THE EFFECT OF STRESS AND SPEECH RATE ON VOWEL QUALITY IN SPONTANEOUS CENTRAL MEXICAN SPANISH

Laura Montoya Hernandez[1], Scott James Perry[2], Benjamin V. Tucker[2,3]

[1]Universidad del Valle, [2]University of Alberta, [3]Northern Arizona University
montoya.laura@correounivalle.edu.co, sperry1@ualberta.ca, benjamin.tucker@nau.edu

ABSTRACT

This study explores the influence of stress and speech rate on vowel quality in spontaneous speech produced by 13 Central Mexican Spanish speakers. Results show that stress more consistently influences F1, and the effect of speech rate is inconsistent, affecting the quality of some vowels and not others. We also find that, for some vowels, the effect of speech rate on vowel centralization is stronger for unstressed vowels than for stressed vowels. Interpreted in the context of overall formant variability, these results show that Spanish vowels undergo phonetic, but not phonological vowel reduction. As our stress results more closely mirror studies on spontaneous Spanish rather than those conducted using laboratory speech, it may be that patterns of phonetic vowel reduction in Spanish are not consistent across speech styles.

Keywords: Spanish, vowel space, spontaneous speech, lexical stress

1. INTRODUCTION

For most varieties of Spanish there is a symmetrical 5-vowel system /i,e,a,o,u/ that is generally assumed to lack phonological vowel reduction in unstressed syllables [1]. Although lexically contrastive vowels may not neutralize in unstressed syllables, studies on many varieties of Spanish have found acoustic differences between stressed and unstressed vowels in terms of duration and vowel quality [2, 3, 4, 5, 6]. Differences in vowel duration are more consistent across the vowel space while vowel quality is sometimes reported to only differ by stress in specific vowels [4, 5]. The primary goal of the present paper is to analyze the spectral qualities of stressed and unstressed vowels in spontaneously produced, conversational speech in Mexican Spanish, building on the modest body of literature investigating phonetic variability in spontaneous Spanish. The second goal is to investigate the effect of speech rate on vowel production and explore how speech rate may or may not interact with stress.

Unstressed vowels in Spanish are shorter than stressed vowels, a robust finding across many studies and varieties [2, 3, 4, 6], even if the difference may be smaller compared to other languages [5]. We did not investigate duration in the present paper due to limitations on boundary placement during forced-alignment [7]. However, we are investigating the effect of speech rate, which we assume impacts vowel quality at least partially due to decreased vowel duration as speech rate increases. As pointed out by Lindblom [8], spectral 'targets' of vowels may not be reached at faster speech rates due to 'articulatory undershoot', as there is less time to articulate sounds. In Spanish, faster speech rates while reading have been associated with increased centralization (F1 and F2), i.e. an overall compression of the vowel space [2].

As mentioned previously, the findings concerning the effect of stress on vowel quality show less consistent patterns across different studies. Nadeu [2] found stress-driven F1 differences for /a/ and F2 differences for /i,a,o/, with stress differences for other vowels not reaching statistical significance. In contrast, Torreira & Ernestus [5] reported that the effect of stress on Spanish vowel quality was largely restricted to movement in F1, with unstressed vowels having lower F1 values. In a study including spontaneous speech from Mexican Spanish speakers [4], differences between stressed and unstressed vowels were larger for F1 than F2, although this was not modelled explicitly. The differences reported above are not surprising, as we expect differences between laboratory and spontaneous speech [7].

While previous findings are somewhat inconsistent, there are many potential reasons for these differences. Some differences are undoubtedly due to dialectal differences, which are attested [4]. Other potential sources are different elicitation methods, and researcher degrees of freedom, as similar studies have employed drastically different approaches to analyzing changes in vowel formants [5, 4]. It is also important to note that some studies claim that vowels in stressed and unstressed vowels are the same based on a non-significant result, confusing the absence of evidence with evidence of absence. As the differences in vowel duration are fairly robust, from an articulatory perspective, there is likely some difference on average between the quality of stressed and unstressed vowels [8].

2. METHOD

The data reported in this paper was recorded as part of a larger research project. Below, we only describe the methods directly relevant to the present paper, omitting other tasks completed by participants.

