UNIVERSITY OF CALIFORNIA
SANTA CRUZ

VOICE QUALITY AND TONE AT THE PHONETICS AND 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 and phonology interface

by

Mykel Loren Brinkerhoff

Theses have elements. Isn't that nice?

To myself,

Mykel Loren Brinkerhoff,

the only person worthy of my company.

Acknowledgments

I want to “thank” my committee, without whose ridiculous demands, I would have graduated so, so, very much faster.

Chapter 1

Introduction

Every dissertation should have an introduction. You might not realize it, but the introduction should introduce the concepts, backgrouand, and goals of the dissertation.

Title	Author
War And Peace	Leo Tolstoy
The Great Gatsby	F. Scott Fitzgerald

Table 1.1: A normalsize table. There has been a complaint that table captions are not single-spaced. This is odd because the code indicates that they should be.

Table 1.2: A small table.

Title	Author
War And Peace	Leo Tolstoy
The Great Gatsby	F. Scott Fitzgerald

Chapter 2

Vowels and suprasegmentals in Santiago Laxopa Zapotec

2.1 Introduction

Santiago Laxopa Zapotec (SLZ; *Dilla'xhunh Laxup* [diʒaˀʐun lːaʂupʰ]) is a Northern Zapotec language spoken by approximately 1000 people in the municipality of Santiago Laxopa, Ixtlán,Oaxaca, Mexico and in diaspora communities throughout Mexico and the United States (Adler & Morimoto 2016, Adler et al. 2018, Foley, Kalivoda & Toosarvandani 2018, Foley & Toosarvandani 2020).

2.2 Vowels

2.2.1 Voice Quality

2.3 Tones

Chapter 3

On using Residual H1*

Chapter 4

The acoustic landscape of voice quality in Santiago Laxopa Zapotec

4.1 Introduction

This chapter details a study of 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 to reduce the dimensionality of a dataset to visualize the relationships between data points. In this study, MDS is used to visualize the acoustic space of voice quality in SLZ. The results of this analysis provide insight into the acoustic correlates of voice quality in SLZ and contribute to our understanding of the phonetic properties of this underdocumented language.

This study is based on the work conducted by Keating et al. (2023) on the acoustic space of voice quality in 11 languages. However, this study focuses on a single language, SLZ, and provides a detailed analysis of the acoustic properties of voice quality in this language. The results of this study will contribute to our understanding of the phonetic properties of SLZ and how the acoustic properties of voice quality in this language compare to 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 the ages of 18 and 60 years and consisted of 5 males and 5 females.

4.2.2 Recordings

The participants were asked to perform a word list elicitation task consisting of 76 words. These words were selected to elicit a wide range of voice quality types, including modal, creaky, and breathy voice. The words were selected based on previous research conducted as part of the Zapotec Language Project at the University of California, Santa Cruz (*Zapotec Language Project — University of California, Santa Cruz* 2022). Because participants were not literate in SLZ, the word list was prompted for them by asking them "How do you say [word in Spanish]?" in Zapotec by myself and another researcher. Participants were asked to respond with the desired word in the carrier phrase *Shnia' [WORD] chonhe lhas* "I say [WORD] three times." which was repeated three times. These utterances were recorded in a quiet environment using a Zoom H4n digital recorder. The recordings were saved as 16-bit WAV files with a sampling rate of 44.1 kHz.

4.2.3 Acoustic measuring

These resulting audio files were then processed in Praat to isolate the vowel portion of each word. The onset of the vowel was set to the second glottal pulse after the onset and the offset of the vowel was set to the last glottal pulse before the decrease in amplitude at the end of the vowel (Garellek 2020). The vowel was then extracted and saved as a separate file for analysis.

These vowels were fed into VoiceSauce (Shue, Keating & Vicenik 2009) to generate the acoustic measures for the studies discussed in this dissertation. Because many of the acoustic measures are based on the fundamental frequency; this measure was calculated using the STRAIGHT algorithm from (Kawahara, Cheveigne & Patterson 1998). The STRAIGHT algorithm estimates the fundamental frequency at one-millisecond (ms) intervals. Once the fundamental frequency is estimated, VoiceSauce then uses an optimization function to locate the harmonics of the spectrum, finding their amplitudes.

