On Residual H1 as a measure of voice quality

Mykel Loren Brinkerhoff & Grant McGuire

Abstract

text.

Keywords:

1 Introduction

2 Santiago Laxopa Zapotec

Santiago Laxopa Zapotec is a Northern Zapotec language of Oto-manguean language family (Adler & Morimoto, 2016; Adler et al., 2018; Brinkerhoff, Duff, & Wax Cavallaro, 2021, 2022; Foley, Kalivoda, & Toosarvandani, 2018; Foley & Toosarvandani, 2020; Sichel & Toosarvandani, 2020a, 2020b). It is spoken by 981 people in the municipality of Santiago Laxopa, Ixtlán, Oaxaca, Mexico ("Santiago Laxopa," n.d.) and a small number of other speakers in diaspora throughout Mexico and the United States. Similar to other Oto-manguean languages, Santiago Laxopa Zapotec is laryngeally complex. According to Blankenship (1997, 2002) and Silverman (1997a, 1997b), laryngeal complexity refers to how these languages make use of both contrastive tone and contrastive voice quality.

Santiago Laxopa Zapotec exhibits the standard five vowel inventory. These five vowels are further distinghuished by the use of contrastive voice quality, as seen in

(1) a. b. c. d.

vowel quality is

3 Methodology

3.1 Elicitation

Ten native speakers of SLZ (five female; five male) participated in a wordlist eliciation. Elication was done in the pueblo of Santiago Laxopa, Ixtlán, Oaxaca, Mexcio during the summer of 2022 on a Zoom H4n handheld recorder (16 bit, 44.5 Khz).

The wordlist consisted of 72 items repeated three times each in isolation and the carrier sentence *Shnia' X chonhe lhas* "I say X three times".[1] Between these 72 words, there were 11 words

[1] See Appendix 1 for wordlist

with breathy voice, 9 with rearticulated, 10 with checked voice, and 42 with modal. Thirteen of the seventy-two words were disyllabic and the majority contained the same phonation type. Of those thirteen only five words contained mixed voicing.

3.2 Data Processing

Each vowel from the target words in the carrier sentence condition was labeled following Garellek (2020) for where the vowel began and ended. Each vowel from the word list was annotated for speaker, word, vowel, tone, voice quality, and utterance number. This labeling was conducted for each of the vowels located in the target word from the elicitation list from the carrier sentences.

These vowels were then extracted and fed into VoiceSauce for acoustic measuring (Shue, Keating, & Vicenik, 2009). Formants were measured using the Snack (Sjölander, 2004) while the fundamental frequency (*f0*) was measured using the STRAIGHT algorithm (Kawahara, Cheveigne, & Patterson, 1998). Spectral slope measures were corrected for formants and bandwidths (Hanson, 1997; Iseli, Shue, & Alwan, 2007).

Because the data contains variables for the grand mean for the different acoustic measures and the means of each tenth of the vowel, the columns were rearranged into a new data frame where each tenth of a vowel's acoustic measurement is located under a single variable with the name of the acoustic measure. This required the creation of a new variable called time. This results in 22890 rows of data After rearranging the data outliers were removed. F0

Data was first grouped by speaker then the z-score was calculated for f0.

If the absolute value of f0 was greater than 3, it was removed. This is because 99.7% of the data in a normally distributed dataset lies within 3 SDs of the mean. Anything greater than 3 is likely an outlier and marked as such. Formants Data was again grouped by speaker, and the Mahalanobis distance (Drumond, Rolo, & Costa, 2019; Mahalanobis, 2018; Martos, Muñoz, & González, 2013) was calculated for F1 and F2. A Mahalanobis distance greater than 6 means that you are a likely outlier This was done by taking the covariance and means of F1 and F2. This gives you a grouping based on the vowels' formants. The Mahalanobis distance was calculated based on F1 and F2 The data was filtered by each vowel and then outliers were determined. Energy If energy was equal to 0 it was converted to NA I then took the log10 of energy across all datapoints because the data is left bounded by 0 and has a long right tail. After determining which items were outliers they were filtered out.

