeal complexity phasing pattern.

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

Previous research on laryngeal complexity

-

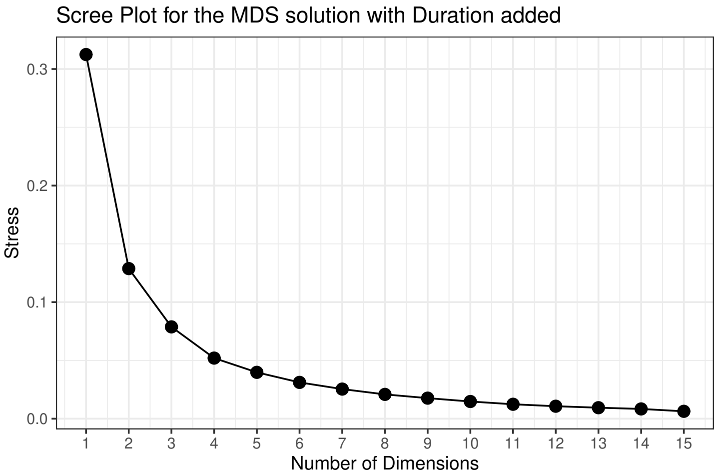
Data

- Data comes from fieldwork on Santiago Laxopa Zapotec (SLZ) from Summer 2023.
- Production data was collected from 10 speakers (5 male/5 female).

MDS analysis

- Multidimensional scaling (MDS; Kruskal & Wish 1978) was used to reduce the dimensionality of the data.

Number of Dimensions needed



Explanation of the plots

- Each point = unique speaker x phonation combination.
- Colors = phonation type.
  - Modal = black
  - Breathy = orange
  - Checked = blue
  - Rearticulated = green
- Axes = dimensions of the acoustic space.

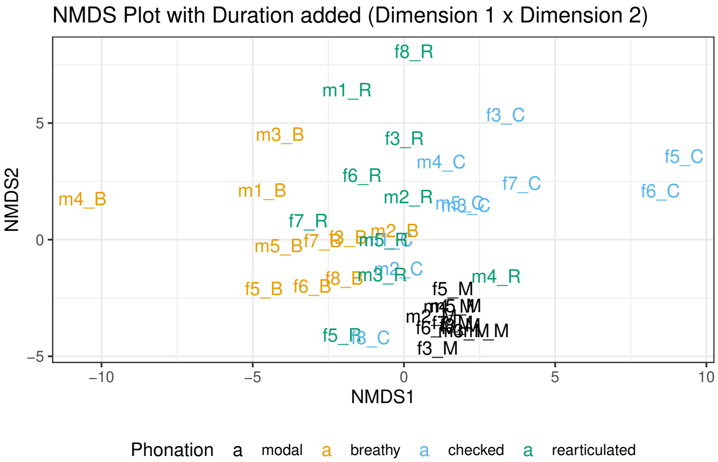
Dimensionality in SLZ

- Scan the QR code to see the three-dimensional space.