VoiceSauce then uses the Snack Sound toolkit (Sjölander 2004) to find the frequencies and bandwidths of the first four formants, also at one-millisecond intervals. The amplitudes of the harmonics closest to these formant frequencies are located and treated as the amplitudes of the formants. These formant frequencies and bandwidths are used to correct the harmonic amplitudes for the filtering effects of the vocal tract, using Iseli, Shue & Alwan's 2007 extension of the method employed by Hanson (1997). Each vowel was measured across ten equal time intervals, resulting in 22890 data points in total. These measures were then z-scored by speaker to reduce the variation between speakers and provide a way to directly compare the different measures on the same scale.

4.2.4 Data processing

Data points with an absolute z-score value greater than 3 were considered outliers and excluded from the analyses in the dissertation. Within each vowel category, the Mahalanobis distance was calculated in the F1-F2 panel. Each data point with a Mahalanobis distance greater than 6 was considered an outlier and excluded from the analysis.

Energy was excluded if it had a value of zero and then log-transformed to normalize its right-skewed distribution. Afterward, the resulting log-transformed data was z-scored and any data point with a z-score larger than 3 was excluded. This outlier removal resulted in 1918 datapoints being removed.

After removal of the outliers, I calculated residual H1* for the remaining data points following Chai & Garellek (2022). First, a linear mixed effects model was generated with the z-scored H1* as the response variable and the z-scored energy as fixed effect. The uncorrelated interaction of the z-scored energy by speaker was treated as random. The energy factor resulting from this linear mixed-effects model was extracted. Finally, the z-scored H1* was the product of the z-scored energy and the energy factor subtracted from it.

Once these steps were completed, the mean of the fifth and sixth intervals was taken for each vowel and speaker. This is similar to what Keating et al. (2023) did by taking the middle of the vowel for their analysis. The reason for this choice is to minimize 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 measures that were negative, which presents a problem for MDS analyses. To correct for this, I added the absolute value of the minimum z-score to each measure. This results in a dataset that still preserves the relative differences in the scores while providing a dataset that is all positive for the MDS analysis.

4.2.5 Statistical analysis

Using a multidimensional scaling (MDS) analysis is a statistical method of reducing the dimensionality of a dataset to visualize the relationships between the data points (Kruskal & Wish 1978). This is especially true when there are many variables that are could contribute to the data. In the case of voice quality this is especially true. As shown in Kreiman et al. (2014, 2021) and Garellek (2020) voice quality is pscyhoacoustically complex and a single measure is not enough to capture the full range of voice quality. Instead, what is required is multiple measures that function as cues for the different types of voice quality. For example, a vowel that is characterized as having a breathy voice has an elevated spectral-slope and a lowered harmonics-to-noise ratio than modal voice. A creaky voice is characterized as having a lowered spectral-slope and a lowered harmonics-to-noise ratio.

Because MDS analyses of many variables can result in rather unmeaningful results, I chose to focus on the speaker x voice quality interaction. This provides us with the opportunity to see how the speakers differ in their production of the different voice qualities. This means that each speakers production of the four phonation contrasts is represented as a single point in the MDS plot. This is similar to what Keating et al. (2023) did in their study of the acoustic space of voice quality in 11 languages. Except that they compared the language x voice quality interaction. Both of these interactions show us similar information, one shows us within a language, while the other shows us between languages.

The MDS analysis was conducted in R using the 'metaMDS' function in the 'vegan' package. The Manhattan distance was used to estimate the physical differences between the pairs of speaker x voice quality. Because the distances are non-Euclidean, the MDS analysis was conducted using the non-metric option.

This algorithm resulted in a solution that involves a number of different dimensions. The number of dimensions that are retained directly affects how well the original data is captured. To many dimensions and the data is overfit, to few and the data is underfit. To determine the number of dimensions to retain, I used a scree plot to plot the stress of each dimension. The 'elbow' of the curve was identified as the correct number of dimensions for the analysis. Figure 4.1 shows that most of the data is captured in a two-dimensional space. With the third dimension, some more subtle information about the voice quality is added.