Standardization The data was grouped by each speaker before calculating the z-scores. Z-scores were calculated for each of the measures except for Strength of Excitation which was normalized according to Garellek et al. (2021) This was done to bring all measurements into the same scale to facilitate better comparisons across speakers for the same measures. This measure works best.

We are not trying to normalize the data but bring everything into the same frame of reference.

Calculating Residual H1* 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 z-scored energy by speaker was treated as random. This is also how residual H1 was calculated in the supplementary material from Chai and Garellek (2022). The resulting residual H1 model's energy factor was extracted Residual h1 was added as a variable to the dataframe by taking the z-scored H1* and subtracting the product of the z-scored energy and the energy factor

3.3 Statistical Modeling

Each model was Acoustic measure ∼Phonation*Position + Tone + (1|Speaker:Word:Iter) + (1|Vowel)

Acoustic measure represents either H1*-H2* (z-score), H1*-A3 (z-score), or residual H1* Phonation*Position is a fixed effect that accounts for the interaction between Phonation and Position Tone is another fixed effect for the five different tones across the vowels This needs to be separated for a few reasons it makes sense for physical reasons tone and phonation are so closely linked that we mostly cover our bases by only including the position interaction with phonation, which avoids perfectly collinear interactions This was revealed by the rank deficiency that occurs when we cross Phonation with Tone. This also makes sense because there are several tone and phonation contrasts that we were not able to capture. (1|Speaker:Word:Iter) is a random effect that accounts for the interaction between these Speaker, Word, and Utterance number. This allows us to capture the fact that each speaker said the same words three times each (1|Vowel) is a random effect that captures the fact that each phonation occurred with different vowels. We also know that certain vowels could adversely affect the raw acoustic measures.

4 Results

4.1 H1*-H2*

4.2 H1*-A3

4.3 Residual H1*

4.4 Model Comparison

According to Casella and Berger (2002), which is the gold standard for statistical inference in the statistics field, the best way to compare models is by comparing the AIC and the likelihood ratio between the models. Using std error to compare models is prone to p-hacking and is heavily frowned upon by statisticians This was done by using the function lrtest() from the lmtest

Table 1: Results of the statistical model for H1*-H2*.

	Estimate	Std. Error	df	t value	p-value	
Intercept	0.02819	0.1033	4.755	0.273	0.796	
Breathy	-3.446e-02	4.21e-02	9.821e+03	-0.818	0.413507	
Checked	-1.402e-01	4.055e-02	1.465e+04	-3.457	0.000547	***
Laryngealized	-1.361e-01	4.969e-02	5.348e+03	-2.740	0.006174	**
Position 2	4.749e-03	1.339e-02	1.902e+04	0.355	0.722877	
Position 3	-1.795e-02	1.259e-02	1.903e+04	-1.426	0.154005	
High tone	1.531e-02	4.460e-02	4.358e+03	0.343	0.731401	
Low tone	-7.839e-02	3.220e-02	8.169e+03	-2.435	0.014931	*
Mid tone	-1.054e-01	4.400e-02	7.421e+03	-2.395	0.016637	*
Rising tone	2.868e-01	6.922e-02	1.022e+04	4.144	3.45e-05	***
Breathy:Position 2	2.104e-02	3.070e-02	1.904e+04	0.685	0.493043	
Checked:Position 2	-3.435e-04	3.598e-02	1.905e+04	-0.010	0.992383	
Laryngealized:Position 2	-6.258e-02	3.227e-02	1.903e+04	-1.940	0.052436	.
Breathy:Position 3	-1.6635e-01	2.901e-02	1.908e+04	5.634	1.78e-08	***
Checked:Position 3	4.822e-02	3.401e-02	1.909e+04	1.418	0.156192	
Laryngealized:Position 3	1.409e-01	3.026e-02	1.906e+04	4.657	3.23e-06	***

Table 2: Results of the statistical model for H1*-A3.