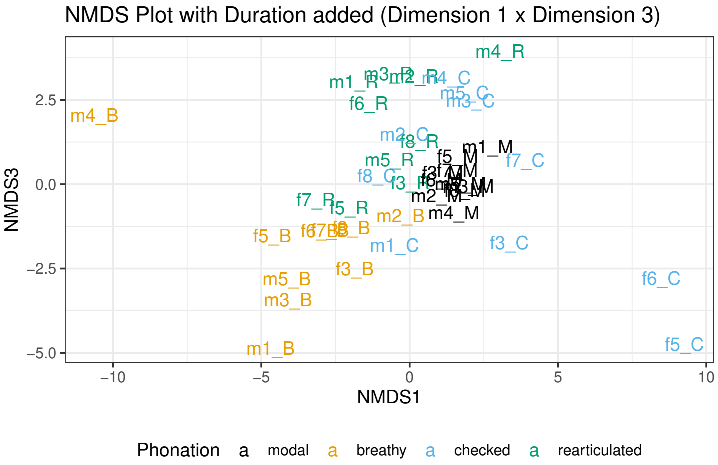
Takeaways from 3D plot

- Shows clear groupings of phonation types
-

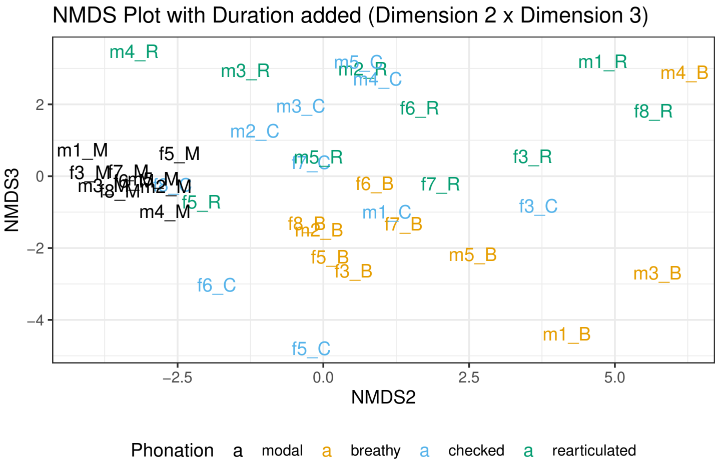
Dimensionality in SLZ



Dimensionality in SLZ



Dimensionality in SLZ



Summary of Dimensions

- Dimension 1 (D1) gives a rough continuum from breathy to creaky.
- Dimension 2 (D2) gives a rough continuum from modal to nonmodal.
- Dimension 3 (D3) gives a rough continuum from breathy to creaky.

Correlation to Acoustic Measures

- D1 correlated with spectral tilt measures:
  - H1*–A1* ($r^2 = -0.83$)
  - H1*–A2* ($r^2 = -0.86$)
  - H1*–A3* ($r^2 = -0.81$)
- D2 correlated with periodicity and energy:
  - HNR<500 Hz ($r^2 = -0.79$)
  - HNR<1500 Hz ($r^2 = -0.80$)
  - Energy ($r^2 = -0.79$)
- D3 correlated with spectral tilt:
  - residual H1* ($r^2 = -0.72$)
  - H2*–H4* ($r^2 = -0.69$)
  - H2* ($r^2 = -0.68$)

Summary of Acoustic Landscape

- SLZ's phonation occupies a three-dimensional space.
- Dimensions are correlated with glottal-airflow continuum (D1/D3) and nonmodal-to-modal continuum (D2).
- Dimensions are similar to those found in Keating et al. (2023).

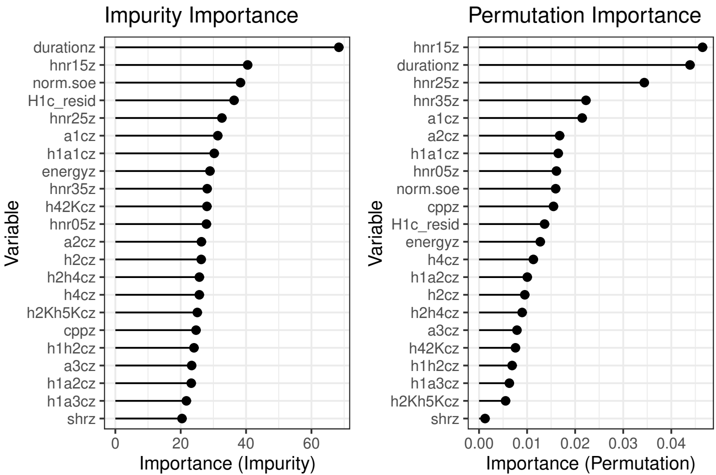
What is a Random Forest?

