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, Foley, Kalivoda & Toosarvandani 2018, Foley & Toosarvandani 2020, Sichel & Toosarvandani 2020a,b, Brinkerhoff, Duff & Wax Cavallaro 2021, 2022). It is spoken by 981 people in the municipality of Santiago Laxopa, Ixtlán, Oaxaca, Mexico (*Santiago Laxopa* 2022) 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 Silverman (1997a,b) and Blankenship (1997, 2002), 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, which are further distinghuished by the use of a four-way contrast in voice quality. This variety is unique for being a Northern Core Zapotec that has developed breathy voice in addition to the two types of laryngealization that characterize the rest of the Zapotec languages, namely checked and rearticulated (see Ariza-García 2018 for a typological study of voice quality distinctions in Zapotec languages). This contrast can be seen in the four-way near-minimal quadruple in (1).

(1) Four-way near-minimal contrast for voice quality in Santiago Laxopa Zapotec.

a. Modal voice
   ya [ʝa˦] ‘temazcal’
b. Breathy voice
   yah [ʝa̤˧] ‘iron’
c. Checked
   cha’ [t͡ʃaa̰˧] ‘pot’
d. Rearticulated
   ya’a [ʝaa̰a˧] ‘market’

Additionally, being an Oto-manguean language, Santiago Laxopa Zapotec is tonal with three level tones (H, M, and L) and two contours (MH and HL) that are allowed to appear on nouns (Brinkerhoff, Duff & Wax Cavallaro 2022).

These contrasts in voice quality present a rich array of information for testing the reliability of the voice quality measures.

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 76 words, there were 11 words with breathy voice, 9 with rearticulated, 10 with checked voice, and 42 with modal. Thirteen of the 76 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). Each vowel was measured with ten equal time intervals, resulting in 22890 data points in total.

The data was cleaned of outliers following the same steps taken by Chai & Garellek (2022) in their study. The H1*, H1*–H2*, and H1*-A3 and *f0* values were z-scored by speaker to reduce the variation between speakers and provide a way to directly compare the different measures on the same scale. Data points with an absolute z-score value larger than 3 were considered as outliers

[1] See Appendix 1 for wordlist

and were excluded from analyses. Within each vowel category, we calculated the Mahalanobis distance on the F1-F2 panel. Each data point with a Mahalanobis distance larger than 6, was considered an outlier and excluded from analysis. This is similar to what was done in Garellek & Esposito (2021), Seyfarth & Garellek (2018), and Chai & Ye (2022). Time points whose f0, F1, or F2 values were outliers were also excluded from H1* and H1*–H2* analyses because H1* and H1*–H2* are calculated based on *f0*, F1, and F2. Energy was excluded if it had a value of zero and then log-transformed to normalize its right-skewed distribution. Afterward the resulted 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 outliers were removed, we 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 z-scored energy by speaker was treated as random. The resulting residual H1 model's energy factor was extracted. Finally, the z-scored H1* had the product of the z-scored energy and the energy factor subtracted from it, leaving us with residual H1*.

After outlier removal and creation of residual H1* the remaining measures were assigned based on their position in the vowel (first, middle, and third) for statistical modeling.

3.3 Statistical Modeling

Three linear mixed effects regression models were fitted, one each for H1*-H2*, H1*-A3, and residual H1*. Each model had the tone and the interaction between voice quality and position in the vowel as fixed effects and vowel and the interaction between speaker, word, and repetition as random intercepts.[2]

The tone found on each data point and the interaction between voice quality and position in the vowel were selected as fixed effects for several reasons. The first being that there are five unique tones that appeared accross the data and it is well established that tone interacts with voice quality in several ways (see Esposito & Khan 2020, Garellek 2019 for discussion). By treating tone as a fixed effect in our model we can account for these interactions. The reason for including the interaction between voice quality and position in the vowel as a fixed effect has to do to the physical differences in the two different laryngealizations; checked and rearticulated vowels. Checked vowels in Zapotec languages have a glottal occlusion or a short period of creaky voice located at the right edge of the vowel. This is in contrast to rearticulated vowels were there is a glottal occlusion or creaky voice in the middle of the vowel. Because this difference between checked and rearticulated vowels is temporal in nature we can account for this difference through