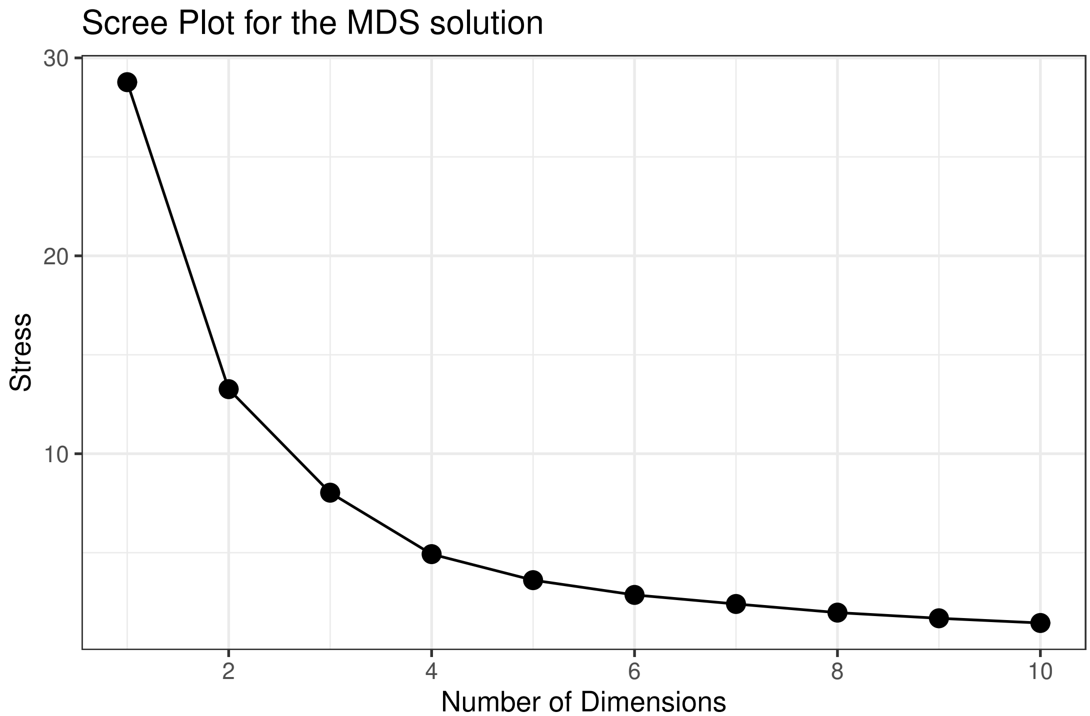

Figure 4.1: Scree plot for the MDS analysis.

4.3 Results

4.3.1 Acoustic space of voice quality

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

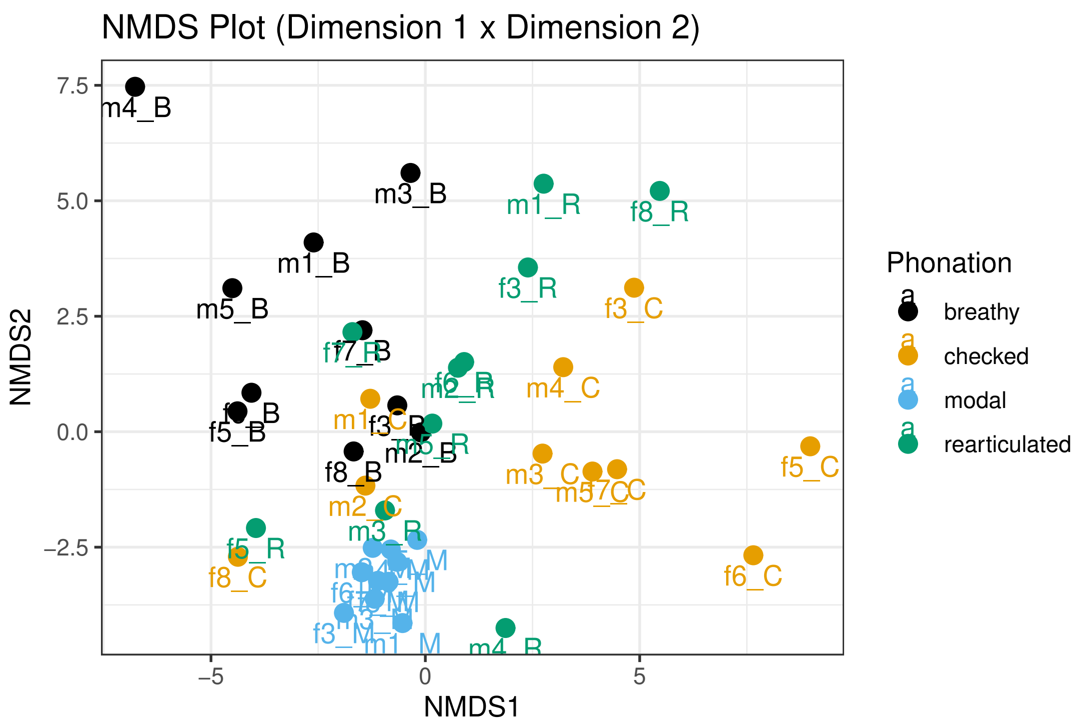

Figure 4.2: Two-dimensional MDS solution showing the first and second dimensions.