- Random Forest is a subclass of decision trees that uses an ensemble of decision trees to make predictions (Breiman et al. 1986, Breiman 1996, 2001).
- More robust than

Interpreting Random Forests

-

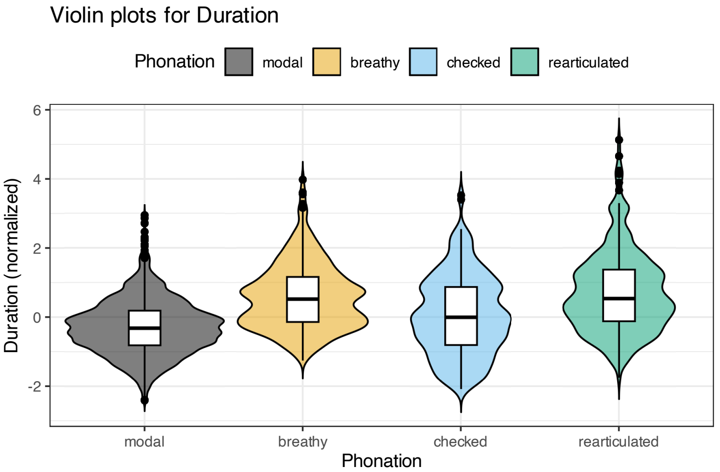
Variable Importance



Variable Importance

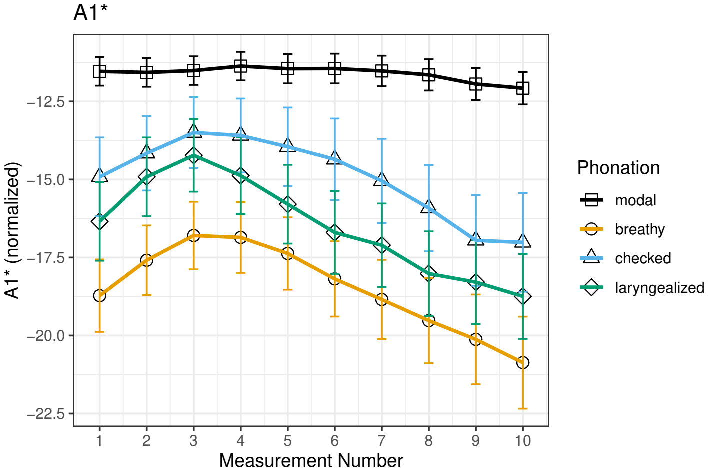
- Variable importance shows that only a handful of measures are important for capturing phonation contrasts.
    1. Duration
    2. A1*
    3. H1*−A1*
    4. Residual H1*
    5. HNR $<$ 1500 Hz
    6. Strength of Excitation (SoE)

Duration



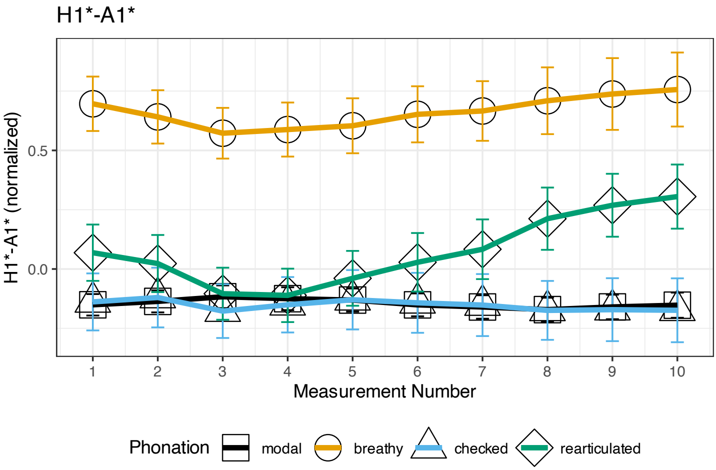
Violin plots for Duration
Phonation
modal
breathy
checked
rearticulated
Duration (normalized)
6
4
2
0
-2
modal
breathy
checked
rearticulated
Phonation