[2] *Measure ∼Phonation*Position + Tone + (1|Speaker:Word:Iter) + (1|Vowel)*

the interaction of voice quality and position in the vowel.[3]

The interaction between speaker, word, and repition was treated as a random intercept because this allows us to take into account that each speaker said each word from the elicitation list three times. This intercept accounts for not only the intraspeaker variablity but the interspeaker variablity as well during each time the word was uttered. Treating vowel as a random intercept allows us to capture the fact that each voice quality occurred with different vowels.

4 Results

4.1 H1*-H2*

Figure 1 is a notched boxplot of the four voice qualities in each third of the vowel. We observe that each of the non-modal voice qualites have lower values than the modal in all positions of the vowel. It does, however, appear as if the value for rearticulated is nearly identical to the value for modal in the final third of the vowel. Additionally, we observe that these measures all overlap to a large degree in each of the three positions.

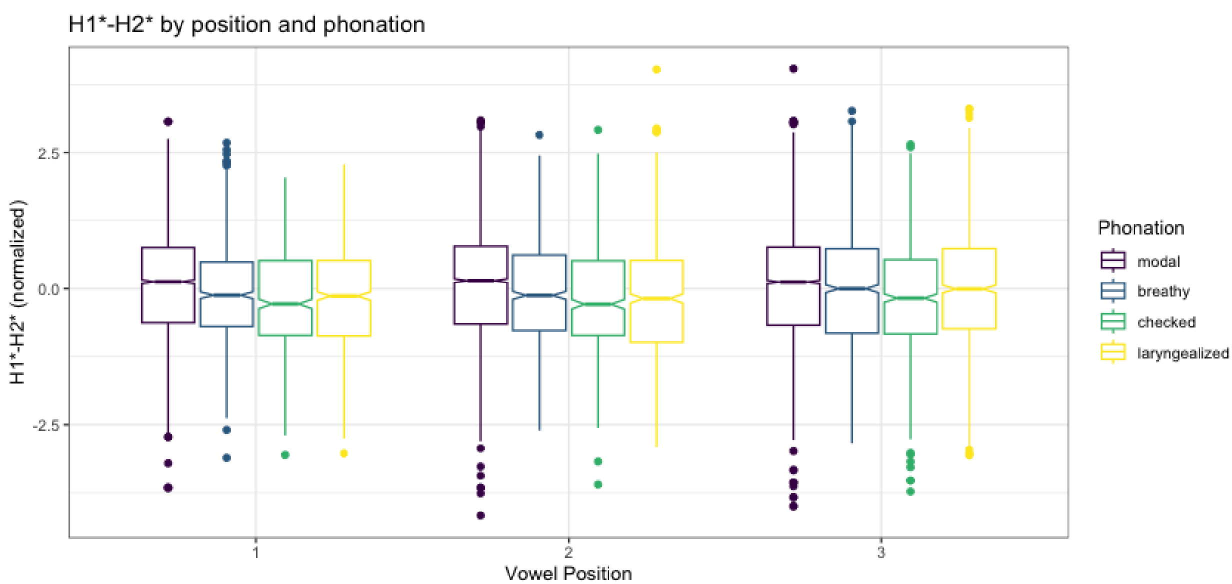

Figure 1: In a notched box plot, the notches extend 1.58*IQR/sqrt(n). This gives a roughly 95% confidence interval for comparing medians.

This is further supported when we compare the plots of each of the voice qualities in a line plot using a loess smooth, as in Figure 2. In this plot, each of the voice qualities is plotted at each of

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

the ten intervals of the vowel. A loess smooth was plotted over each to show how the acousitc measure functions across the ten intervals. We see that breathy, checked, and rearticulated all have a lower value than the modal at each of the first nine intervals. In the final interval, breathy and rearticulated are essentially equal to modal's value. In contrast, checked's value remains lower than the modal's value throughout the entire vowel.

