VOWEL DURATION IN ENENLHET

Paige Erin Wheeler

The University of Texas at Austin
pewheeler@utexas.edu

ABSTRACT

This study describes vowel duration in Enenlhet (ISO 639-3 tmf; Paraguay), based on a corpus of 3.5 hours of naturalistic speech. Enenlhet is both under-documented and endangered, having only around 1200 speakers. The Enenlhet corpus was aligned using untrained forced alignment, and vowel boundaries were hand corrected. Duration data were analysed with a linear mixed-effects model which examined factors shown by cross-linguistic research to regularly affect vowel duration. Results show that vowels are longer in pre-pausal syllables, in open syllables, and before voiced consonants in closed syllables; these three factors interact. The mid vowel /e/ is shorter than both /o/ and /a/.

Keywords: vowel duration, Enenlhet (Enlhet-Enenlhet) language description, naturalistic speech

1. INTRODUCTION

Enenlhet is one of six Enlhet-Enenlhet languages, all spoken in the Gran Chaco region of Paraguay. All six languages are endangered, and the family is under-described [1]. There are no detailed linguistic studies of Enenlhet, and only two phonetic studies of related languages (Enxet and Sanapaná) [2, 3]. The Enenlhet vowel inventory contains three vowels: /a, e, o/. Though sister languages have been argued to have phonemic vowel length, Enenlhet has not been described to have it [4, 5].

This study examines Enenlhet vowel duration through a corpus of naturalistic speech data [6]. In addition to contributing to the description of Enenlhet, it also provides information about vowel duration in spontaneous, as opposed to elicited speech, which may differ [7, 8]. Factors often found to increase vowel duration are considered: utterance-final position, stress, voiced following consonants, open syllables, and low vowels.

Utterance-final lengthening is widely attested [9, 10, 11, 12], though the domain may vary (e.g., final syllable, final foot, final word [13]). Though much of the literature on stress focuses on European languages, greater duration has been found to be a robust acoustic correlate of stress [14, 15, 16, 17]. Vowels in many languages are also longer preceding a voiced consonant [18, 19] (though cf. [20]). Open-syllable lengthening is also present in many languages [21, 22], though some studies do not find longer vowels in open syllables [23]. Low vowels have also been shown to have a longer inherent duration than mid or high vowels, perhaps due to articulatory constraints [24, 25, 26].

3. METHODS

3.1. Corpus

This study uses naturalistic speech collected as a part of a project to create an Enenlhet dictionary and available in the Archive of the Indigenous Languages of Latin America [6]. Recordings (24bit/96kHz) were made with a Zoom h4n recorder and an XLR lavalier microphone. Transcription, translation, and utterance-level segmentation were done by the corpus compilers. Utterance-level segmentation was done based on syntactic criteria; most utterances were also bounded by silence.

Interviews or narratives were selected for this study. The selected recordings amounted to 3.5 hours of speech from 8 speakers (4 men, 4 women), with a total of 15,031 vowels from 2013 unique lexical items. Most words contained between one and three syllables; the longest words in the corpus had eight.

3.2. Alignment and annotation

To speed data processing, the corpus was force-aligned using the Praat [27] plugin Easyalign [28]. Since there are no trained acoustic models of Enenlhet, this force-alignment used a model for Spanish (with *seseo*). Transcriptions were adjusted to account for deletion and speech disfluencies, both of which were frequent.

Utterance boundaries were marked at all utterance boundaries in the original transcription, as well as after any vowel ending >250ms before the start of the next vowel (i.e., followed by silence). Each word was marked as either pre-pausal or non-pre-pausal. Vowels were marked for their position within a word, numbered from right word edge and then again from the left. Enclitics [29] were counted as part of the host word. Vowels were marked for the voicing of the following consonant. Syllables were marked as either open or closed using a Python [30] script which segmented consonants as onsets whenever possible (i.e., CV.CV not CVC.V).