A1*



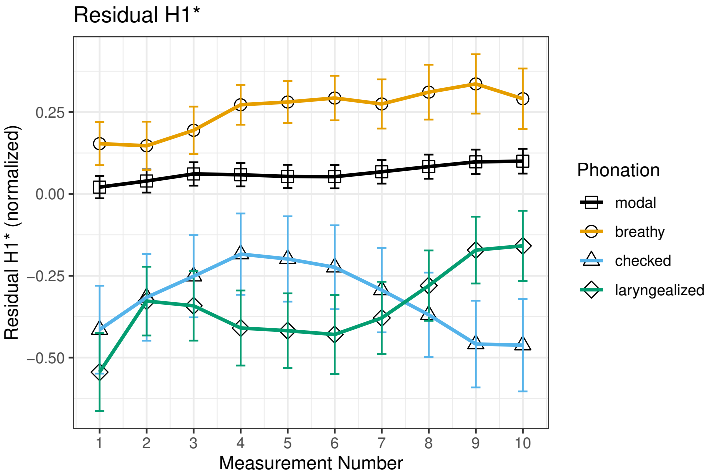
A1*
A1* (normalized)
−12.5
−15.0
−17.5
−20.0
−22.5
1
2
3
4
5
6
7
8
9
10
Measurement Number
Phonation
modal
breathy
checked
laryngealized

H1*—A1*



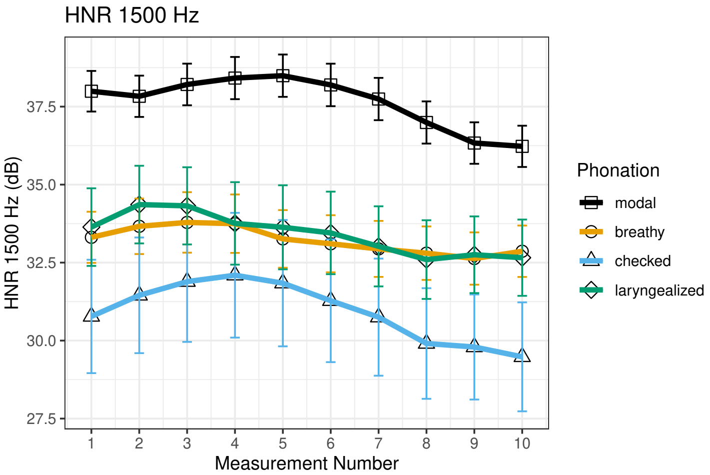
H1*-A1*
H1*-A1* (normalized)
0.5
0.0
1
2
3
4
5
6
7
8
9
10
Measurement Number
Phonation
modal
breathy
checked
rearticulated

Residual H1*

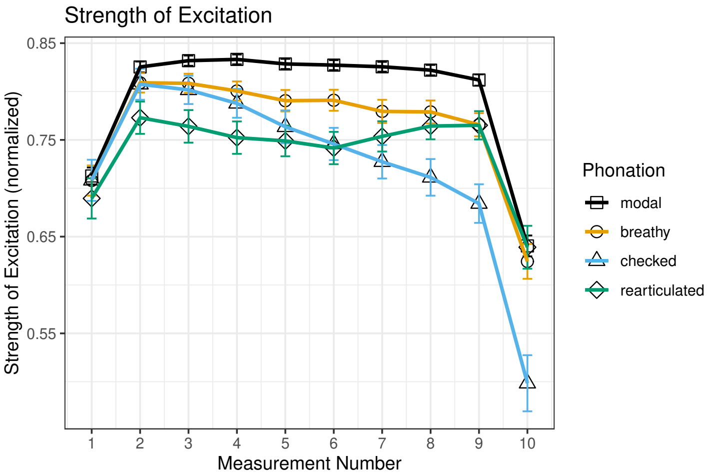


Residual H1*
Residual H1* (normalized)
0.25
0.00
−0.25
−0.50
1
2
3
4
5
6
7
8
9
10
Measurement Number
Phonation
modal
breathy
checked
laryngealized

HNR < 1500 Hz


HNR 1500 Hz
HNR 1500 Hz (dB)
37.5
35.0
32.5
30.0
27.5
1
2
3
4
5
6
7
8
9
10
Measurement Number
Phonation
modal
breathy
checked
laryngealized

Strength of Excitation (SoE)



Summary of Random Forest Results

- Only a handful of measures are important for capturing phonation contrasts.
- Duration, A1*, H1*−A1*, residual H1*, HNR $<$ 1500 Hz, and SoE are the most important measures.
- There is a lot of overlap between MDS' correlated measures and the important measures from the Random Forests.