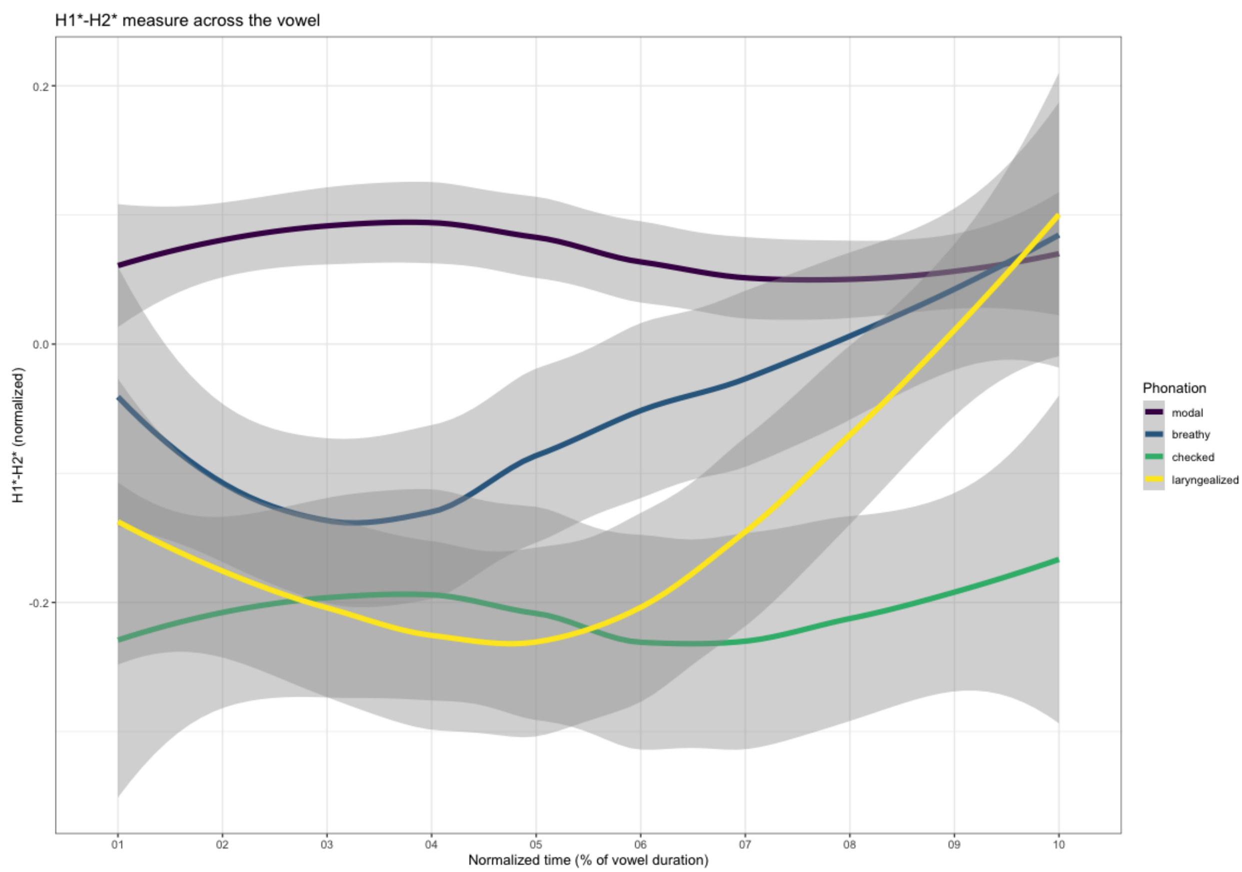

Figure 2: <caption>

4.2 H1*-A3

4.3 Residual H1*

4.4 Model Comparison

According to Casella & 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

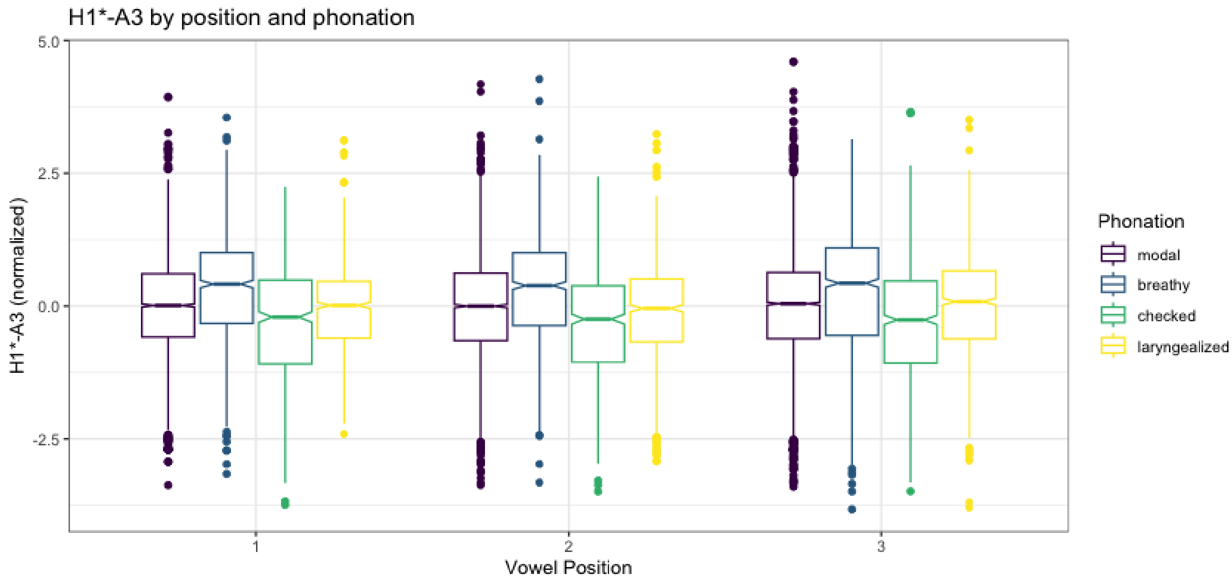

Figure 3: <caption>

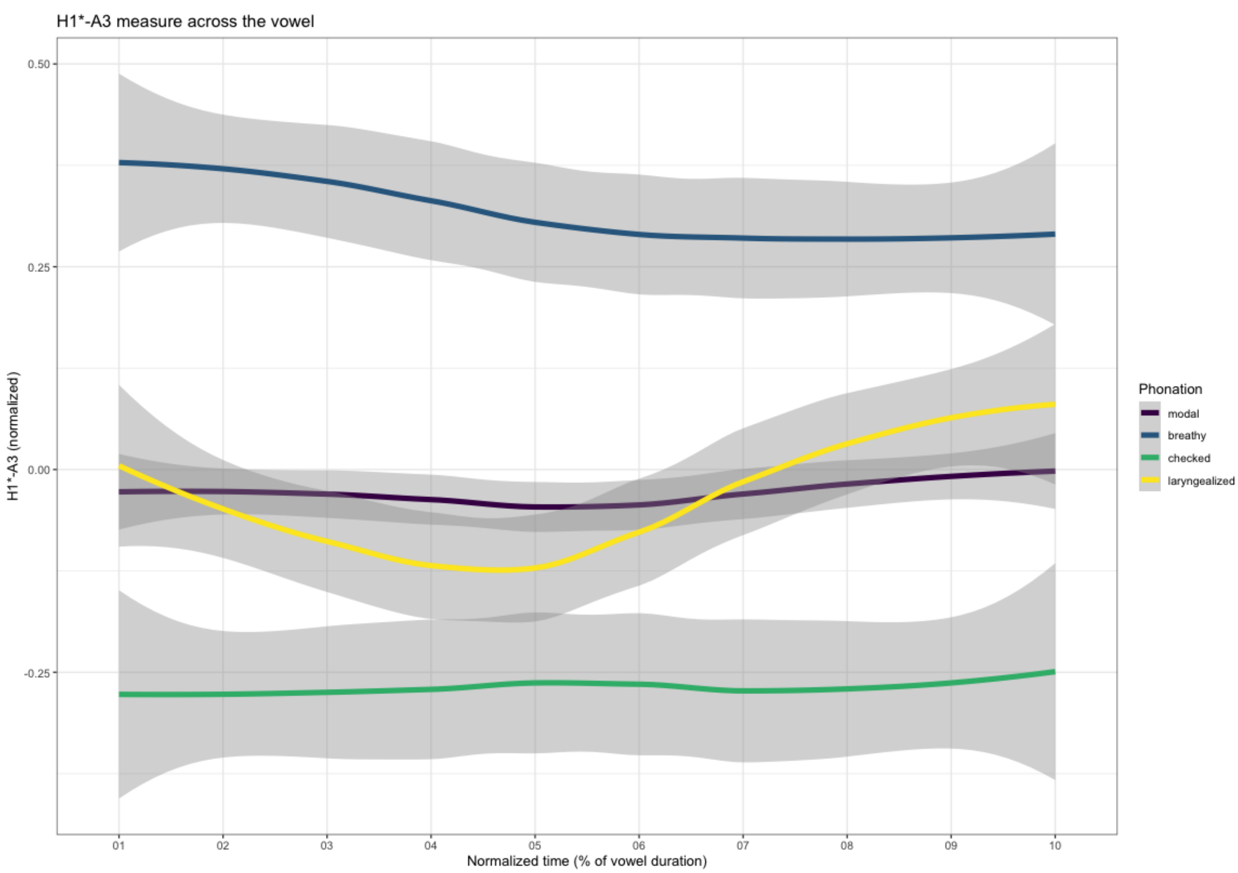

Figure 4: <caption>

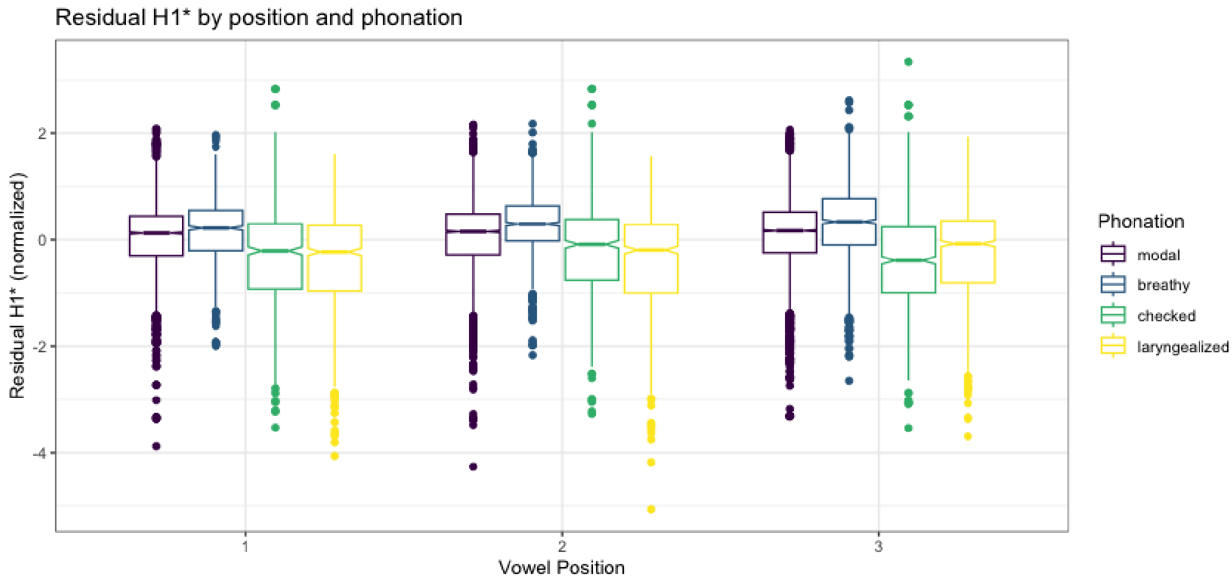

Figure 5: <caption>

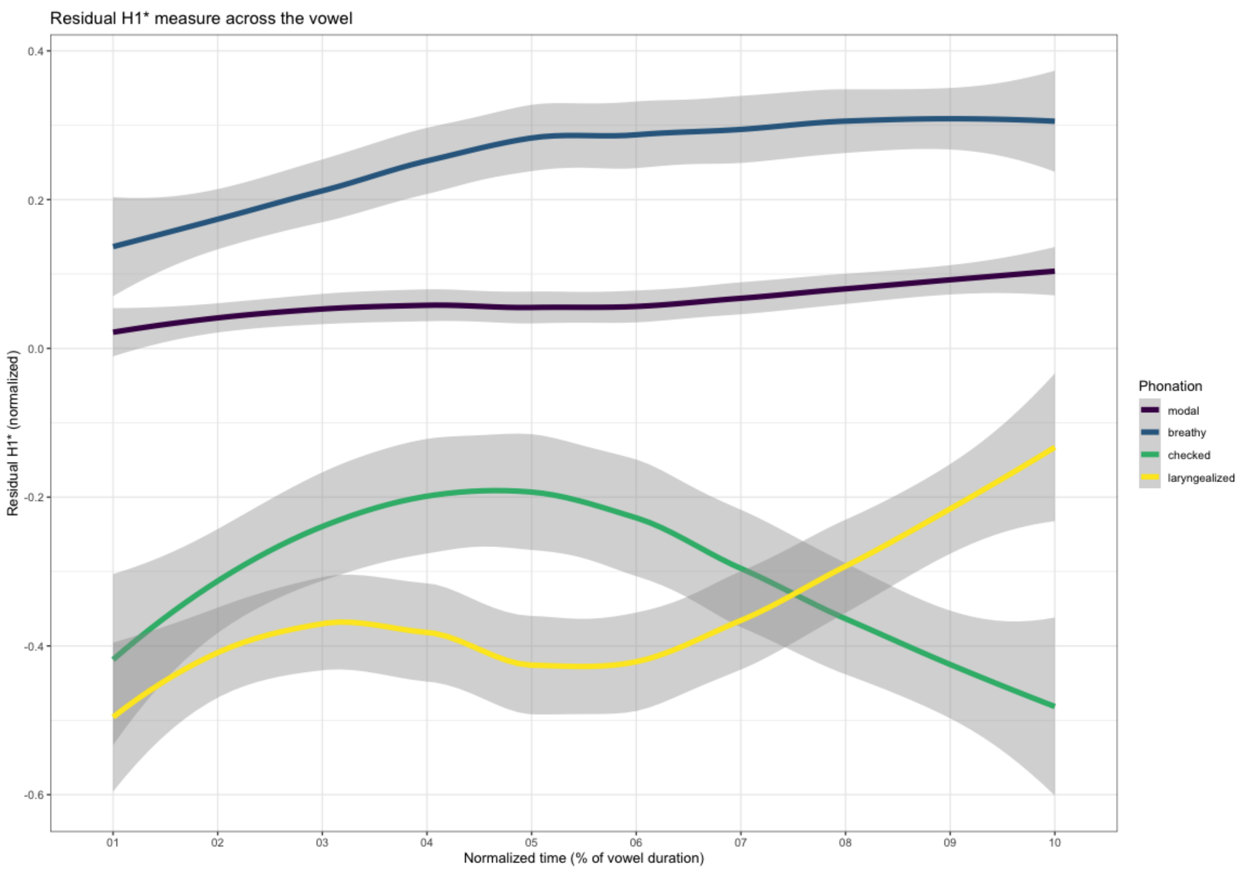

Figure 6: <caption>

Table 1: Results of the linear mixed-effects model for z-scored H1*-H2*.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	0.02876	0.09336	6.17877	0.30809	0.76814	