3.3. Boundary correction and measurement

Vowel boundaries were manually adjusted based on the following criteria: following a closure, release bursts were included as part of the vowel; preceding a closure, vowel offset was placed at the end of F2; adjacent to nasals, boundaries were placed at the beginning of a sharp intensity drop/rise or, if not visible, the onset and offset of antiformants; adjacent to frication, onset/offset were marked at the end/beginning of aperiodic noise; after approximants, onset was marked at the start of the amplitude increase or, if not visible, the beginning of the F1 rise; vowel offset before approximants was placed at the start of the amplitude decrease marking the start of the constriction.

The duration of each vowel segment was extracted using a Praat script, which also extracted the word label associated with each vowel. This analysis considers only tokens transcribed as a single vowel. The original transcriptions also included some adjacent vowels not separated by glides and /VʔV/ sequences with identical vowels. In related languages, these cases have been analyzed as sequences of segments rather than single units [3, 5], and therefore they warrent a separate analysis.

3.4. Hypotheses

Because utterance-final lengthening is so widely attested, Enenlhet vowels were expected to be longer before a pause (utterance boundary) than utterance-medially. If Enenlhet has fixed stress, it was expected to be identifiable as a lengthening effect associated with a vowel's position within a word. Lexically idiosyncratic stress would be indistinguishable from phonemic length. Vowels were expected to be longer before the voiced consonants – nasals (/m, n, ŋ/), glides (/j, w/), and the lateral approximant (/l/) – than the voiceless ones (/p, t, k, q, ʔ, ɬ, s, h/). Consonant clusters are rare in Enenlhet, with most syllables being either CV or CVC. Vowels were anticipated to be longer in open (C(C)V) syllables than in closed ((C)CVC(C)) ones. Enenlhet has no high vowels, and studies of vowel quality in sister languages have found more variation in the F2 dimension compared to the F1 [2, 3]. Nevertheless, the low vowel /a/ was predicted to be longer than the mid vowels /e/ and /o/. The mid vowels were not expected to differ.

3.5. Statistical analysis

A linear mixed effects model was calculated in R [31] using the *lme4* package [32]. The model treated duration as the dependent variable and included vowel quality (/a, e, o/), following consonant voicing (voiced or voiceless), syllable structure (open or closed), word position, and utterance position of the word (pre-pausal vs. medial) as fixed effects. Lexical item and speaker were random intercepts. It also included all possible two-way interactions between the fixed effects. The baselines for each variable were: vowel quality /e/, medial word, final syllable, voiceless following consonant, closed syllable.

This model showed significant effects of word position, utterance position, and their interaction. These two effects were investigated separately by examining two overlapping subsets of the corpus: all word-final vowels, in both medial and pre-pausal words; and all medial words, with vowels in any position. The model including only medial words did not show any significant word position effects, regardless of whether word position was anchored to the left or the right word-edge, so the word position and utterance position variables were subsequently recoded and combined. Each syllable was marked as either immediately pre-pausal or non-pre-pausal, and a third model was run using this variable in place of utterance and word position separately. This model also included all possible two-way interactions between consonant voicing, syllable structure, and pre-pausal position; it used the same baseline variables as the previous one.

4. RESULTS

Table 1 shows the significant effects and their estimated intercepts from the model with word position and utterance position pooled. Effects which were not significant are not shown in the tables. All results are significant at $p < 0.0001$.

Effect	Std. error	t-value	Intercept (ms)
vowel quality: /o/	1.04	9.92	10.28
vowel quality: /a/	0.61	17.82	10.88
pre-pausal syllable	1.04	15.61	16.28
vcd following C	0.81	10.31	8.35
open syllable	0.87	6.53	5.70
vcd C * pre-pausal syll	1.40	-7.44	-10.39
open syll * pre-pausal	1.36	8.10	11.00
vcd C * open syll	1.11	-8.17	-9.12

Table 1: Significant effects and interactions from model combining word and utterance position variable (n=14,810)

Table 1 shows a significant effect of vowel quality, with /o/ and /a/ both being longer than /e/. Two-way comparisons with *emmeans* [33] indicate a significant difference between /a/ and /e/, and between /e/ and /o/, but not between /o/ and /a/.