2.1. Speakers

The speech analyzed for the present study consists of recordings in Spanish of L1 Spanish/L2 English speakers from Querétaro, México (N=13, 9 females and 4 males).

2.2. Materials and procedure

Participants were recorded in a sound booth while speaking to a relative or friend by phone for 15 minutes wearing a head-mounted microphone. The speech produced by the participants in the booth was orthographically transcribed by a native Spanish speaker. We used these transcripts to generate a pronunciation dictionary using a function of the Montreal Forced Aligner (MFA; [9]).

The dictionary was reviewed by the native speaker and a phonetically-trained non-native speaker of Spanish to check that the pronunciations were accurate and consistent with the dialect under analysis. During the transcription process, pronunciation variants noted were added to the dictionary. We then used the MFA to align the data at the word and phone levels using only our data, as we had more than one hour of speech required to produce desirable alignment.

2.3. Measurement

In order to extract spectral information from vowels in our data, we first determined formant tracker settings for each speaker manually by looking at a small number of tokens of each vowel. Using this information, we extracted the formant measures (at multiple time-points for each vowel), duration, previous/following segments, word, and speech rate using a custom script in Praat [10]. We calculated speech rate for each utterance by taking the number of syllables divided by the duration.

As we expected a certain amount of measurement error due to both formant tracking and the forced alignment process, we elected to use a robust method of calculating multivariate outliers - the minimum covariance determinant. For each speaker, the outliers for each vowel were calculated using raw formant values at each of 30, 40, 50, 60, and 70 percent of the vowels' duration. We then calculated a multivariate normal distribution for each vowel by taking a number of subsamples of the data and determining the ellipsis that minimized the covariance structure. This prevents measurement errors from skewing the ellipsis (as would happen with Mahalanobis distance). We calculated this using `rrcov` (v 1.6-0, [11]), and vowels that fell in the 2.5% tail of the resulting chi-square distribution were removed from our analysis. This removed 22.97% of the data, leaving us with 26,292 tokens. We then removed tokens from utterances that were measured to have speech rates more than three standard deviations below the average speech rate, which left us with 25,925 vowel tokens for our analysis. The distribution of these tokens was as follows: /i/ - 10.5% (8.4% stressed), /e/ - 34.1% (21.9% stressed), /a/ - 29.3% (15.7% stressed), /o/ - 22.1% (10.8% stressed), /u/ - 3.8% (2.7% stressed).

2.4. Statistical Analyses

All statistical analyses were carried out in `R` (v 4.1.1; [12]). The formant normalization process, using the `phonTools` (v 0.2-2.1; [13]) implementation of the Nearey Intrinsic [14] vowel normalization procedure, was done to minimize variation in spectral properties between speakers due to anatomical differences. This method was chosen as it minimized within-category variation compared to other vowel normalization methods.

All models fit to address our research questions were linear mixed-effect models fit using `lme4` (v 1.1-29; [15]). The dependent variables were normalized F1 and F2 values at vowel midpoint. The independent variables were stress, vowel, and speech rate. Stress was a two-level factor (stressed vs. unstressed) treatment-coded into the regression model with 'stressed' as the reference level. Vowel was a five-level factor (a,e,i,o,u) treatment-coded with 'a' as the reference level. Speech rate was the average speech rate of the utterance being produced and was log-transformed before being centered and scaled. All models had varying intercepts for speaker, previous segment, and following segment.

The first set of models explored the effect of stress on formant values at vowel midpoint, and included interactions between stress and vowel, with speech rate as a covariate. Both the model for F1 and F2 included random slopes for stress by participant. We then further explored the effect of speech rate by including a three-way interaction between vowel, stress, and speech rate. The model for F1 included random slopes for stress and speech rate, and the model for F2 included a random slope for stress. Interactions were interpreted with the aid of the package `emmeans` (v 1.6.3; [16]).