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

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

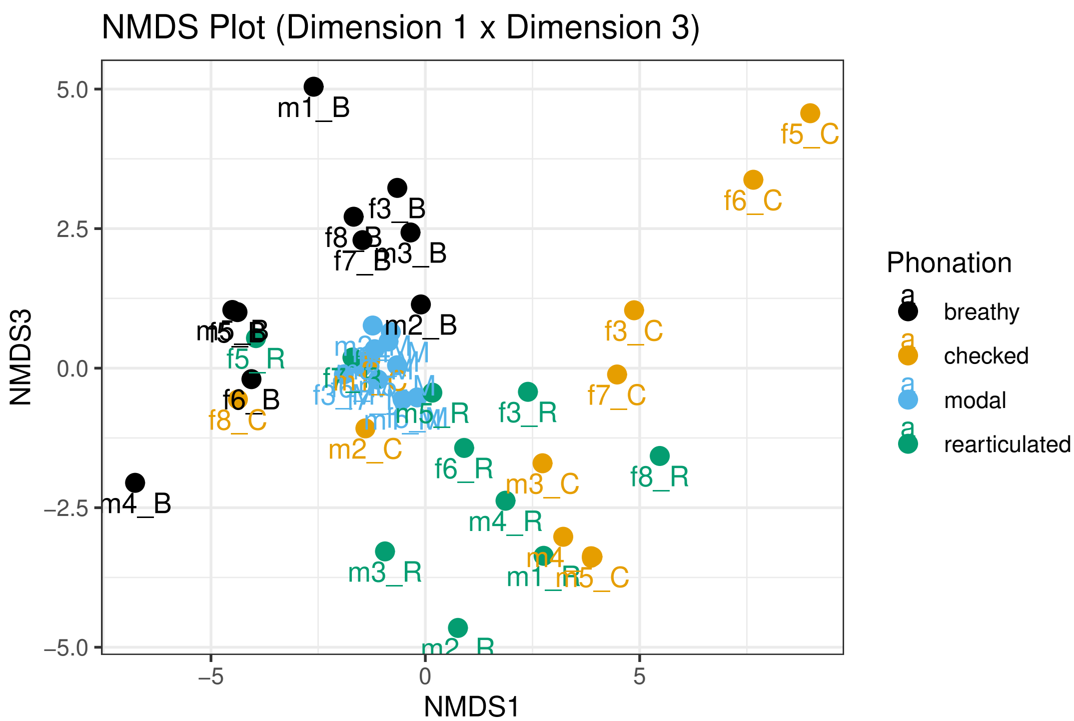

Figure 4.3: Two-dimensional MDS solution showing the first and third dimensions.