Quanitfying Laryngeal Complexity

- Laryngeal complexity is the number of phonation contrasts in a language.
- More phonation contrasts means more complex laryngeal system.
- Laryngeal complexity can be measured by the number of phonation contrasts in a language.
- Laryngeal complexity can be measured by the dimensionality of the acoustic space.

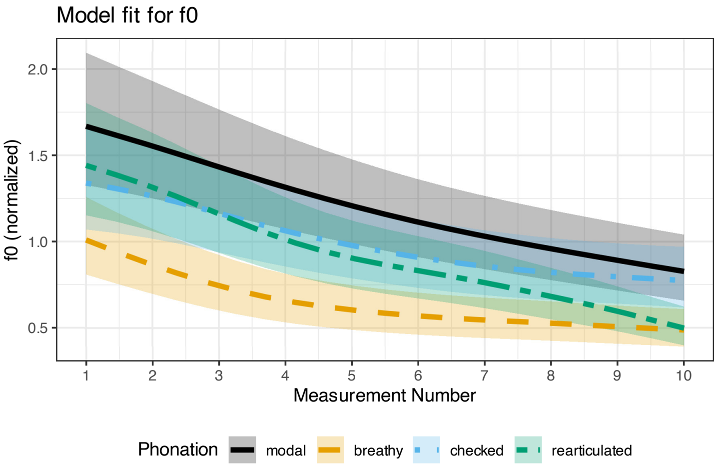
What are Generalized Additive Mixed Models (GAMMs)

- GAMMs are a type of statistical model that can be used to analyze non-linear relationships between variables.

Evaluating GAMMs

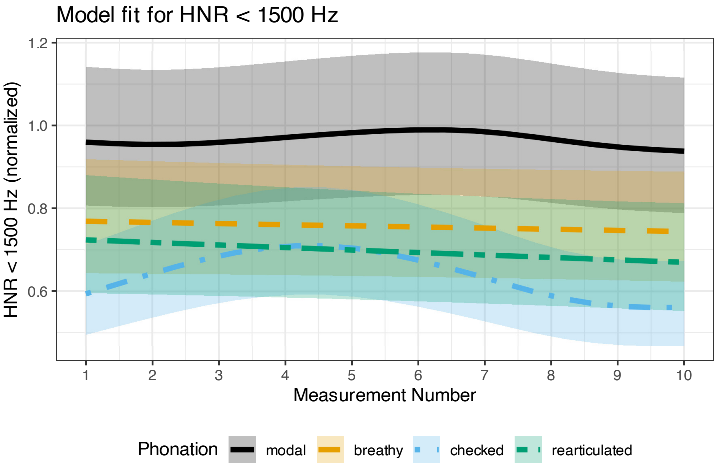
- Laryngeal complexity is measured by the number of phonation contrasts in a language.
- SLZ has four phonation contrasts.
- The more phonation contrasts a language has, the more complex its laryngeal system is.