Table 1 also shows a negative interaction between consonant voicing and pre-pausal syllable. To investigate these factors separately, an additional model containing only utterance-medial vowels in closed syllables was run; results appear in Table 2.

Effect	Std. error	t-value	Intercept (ms)
vowel quality: /o/	1.56	4.74	7.37
vowel quality: /a/	0.78	8.88	6.96
vcd following C	0.75	9.83	7.34

Table 2: Significant effects in model with medial vowels in closed syllables (n=6238)

Table 3 shows the results of a model containing only medial vowels followed by voiceless consonants, which examined the effect of syllable structure.

Effect	Std. error	t-value	Intercept (ms)
vowel quality: /o/	1.47	7.74	11.35
vowel quality: /a/	0.87	14.31	12.48
open syllable	0.86	6.15	5.32

Table 3: Significant effects in model with only medial vowels followed by voiceless consonants (n=5494)

Finally, Table 4 shows the results of a model containing only vowels in closed syllables followed by voiceless consonants, both medial and pre-pausal.

Effect	Std. error	t-value	Intercept (ms)
vowel quality: /o/	2.54	5.62	14.30
vowel quality: /a/	1.75	8.23	14.42
pre-pausal syllable	1.44	19.82	28.44

Table 4: Significant effects in model containing closed syllables with voiceless consonant codas (n=3522)

Tables 2, 3, and 4 show significant effects of voiced following consonants, open syllables, and pre-pausal position, respectively. To investigate the interaction between these three factors, another pair of models were run, one including only utterance-medial syllables, and one including only voiceless consonants. Table 5 shows the results for the model containing only utterance-medial vowels.

Effect	Std. error	t-value	Intercept (ms)
vowel quality: /o/	1.01	10.66	10.80
vowel quality: /a/	0.58	17.41	10.14
vcd following C	0.75	10.26	7.69
open syllable	0.81	6.11	4.97
vcd C * open syll	1.09	-6.77	-7.36

Table 5: Significant effects and interactions from model containing utterance-medial vowels (n=11,661)

The significant negative interaction between consonant voicing and syllable structure shown in Table 5 suggests that the lengthening effect due to voiced following consonants is smaller in open syllables than in closed ones. Table 1 also includes a negative interaction between consonant voicing and utterance position, which suggests that the lengthening associated with voiced following consonants is limited in pre-pausal position.

Table 6 shows the results from the model containing only voiceless following consonants, which investigated the effect of syllable structure and its interaction with utterance position.

Effect	Std. error	t-value	Intercept (ms)
vowel quality: /o/	1.51	5.87	8.84
vowel quality: /a/	0.94	12.93	12.20
pre-pausal syllable	1.23	12.50	14.14
open syllable	1.01	4.48	4.52
pre-pausal * open syll	1.78	8.17	14.52

Table 6: Significant effects and interactions from model containing voiceless following consonants (n=7717)

Table 6 shows a significant positive interaction between pre-pausal position and syllable structure, suggesting more lengthening in pre-pausal open syllables than medial ones.

5. DISCUSSION AND CONCLUSION

5.1. Pre-pausal lengthening

The largest effect in Table 1 is that of pre-pausal position. Pre-pausal vowels are, as expected, longer than vowels not adjacent to the pause. This effect is the only positional effect observed in the corpus; there is not an independent effect of a vowel's position within a word.

These results are congruent with previous work on the effect of utterance position on duration, though the effect in Enenlhet is smaller than has been reported for other languages (e.g., [9, 10]). Unlike studies which find word-final lengthening as well as utterance-final lengthening [13] or progressive lengthening over the last several syllables in an utterance [12], utterance-final lengthening in Enenlhet is restricted to the final vowel in the utterance-final word.