3. RESULTS

For the models investigating stress, the interaction between stress and vowel significantly contributed to model fit, as evidenced by likelihood ratio test, for both the model predicting F1 values ($\chi^2(4) = 35.18, p < 0.0001$) and F2 values ($\chi^2(4) = 40.52, p < 0.0001$). Figure 1 plots the average formant values for each vowel by stress condition. Stress did not have a statistically significant effect on F1 of /i/, but all other vowels had significantly lower F1 values in the unstressed position: /a/ ($\hat{\beta} = -0.028, SE = 0.004, p < 0.001$), /e/ ($\hat{\beta} = -0.012, SE = 0.004, p = 0.004$), /o/ ($\hat{\beta} = -0.009, SE = 0.004, p = 0.045$), /u/ ($\hat{\beta} = -0.035, SE = 0.009, p < 0.001$). For F2 values, stress did not significantly affect the F2 for the /e, i, u/ vowels. Only the /o/ vowel ($\hat{\beta} = 0.018, SE = 0.006, p = 0.003$) and the /a/ vowel ($\hat{\beta} = 0.021, SE = 0.006, p < 0.001$) showed evidence of higher F2 values in unstressed vowels.

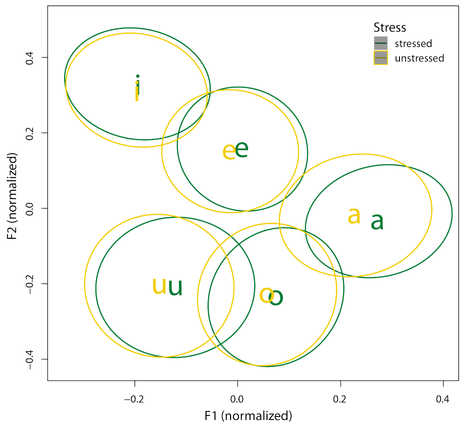

Figure 1: Normalized mean F1 and F2 values for stressed (green) and unstressed (yellow) vowels in Central Mexican Spanish. Ellipses represent one SD.

For the model which further explored the role of speech rate on vowel quality, the three-way interaction between stress, vowel, and speech rate was significant for both the F1 ($\chi^2(4) = 21.409, p < 0.001$) and F2 ($\chi^2(4) = 11.435, p < 0.05$).

First, we checked if speech rate was predictive of F1 values in the context of the vowel:stress interaction. The F1 values of the /e,i,o/ vowels were not significantly predicted by speech rate regardless of stress. However, both stressed /a/ ($\hat{\beta} = -0.016, SE = 0.003, p < 0.001$) and unstressed /a/ ($\hat{\beta} = -0.026, SE = 0.003, p < 0.001$) had significantly lower values of F1 as speech rate increased. For the /u/ vowel, increased speech rate led to higher values of F1 for stressed ($\hat{\beta} = 0.013, SE = 0.005, p = 0.023$) and unstressed ($\hat{\beta} = 0.050, SE = 0.008, p < 0.001$) vowels. The interaction between stress and speech rate reached statistical significance for both the /a/ ($\hat{\beta} = 0.010, SE = 0.003, p = 0.006$) and /u/ ($\hat{\beta} = -0.037, SE = 0.010, p < 0.001$) vowels. The three-way interaction for the model predicting normalized F1 values is visualized in Figure 2.

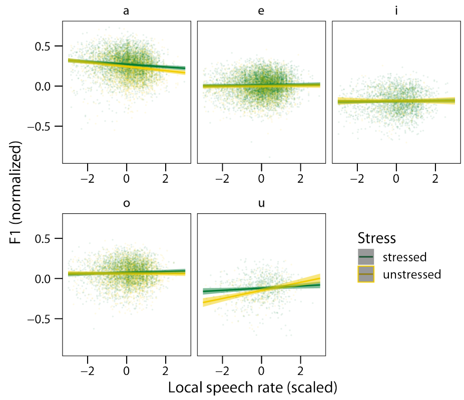

Figure 2: F1 variation of stress (green) and unstressed (yellow) vowels as speech rate increases.

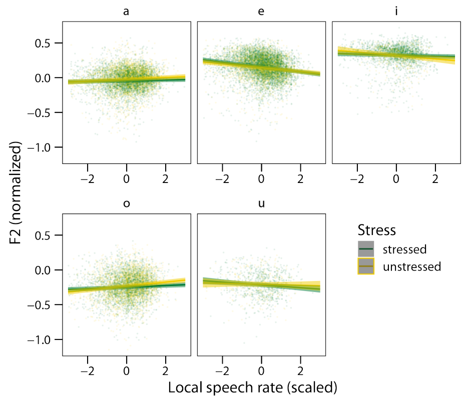

Figure 3: F2 variation of stress (green) and unstressed (yellow) vowels as speech rate increases.