Breathy	-0.03569	0.04210	9749.84999	-0.84781	0.39656	
Checked	-0.14120	0.04050	14402.97700	-3.48623	<0.001	***
Rearticulated	-0.13719	0.04964	5353.41205	-2.76340	0.00574	**
Position 2	0.00475	0.01339	19027.80227	0.35456	0.72293	
Position 3	-0.01795	0.01259	19039.57085	-1.42613	0.15385	
Tone H	0.01597	0.04454	4340.72994	0.35855	0.71995	
Tone L	-0.07794	0.03218	8176.59285	-2.42248	0.01544	*
Tone M	-0.10568	0.04397	7428.42296	-2.40365	0.01626	*
Tone R	0.28710	0.06918	10218.05414	4.15025	<0.001	***
Breathy:Position 2	0.02104	0.03069	19046.99970	0.68566	0.49294	
Checked:Position 2	-0.00034	0.03597	19063.65965	-0.00949	0.99242	
Rearticulated:Position 2	-0.06258	0.03226	19036.21224	-1.94021	0.05237	.
Breathy:Position 3	0.16346	0.02900	19087.82852	5.63588	<0.001	***
Checked:Position 3	0.04822	0.03400	19096.31688	1.41845	0.15608	
Rearticulated:Position 3	0.14092	0.03025	19072.44284	4.65848	<0.001	***

Table 2: Results of the linear mixed-effects model for z-scored H1*-A3.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-0.31027	0.07720	6.62348	-4.01891	0.00568	**