Model fit for $f0$



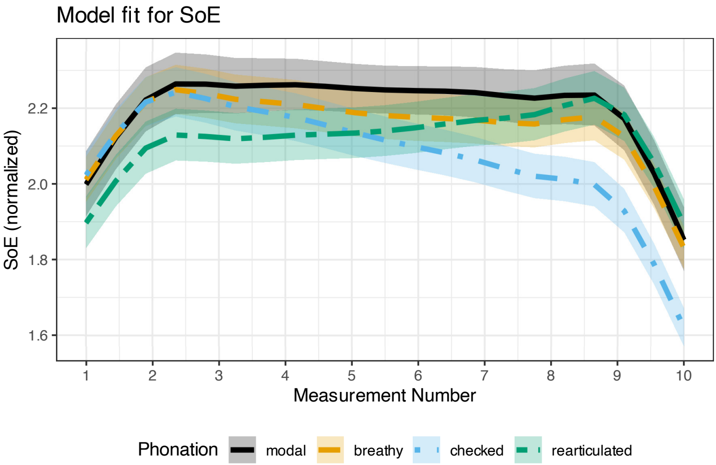
Model difference for $f0$

Model fit for HNR

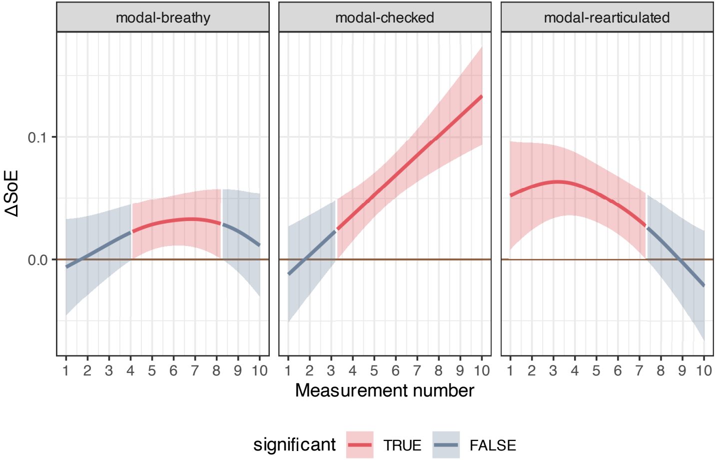


Model difference for HNR

Model fit for SoE



Model difference for SoE



Summary of Laryngeal Complexity

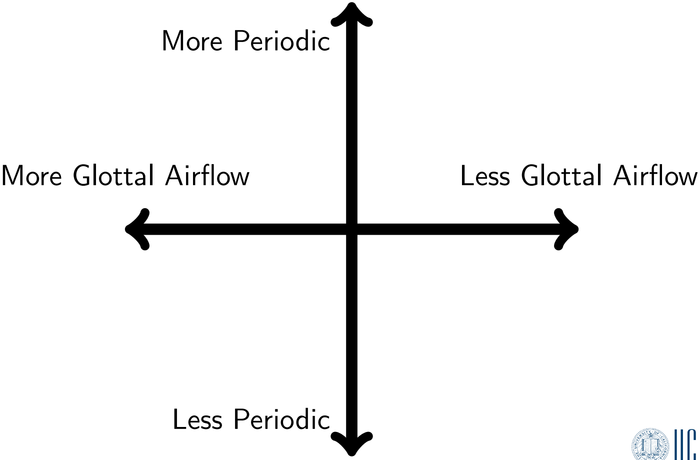
- SLZ has four phonation contrasts.
- SLZ's phonation occupies a three-dimensional space.
- Dimensions are correlated with glottal-airflow continuum (D1/D3) and nonmodal-to-modal continuum (D2).
- Dimensions are similar to those found in Keating et al. (2023).

Summary of Results

- SLZ's phonation occupies a three-dimensional space.
- Dimensions are correlated with glottal-airflow continuum (D1/D3) and nonmodal-to-modal continuum (D2).
    - Collaborated with acoustic measure correlations.
- Dimensions are similar to those found in Keating et al.

Summary

- Acoustic space can be reduced to two dimensions.
- More dimensions add more information about these two dimensions.



Summary

- Dimensionality reduction also occurs in a single language.
- Dimensions correspond to glottal-airflow and nonmodal-to-modal continua within a language and cross-linguistically.
- If additional dimensions are added, they only add additional information about these two dimension.

- Outlook
  - What are the perceptual cues that speakers use to distinguish between phonation types?
  - How do these dimensions relate to the phonology?

Duxhklhenhu' lhe' (Thank you)

Acknowledgements

- Thank you to the speakers in Santiago Laxopa for sharing their time and language expertise.
- Thank you to my Grant McGuire, Jaye Padgett, Marc Garellek, Ryan Bennett, Jack Duff, Maya Wax Cavallaro, and many others for their help and discussions during all stages of this project.

Acknowledgements

This work is supported by funding from:

- The National Science Foundation under Grant No. 2019804
- The Humanities Institute at UC Santa Cruz
- The Jacobs Research Funds

References I

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

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

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

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

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

References II

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

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

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

Brunelle, Marc & 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.

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

References III

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

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

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

References IV

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.

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

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

References V

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

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.

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

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.

References VI

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

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

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

Ladefoged, Peter. 1971. *Preliminaries to linguistic phonetics*. Chicago: University of Chicago.

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

References VII

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

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

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

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

References VIII

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

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

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