For F2, we find that speech rate is predictive for stressed /e/ ($\hat{\beta} = -0.035, SE = 0.002, p < 0.001$), /i/ ($\hat{\beta} = -0.008, SE = 0.003, p = 0.0297$),

/o/ ($\hat{\beta} = 0.011, SE = 0.003, p < 0.001$), and /u/ ($\hat{\beta} = -0.020, SE = 0.006, p = 0.003$). For unstressed vowels, speech rate was predictive of F2 for /a/ ($\hat{\beta} = 0.013, SE = 0.003, p < 0.001$), /e/ ($\hat{\beta} = -0.028, SE = 0.003, p < 0.001$), /i/ ($\hat{\beta} = -0.024, SE = 0.008, p = 0.003$), and /o/ ($\hat{\beta} = 0.028, SE = 0.003, p < 0.001$). The interaction between stress and speech rate was only statistically significant for the /o/ vowel ($\hat{\beta} = -0.017, SE = 0.004, p < 0.001$). The three-way interaction from the model predicting normalized F2 values is visualized in Figure 3.

4. DISCUSSION

In the current paper, we analyzed a corpus of spontaneous Central Mexican Spanish using automated methods to investigate how stress and speech rate influence vowel quality. We began by analyzing the effect of stress conditional on each vowel to compare these results with previous studies. Our second analysis concerned whether or not the effect of speech rate was the same across stressed and unstressed vowels.

During the initial analysis of the effect of stress, we found that stress is related to changes in F1 in all vowels except for /i/, while stress was only related to increased F2 values for /o/ and /a/. This finding of stress in Spanish being more related to vowel raising rather than centralization is similar to results found for spontaneous Peninsular Spanish by [5] and Mexican Spanish by [4]. Comparing our results concerning F2 with previous research, it seems that changes in F2 by stress are inconsistent across studies, with different vowels having a significant effect of stress for F2 values. Speech rate was only related to changes in F1 values for /a/ and /u/ vowels, with faster speech rates associated with increased centralization. For these two vowels, we also found stronger effects of speech rate for unstressed vowels, which showed more raising with increased speech rate compared to stressed vowels.

For F2, speech rate either had no significant effect on formants or had the predicted effect of leading to more centralized vowels at higher speech rates. The exception is the stressed /u/ vowel, which became more peripheral at higher speech rates. While this is contrary to our predictions, a number of studies have argued that stressed syllables tend to be produced with more articulatory effort [17]. The overall change in the average F2 is small compared to the overall variation in formant values for this vowel. For most vowels, speech rate had a similar effect on F2 for stressed and unstressed vowels, with the exception of /o/, where centralization of unstressed vowels at faster speech rates increased.

While we must be cautious interpreting a statistically insignificant interaction, speech rate seems to have a relatively similar effect on vowel quality in both stressed and unstressed syllables. When there is a difference, it comes in the form of increased centralization for unstressed vowels. This is particularly interesting for the F1 of the /u/ vowel, where stress and speech rate seem to have opposing effects, with unstressed /u/ being raised in contrast to a lowering effect from increased speech rate.

In some cases, our results match previous literature [18] and [2] such as some unstressed vowels like /i,e/ tend to lower their F2 values while the rest seem to increase it. Most formant movement we observe is the lowering of F1 for all the unstressed vowels except /i/, which is the highest vowel on average. It's worth noting here that /u/ having lower average F1 values than /i/ was also reported for other Spanish varieties [2], so in a sense, /u/ had more room to raise than /i/.

We think it is important to acknowledge that we conducted an analysis of vowel quality at the vowel midpoint. We also conducted a non-linear regression on the formant trajectories between 30-70% of vowel duration and found that some vowels did show a pattern of movement over the course of the vowel. These trends were small compared to the overall variation in formant values, so we did not pursue this further due to space constraints. Future research could examine potential vowel inherent spectral changes [19] in Spanish vowels.