Breathy	0.38600	0.03945	9730.33732	9.78394	<0.001	***
Checked	-0.54747	0.03689	11945.83254	-14.84188	<0.001	***
Rearticulated	-0.09924	0.04880	5860.06520	-2.03384	0.04201	*
Position 2	-0.01965	0.01153	19006.00934	-1.70359	0.08847	.
Position 3	0.00304	0.01085	19013.63370	0.28054	0.77907	
Tone H	0.26138	0.04422	4799.12698	5.91144	<0.001	***
Tone L	0.35203	0.03053	10954.20080	11.53172	<0.001	***
Tone M	0.08136	0.04209	9223.72777	1.93308	0.05326	.
Tone R	0.73152	0.06454	14312.23532	11.33400	<0.001	***
Breathy:Position 2	-0.01424	0.02644	19019.42711	-0.53839	0.59031	
Checked:Position 2	0.02067	0.03099	19030.35345	0.66688	0.50485	
Rearticulated:Position 2	-0.04613	0.02779	19011.33138	-1.66014	0.09690	.
Breathy:Position 3	-0.07100	0.02499	19048.14423	-2.84079	0.00451	**
Checked:Position 3	-0.00026	0.02930	19051.95269	-0.00876	0.99301	
Rearticulated:Position 3	0.09593	0.02607	19036.32796	3.68044	<0.001	***

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 linear mixed-effects model for Residual H1*.

