On Residual H1 as a measure of voice quality

Mykel Loren Brinkerhoff & Grant McGuire

Abstract

text.

Keywords:

1 Introduction

2 Santiago Laxopa Zapotec

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 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).

[1] See Appendix 1 for wordlist

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 Mahalanobis distance 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 & 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

4 Results

4.1 H1*-H2*

4.2 H1*-A3

4.3 Residual H1*

5 Discussion

6 Conclusion

References

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

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.

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.

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

Appendix 1: Elicitation word list