5. CONCLUSION

The present study has reported shifts in vowel quality according to stress and speech rate in spontaneous Spanish from Central Mexico. We find that stress has a more consistent effect on F1 than F2, echoing findings from work on spontaneous Spanish from other dialects. We also find that the effect of speech rate is inconsistent across the vowel inventory and that for some vowels there is a stronger effect of speech rate for unstressed vowels. Comparing this study with previous work on Spanish vowels, it seems that the difference between stressed and unstressed vowels may be modulated by speech style, although it is difficult to make these comparisons across different dialects. As such, future research employing a within-speaker design examining the effect of register and elicitation method would be ideal for documenting these differences in Spanish.

6. ACKNOWLEDGMENTS

The data analyzed for this study was part of National Science Foundation grant 1022313 to Natasha Warner and Miquel Simonet. We thank the Mitacs organization for supporting the first author through a Globalink Research Internship and the Social Sciences and Humanities Research Council for supporting the second author in the form of a CGS-D scholarship.

7. REFERENCES

[1] J. I. Hualde, *The Sounds of Spanish*. Cambridge, UK: Cambridge University Press, 2005.
[2] M. Nadeu, "Stress- and speech rate-induced vowel quality variation in Catalan and Spanish," *Journal of Phonetics*, vol. 46, pp. 1–22, 2014.
[3] M. Ortega-Llebaria and P. Prieto, "Acoustic correlates of stress in Central Catalan and Castilian Spanish," *Language and Speech*, vol. 54, no. 1, pp. 73–97, 2011.
[4] F. Santiago and P. Mairano, "The role of lexical stress on vowel space and duration in two varieties of Spanish," in *9th International Conference on Speech Prosody*. Poznan, Poland: ISCA, 2018, pp. 453–457.
[5] F. Torreira and M. Ernestus, "Vowel elision in casual French: the case of vowel /e/ in the word c'etait," *Journal of Phonetics*, vol. 39, pp. 50–58, 2011.
[6] A. M. Delforge, "Unstressed vowel reduction in Andean Spanish," in *Selected proceedings of the 3rd conference on laboratory approahces to Spanish phonology*, L. Colantoni and J. Steele, Eds. Somerville, MA: Cascadilla Proceedings Project, 2008, pp. 107–124.
[7] B. V. Tucker and Y. Mukai, *Spontaneous Speech*, ser. Elements in Phonetics. Cambidge University Press, 2023.
[8] B. Lindblom, "Spectrographic study of vowel reduction," *The Journal of the Acoustical Society of America*, vol. 35, no. 11, pp. 1773–1781, 1963.
[9] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, "Montreal Forced Aligner: Trainable text-speech alignment using Kaldi," in *Proceedings of the annual conference of the International Speech Communication Association, INTERSPEECH*, 2017.
[10] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," 2019. [Online]. Available: www.praat.org
[11] V. Todorov and P. Filzmoser, "An Object-Oriented Framework for Robust Multivariate Analysis," *Journal of Statistical Software*, vol. 32, no. 3, pp. 1–47, 2009.
[12] R. C. Team, "R: A language and environment for statistical computin," Vienna, Austria, 2021. [Online]. Available: http://www.R-project.org/
[13] S. Barreda, "phonTools: Functions for phonetics in R." 2015.
[14] T. M. Nearey, *Phonetic feature systems for vowels*. Indiana University Linguistics Club, 1978.
[15] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
[16] R. V. Lenth, "emmeans: Estimated marginal means, aka least-squares means," 2021.
[17] T. Gay, "Physiological and acoustic correlates of perceived stress," *Language and Speech*, vol. 21, no. 4, pp. 347–353, 1978.
[18] R. E. Ronquest, "An acoustic examination of unstressed vowel reduction in heritage Spanish," in *Selected proceedings of the 15th Hispanic linguistics symposium*. Somerville, MA: Cascadilla Proceedings Project, 2013, pp. 151–171.
[19] T. M. Nearey and P. F. Assmann, "Modeling the role of inherent spectral change in vowel identification," *The Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1297–1308, 1986.