4.3.2 Acoustic correlates of voice quality

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

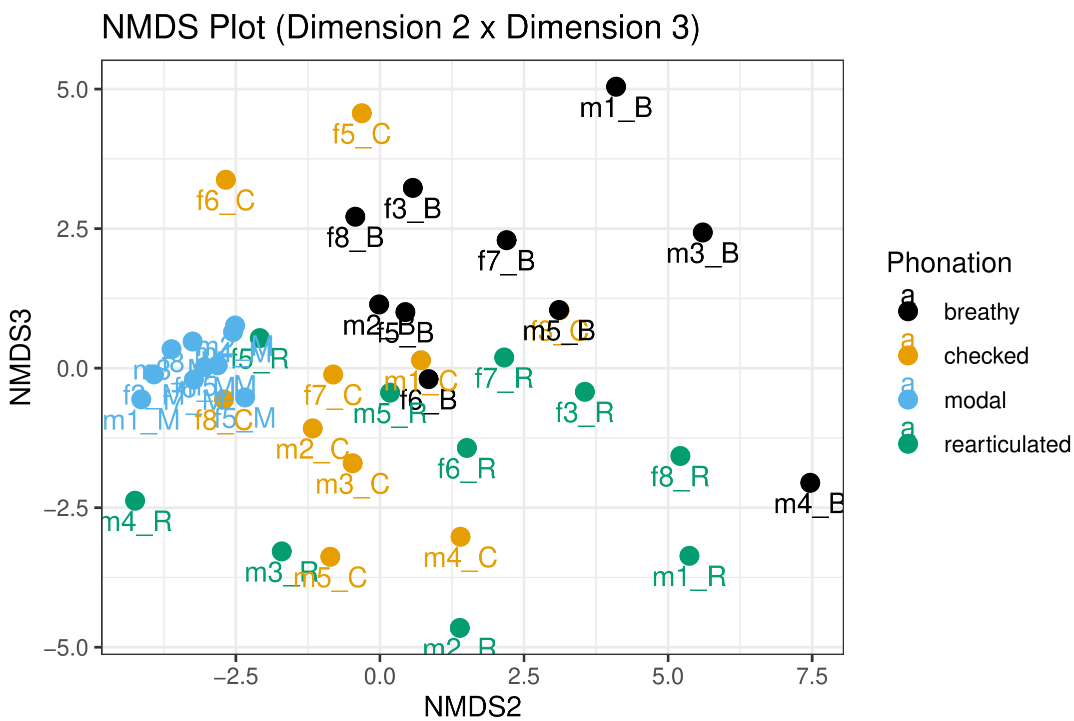

Figure 4.4: Two-dimensional MDS solution showing the second and third dimensions.

are in boldface (weights > 3.0).

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

Table 4.1: Weight of each acoustic measure along each dimension of the three-dimensional MDS solution (D1, D2, D3). Parameters that have weights higher than those of other parameters on each dimension are in boldface (weights > 4.0 for D1 and D2, and weights > 3.0 for D3).

Acoustic Measure	D1	D2	D3
h1h2	1.03	1.01	0.39
h2h4	1.15	3.98	2.13
h1a1	2.22	**5.15**	1.84
h1a2	2.93	**4.66**	1.00
h1a3	2.37	3.24	0.90
h42k	1.47	0.31	1.59
h2kh5k	3.73	0.73	0.84
h1_resid	1.75	0.97	**4.24**
h2	1.76	0.94	**4.09**
h4	0.79	4.28	0.10
a1	**4.96**	**5.48**	0.17
a2	**5.30**	**4.90**	1.38
a3	**4.54**	2.91	1.11
cpp	4.08	0.10	1.68
hnr05	**5.66**	1.47	1.81
hnr15	3.95	2.68	**3.08**
hnr25	3.15	1.63	**3.42**
hnr35	2.86	0.55	**3.19**
soe	2.09	0.78	0.36
shr	2.39	0.50	0.47
energy	2.22	3.91	0.64

4.4 Discussion

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

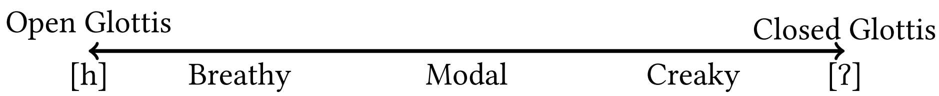

Figure 4.5: A diagram showing the relationship between breathy, modal, and creaky phonation types. Based on Gordon & Ladefoged (2001).

As mentioned above, the measures that contribute the most this dimension are the amplitudes of the first three formants and HNR < 500 Hz. It is interesting even though this dimension is similar to the open-quotient model, we do not observe measures that are traditionally associated with this measure (i.e., spectal-slope). Instead of seeing traditional spectral-slope measures, we find the three formant amplitudes which are 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 amplitude of the formants than the spectral-slope of the signal. This is combined with the HNR < 500 Hz which is a measure of the harmonics-to-noise ratio for the first 500 Hz of the signal.

The second dimension divides the space into modal versus non-modal voice quality.

The third dimension adds more information about non-modal voice quality. As seen in Figure 4.3 and Figure 4.4, this

This dimension is characterized by the harmonics-to-noise ratio for the first 1500 Hz, 2500 Hz, and 3500 Hz. This suggests that the third dimension is more about the spectral quality of the signal than the amplitude of the formants. This is combined with the residual H1* and H2* which are measures of the spectral-slope of the signal.

4.5 Conclusion

Chapter 5

Trees reveal the importance of measures in SLZ

Chapter 6

Testing the laryngeal complexity hypothesis

6.1 Laryngeal Complexity

Some other research was once performed.