	Estimate	Std. Error	df	t value	p-value	
Intercept	-3.108e-01	8.527e-02	5.029e+00	-3.645	0.014674	*
Breathy	3.870e-01	3.951e-02	1.014e+04	9.795	< 2e-16	***
Checked	-5.464e-01	3.694e-02	1.274e+04	-14.790	< 2e-16	***
Laryngealized	-9.881e-02	4.886e-02	5.909e+03	-2.022	0.043188	*
Position 2	-1.964e-02	1.154e-02	1.899e+04	-1.703	0.088627	.
Position 3	3.045e-03	1.085e-02	1.900e+04	0.281	0.778964	
High tone	2.596e-01	4.430e-02	4.933e+03	5.859	4.95e-09	***
Low tone	3.514e-01	3.055e-02	1.095e+04	11.504	< 2e-16	***
Mid tone	8.087e-02	4.212e-02	9.234e+03	1.920	0.054923	.
Rising tone	7.312e-01	6.457e-02	1,431e+04	11.323	< 2e-16	***
Breathy:Position 2	-1.424e-02	2.645e-02	1.901e+04	-0.538	0.590425	
Checked:Position 2	2.067e-02	3.100e-02	1.902e+04	0.667	0.504998	
Laryngealized:Position 2	-4.614e-02	2.780e-02	1.900e+4	-1.660	0.097006	.
Breathy:Position 3	-7.101e-02	2.500e-02	1.904+04	-2.840	0.004516	***
Checked:Position 3	-2.534e-04	2.931e-02	1.904e+04	-0.009	0.993102	
Laryngealized:Position 3	9.593e-02	2.607e-02	1.902e+04	3.679	0.000235	***

package and using the function AIC() from the base stat package The model with the highest Log-Likelihood ratio and lowest AIC is the most robust.

Table 3: Results of the statistical model for Residual H1*.

	Estimate	Std. Error	df	t value	p-value	
Intercept	-1.051e-01	4.346e-02	7.543e+00	-2.418	0.043781	*
Breathy	1.502e-01	3.184e-02	5.952e+03	4.716	2.46e-06	***
Checked	-3.848e-01	3.070e-02	6.691e+03	-12.535	< 2e-16	***
Laryngealized	-4.856e-01	3.747e-02	4.769e+03	-12.960	< 2e-16	***
Position 2	1.369e-02	1.027e-02	1.902e+04	1.333	0.182633	
Position 3	4.386e-02	9.660e-03	1.903e+04	4.541	5.64e-06	***
High tone	9.129e-02	3.345e-02	3.138e+03	2.729	0.006393	**
Low tone	1.854e-01	2.441e-02	7.861e+03	7.595	3.44e-14	***
Mid tone	5.386e-02	3.332e-02	7.052e+03	1.616	0.106031	
Rising tone	3.346e-01	5.260e-02	9.818e+03	6.360	2.10e-10	***
Breathy:Position 2	1.160e-01	2.355e-02	1.904e+04	4.926	8.47e-07	***
Checked:Position 2	1.006e-01	2.760e-02	1.905e+04	3.645	0.000268	***
Laryngealized:Position 2	-2.675e-02	2.475e-02	1.903e+04	-1.081	0.279849	
Breathy:Position 3	1.149e-01	2.225e-02	1.908e+04	5.163	2.45e-07	***
Checked:Position 3	-1.033e-01	2.608e-02	1.909e+04	-3.959	7.55e-05	***
Laryngealized:Position 3	1.203e-01	2.321e-02	1.906e+04	5.183	2.21e-07	***

Table 4: log likelihood scores and AIC for the three statistical models.