5.2. Stress

There is no independent word position effect in the corpus, and therefore no evidence for fixed stress. Future research is needed to determine if Enenlhet lacks word-level stress entirely. Alternatively, stress may be lexically specified, variable based on word class, or some morphemes may fall outside the stress domain and therefore obscure stress in this analysis.

5.3. Consonant voicing

The effect of consonant voicing is consistent with results from other languages. Enenlhet vowels are significantly longer before voiced consonants. The negative interaction between syllable structure and consonant voicing suggests that this effect is primarily limited to closed syllables, when the following consonant forms a coda.

Consistent with studies that find that other duration effects are not strictly additive [25, 35], the effect of coda consonant voicing is restricted in pre-pausal syllables. This finding suggests a potential global limit to the amount of utterance-final lengthening which is permitted.

Because Enenlhet does not make a voicing distinction in the obstruent class, voicing also splits the obstruents from the sonorants. However, there may be differences in consonant behaviour dependent on manner of articulation (e.g., nasals vs. approximants or stops vs. fricatives) akin to findings from other languages [25, 34].

5.4. Syllable structure

As shown in Table 3, there is a significant effect of syllable structure. As expected based on other languages, vowels in open syllables are longer than those in closed syllables. In addition to the relationship between syllable structure and consonant voicing, Table 1 shows an interaction between syllable structure and utterance position. In contrast to the effect of the voicing of the following consonant, the interaction between syllable structure and utterance position in Table 6 shows that open syllable lengthening is greater in final syllables than medial ones.

5.5. Vowel quality

Vowel quality has an unexpected effect on vowel duration. Pairwise comparison confirms that /e/ is significantly shorter than /a/, as expected based on the shorter inherent durations of mid vowels than low vowels. However, /e/ and /o/ are also significantly different, and /o/ and /a/ are not.

The estimated differences between /e/ and /o, a/ are much smaller than has been reported previously [24, 25], amounting to only about 8% of the mean duration of /e/. Even so, this finding suggests that /e/ is higher than /o/ and that /o/ and /a/ have about the same height. However, the average formant values for each speaker, shown in Figure 1, suggest that each vowel occupies a different area of the acoustic space.

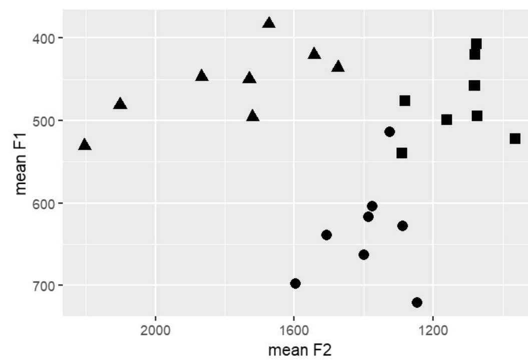

Figure 1: Mean F1 and F2 values for /a, e, o/ for each speaker (triangles: /e/, squares: /o/, circles: /a/)

For most speakers, the F1 values for /e/ and /o/ fall in the range of 400–500 Hz, and for all but one speaker, the F1 values for /a/ (600–700Hz) are higher than the means for the mid vowels. These mean F1 and F2 frequencies suggest that the vowel height cannot fully explain the duration difference between /e/ and /o/ and the similarity between /o/ and /a/.

5.5. Conclusion

This study describes vowel duration in Enenlhet, finding that vowels are longer utterance-finally, in open syllables, and before voiced consonants. The syllable structure effect is enhanced utterance-finally, while the effect of voiced consonants is limited. Lengthening due to voiced following consonants is greater in closed syllables than in open ones. Compared to other languages, the effects in Enenlhet are quite small, though still significant. Utterance-final lengthening, open syllable lengthening, and lengthening preceding voiced consonants are also found in unrelated languages, suggesting that they are relatively universal phenomena, albeit with language-specific implementations.