	Estimate	Std. Error	df	t value	Pr(>|t|)	
(Intercept)	-0.10483	0.04015	10.71725	-2.61084	0.02469	*
Breathy	0.14997	0.03175	5393.82513	4.72315	<0.001	***
Checked	-0.38554	0.03061	5905.98167	-12.59335	<0.001	***
Rearticulated	-0.48437	0.03740	4633.21778	-12.95239	<0.001	***
Position 2	0.01369	0.01027	19029.30204	1.33294	0.18257	
Position 3	0.04386	0.00966	19041.55355	4.54158	<0.001	***
Tone H	0.09272	0.03335	2975.18807	2.78046	0.00546	**
Tone L	0.18568	0.02439	7860.40311	7.61295	<0.001	***
Tone M	0.05475	0.03328	7005.78449	1.64525	0.09996	.
Tone R	0.33507	0.05257	9813.69066	6.37372	<0.001	***
Breathy:Position 2	0.11599	0.02354	19049.14884	4.92755	<0.001	***
Checked:Position 2	0.10060	0.02759	19066.47122	3.64602	<0.001	***
Rearticulated:Position 2	-0.02675	0.02474	19038.08812	-1.08092	0.27975	
Breathy:Position 3	0.11491	0.02225	19091.32792	5.16510	<0.001	***
Checked:Position 3	-0.10327	0.02608	19100.43388	-3.96047	<0.001	***
Rearticulated:Position 3	0.12029	0.02320	19075.58487	5.18437	<0.001	***

Table 4: Model comparison between H1*–H2* and Residual H1* in distinguishing SLZ phonation types.

Phonation contrast	Model	β	Standard error	t-value	p	
Breathy vs Modal	H1*-H2* model	-0.03569	0.04210	-0.84781	0.39656	
	Res. H1* model	0.14997	0.03175	4.72315	<0.001	***
Checked vs Modal	H1*-H2* model	-0.14120	0.04050	-3.48623	<0.001	***
	Res. H1* model	-0.38554	0.03061	-12.59335	<0.001	***
Rearticulated vs Modal	H1*-H2 model	-0.13719	0.04964	-2.76340	0.00574	**
	Res. H1* model	-0.48437	0.03740	-12.95239	<0.001	***

Table 5: LRT and AIC for the H1*-H2* and residual H1* models.

Model	Log-Likelihood Ratio	AIC	ΔAIC
H1*-H2* model	-21674	43386.99	11214.54
Residual H1* model	-16067	32172.45	0

5 Discussion

6 Conclusion

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, 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.

Ariza-García, Andrea. 2018. Phonation types and tones in Zapotec languages: A synchronic comparison. *Acta Linguistica Petropolitana* XIV(2). 485–516. https://doi.org/10.30842/alp2306573714220.

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.

Brinkerhoff, Mykel Loren, John Duff & Maya Wax Cavallaro. 2021. Downstep in Santiago Laxopa Zapotec and the prosodic typology of VSO languages. In *Manchester Phonology Meeting*. Manchester, England.

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

Casella, George & Roger L. Berger. 2002. *Statistical inference*. 2. ed. Pacific Grove, Calif: Duxbury. 660 pp.

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.

Chai, Yuan & Shihong Ye. 2022. Checked Syllables, Checked Tones, and Tone Sandhi in Xiapu Min. *Languages* 7(1). 47. https://doi.org/10.3390/languages7010047.

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.

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 & 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.

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 sourcea). *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.

Santiago Laxopa. 2022. *Santiago Laxopa: Economy, Employment, Equity, Quality of Life, Education, Health and Public Safety*. Data México. https://datamexico.org/en/profile/geo/santiago-laxopa (25 April, 2022).

Seyfarth, Scott & Marc Garellek. 2018. Plosive voicing acoustics and voice quality in Yerevan Armenian. *Journal of Phonetics* 71. 425–450. https://doi.org/10.1016/j.wocn.2018.09.001.

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.

Sichel, Ivy & Maziar Toosarvandani. 2020a. Pronouns and Attraction in Sierra Zapotec. In Andrew Hedding & Morwenna Hoeks (eds.), *Syntax and semantics at Santa Cruz, Volume IV* (Syntax & Semantics at Santa Cruz (SASC) 4). Santa Cruz, CA: Linguistics Research Center.

Sichel, Ivy & Maziar Toosarvandani. 2020b. The featural life of nominals. lingbuzz/005523.

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.

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

Appendix 1: Elicitation word list