Model	Log-Likelihood Ratio	AIC
H1-H2 model	-21716	43469.29
H1- A3 model	-19048	38134.12
Residual H1 model	-16113	32264.81

5 Discussion

6 Conclusion

References

Adler, J., Foley, S., Pizarro-Guevara, J., Sasaki, K., & Toosarvandani, M. (2018). The derivation of verb initiality in Santiago Laxopa Zapotec. In J. Merchant, L. Mikkelsen, D. Rudin, & K. Sasaki (Eds.), *A reasonable way to proceed: Essays in honor of Jim McCloskey* (pp. 31–49). University of California.

Adler, J., & Morimoto, M. (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

Blankenship, B. (1997). *The time course of breathiness and laryngealization in vowels* [Doctoral dissertation, University of California, Los Angeles].

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

Brinkerhoff, M. L., Duff, J., & Wax Cavallaro, M. (2021). Downstep in Santiago Laxopa Zapotec and the prosodic typology of VSO languages. *Manchester Phonology Meeting*.

Brinkerhoff, M. L., Duff, J., & Wax Cavallaro, M. (2022). *Tonal patterns and their restrictions in Santago Laxopa Zapotec* (Presentation). Virtual.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2. ed). Duxbury.

Chai, Y., & Garellek, M. (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

Drumond, D. A., Rolo, R. M., & Costa, J. F. C. L. (2019). Using Mahalanobis Distance to Detect and Remove Outliers in Experimental Covariograms. *Natural Resources Research*, *28*(1), 145–152. https://doi.org/10.1007/s11053-018-9399-y

Foley, S., Kalivoda, N., & Toosarvandani, M. (2018). Forbidden clitic clusters in Zapotec. In D. Edmiston, M. Ermolaeva, E. Hakgüder, J. Lai, K. Montemurro, B. Rhodes, A. Sankhagowit, & M. Tabatowski (Eds.), *Proceedings of the Fifty-third Annual Meeting of the Chicago Linguistic Society* (pp. 87–102).

Foley, S., & Toosarvandani, M. (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, M. (2020). Acoustic Discriminability of the Complex Phonation System in !Xóõ. *Phonetica*, *77*(2), 131–160. https://doi.org/10.1159/000494301

Garellek, M., Chai, Y., Huang, Y., & Van Doren, M. (2021). Voicing of glottal consonants and non-modal vowels. *Journal of the International Phonetic Association*, 1–28. https://doi.org/10.1017/S0025100321000116

Hanson, H. 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, M., Shue, Y.-L., & Alwan, A. (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice sourcea). *The Journal of the Acoustical Society of America*, *121*(4), 2283–2295. https://doi.org/10.1121/1.2697522

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

Mahalanobis, P. C. (2018). On the Generalized Distance in Statistics. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, *80*, S1–S7.

Martos, G., Muñoz, A., & González, J. (2013). On the Generalization of the Mahalanobis Distance. In J. Ruiz-Shulcloper & G. Sanniti di Baja (Eds.), *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications* (pp. 125–132). Springer. https://doi.org/10.1007/978-3-642-41822-8_16

Santiago Laxopa: Economy, employment, equity, quality of life, education, health and public safety. (n.d.). Data México. Retrieved April 26, 2024, from https://www.economia.gob.mx/datamexico/en/profile/geo/santiago-laxopa

Shue, Y.-L., Keating, P., & Vicenik, C. (2009). VOICESAUCE: A program for voice analysis. *The Journal of the Acoustical Society of America*, *126*(4), 2221. https://doi.org/10.1121/1.3248865

Sichel, I., & Toosarvandani, M. (2020a). *The featural life of nominals* [lingbuzz/005523].

Sichel, I., & Toosarvandani, M. (2020b). Pronouns and Attraction in Sierra Zapotec. In A. Hedding & M. Hoeks (Eds.), *Syntax and semantics at Santa Cruz, Volume IV*. Linguistics Research Center.

Silverman, D. (1997a). *Phasing and recoverability*. Garland Pub.

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

Sjölander, K. (2004). *Snack sound toolkit*. Stockholm, Sweden.

Appendix 1: Elicitation word list