6.2

Figure 6.1: A first figure.

Figure 6.2: A second figure.

Chapter 7

Conclusion

Bibliography

Adler, Jeff, Steven Foley, Jed Pizarro-Guevara, Kelsey Sasaki & Maziar Toosarvandani. 2018. The derivation of verb initiality in Santiago Laxopa Zapotec. In Jason Merchant, Line Mikkelsen, Deniz Rudin & Kelsey Sasaki (eds.), *A reasonable way to proceed: Essays in honor of Jim McCloskey*, 31–49. Santa Cruz, Berkeley, Chicago: University of California.

Adler, Jeff & Maho Morimoto. 2016. Acoustics of phonation types and tones in Santiago Laxopa Zapotec. *The Journal of the Acoustical Society of America* 140(4). 3109–3109. https://doi.org/10.1121/1.4969713.

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.

Black, Cheryl A. 1996. A backwards binding construction in Zapotec. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 40(1). https://doi.org/10.31356/silwp.vol40.01.

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.

Brunelle, Marc. 2012. Dialect experience and perceptual integrality in phonological registers: Fundamental frequency, voice quality and the first formant in Cham. *The Journal of the Acoustical Society of America* 131(4). 3088–3102. https://doi.org/10.1121/1.3693651.

Brunelle, Marc & James Kirby. 2016. Tone and Phonation in Southeast Asian Languages. *Language and Linguistics Compass* 10(4). 191–207. https://doi.org/10.1111/lnc3.12182.

Bueno Holle, Juan José. 2019. *Information structure in Isthmus Zapotec narrative and conversation* (Topics at the Grammar-Discourse Interface 3). https://doi.org/10.5281/ZENODO.2538324.

Campbell, Eric W. 2017. Otomanguean historical linguistics: Past, present, and prospects for the future. *Language and Linguistics Compass* 11(4). e12240. https://doi.org/10.1111/lnc3.12240.

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

Chávez-Peón, Mario E. 2010. *The interaction of metrical structure, tone, and phonation types in Quiaviní Zapotec*. University of British Columbia dissertation. https://doi.org/10.14288/1.0071253.

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.

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, 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.

Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29(4). 383–406. https://doi.org/10.1006/jpho.2001.0147.

Hanson, Helen M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America* 101(1). 466–481. https://doi.org/10.1121/1.417991.

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

Kawahara, Hideki, Alain De Cheveigne & Roy D. Patterson. 1998. An instantaneous-frequency-based pitch extraction method for high-quality speech transformation: revised TEMPO in the STRAIGHT-suite. In *5th International Conference on Spoken Language Processing (ICSLP 1998)*, paper 0659–. ISCA. https://doi.org/10.21437/ICSLP.1998-555.

Keating, Patricia, Marc Garellek & Jody Kreiman. 2015. Acoustic properties of different kinds of creaky voice. In *Proceedings of the 18th International Congress of Phonetic Sciences.*

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.

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.

Lee, Felicia. 2006. *Remnant raising and VSO clausal architecture: A case study of San Lucas Quiaviní Zapotec*. Vol. 66 (Studies in Natural Language and Linguistic Theory (SNLT)). Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-4308-2.

Merrill, Elizabeth D. 2008. Tilquiapan Zapotec. *Journal of the International Phonetic Association* 38(01). https://doi.org/10.1017/S0025100308003344.

Nellis, Donald G. & Barbara E. Hollenbach. 1980. Fortis versus Lenis in Cajonos Zapotec Phonology. *International Journal of American Linguistics* 46(2). 92–105. https://doi.org/10.1086/465639.

Operstein, Natalie. 2017. Suppletion in Zapotec. *Linguistics* 55(4). https://doi.org/10.1515/ling-2017-0010.

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.

Sicoli, Mark. 2007. *Tono: A linguistic ethnography of tone and voice in a Zapotec region (Mexico)*. University of Michigan dissertation.

Sicoli, Mark A. 2010. Shifting voices with participant roles: Voice qualities and speech registers in Mesoamerica. *Language in Society* 39(4). 521–553. https://doi.org/10.1017/S0047404510000436.

Sjölander, Kåre. 2004. *Snack Sound Toolkit*. Stockholm, Sweden: KTH.

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.

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.

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

Appendix A

Some Ancillary Stuff

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