Vowel quality has an unexpected relationship to duration: /e/ is shorter than /a/ and /o/. Further study of the relationship between quality and duration is needed, as vowel height alone does not appear to explain this difference.

This study used a corpus of naturalistic speech, which is underutilized in phonetic research. Future work comparing these results with those from experimental work will be fruitful in gaining a greater understanding of how running speech differs from careful production.

6. ACKNOWLEDGEMENTS

I would like to thank Scott Myers, Pattie Epps, Elizabeth Wood, Raina Heaton, and the members of the UT Austin research group in phonetics and phonology; their valuable comments have substantially improved this analysis. This research was supported by the National Science Foundation (Doctoral Dissertation Research Improvement Grant No. 2024000). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect views of the National Science Foundation. All remaining errors are my own.

7. REFERENCES

[1] Enenlhet. 2023. *Catalogue of Endangered Languages*. University of Hawai'i at Manoa. https://endangeredlanguages.com/lang/2141

[2] Elliott, J. 2016. For bilinguals, Enxet vowel spaces smaller than Spanish. *J. Acoust. Soc. Am*. 140. 3107.

[3] van Gysel, J. 2022. The influence of language shift on Sanapaná vowels: An exemplar based perspective. *Lang. Var. and Change* 34. 1–29.

[4] Unruh, E. & Kalisch, H. 2003. Enlhet-Enenlhet: Una familia lingüística chaqueña. *Thule* 14/15, 207–231.

[5] Elliott, J. 2021. *A grammar of Enxet Sur*, Manoa.

[6] Heaton, R. 2019. Enenlhet Documentation. *The Archive of the Indigenous Languages of Latin America.* https://ailla.utexas.org. PID: ailla:266555.

[7] DiCanio, C., Whalen, D. 2015. The interaction of vowel length and speech style in an Arapaho speech corpous. *Proc. 18th ICPhS* Glasgow.

[8] DiCanio, C., Nam, H., Amith, J., Castillo García, R., Whalen, D. 2015. Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec. *J. Phon*. 48, 45–59.

[9] Oller, D. 1973. The effect of position in utterance on speech segment duration in English. *J. Acoust. Soc. Am.* 54. 1235–1247.

[10] Berkovits, R. 1994. Durational effects in final lengthening, gapping, and contrastive stress. *Lang. and Speech* 37. 237–250.

[11] Hockey, B. Faygal, Z. 1999. Phonemic length and pre-boundary lengthening: An experimental investigation of the use of durational cues in Hungarian. *Proc. 14th ICPhS* San Francisco. 313–316.

[12] Nakai, S., Kunnari, S., Turk, A., Suomi, K., Ylitalo, R. 2009. Utterance-final lengthening and quantity in Northern Finnish. *J. Phon.* 37. 29–45.

[13] Beckman, M., Edwards, J. 1990. Lengthenings and shortenings and the nature of prosodic constituency. *Between the grammar and physics of speech.* Cambridge.

[14] Fry, D. 1955. Duration and intensity as physical correlates of linguistic stress. *J. Acoust. Soc. Am*. 27. 765–768.

[15] Lindblom, B., Rapp, K. 1973. *Some temporal regularities of spoken Swedish*, Stockholm

[16] Ortega-Llebaria, M., Prieto, P. 2011. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Lang. and Speech* 54. 73–97.

[17] Gordon, M., Roettger, T. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Lin. Vanguard.* https://doi.org/10.1515/lingvan-2017-0007

[18] Fintoft, K. 1961. The duration of some Norwegian speech sounds. *Phonetica* 7. 19–39.

[19] Chen, M. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22. 129–159.

[20] Mitleb, F. 1984. Voicing effect on vowel duration is not an absolute universal. *J. Phon.* 12. 23–7.

[21] Benguerel, A. 1971. Duration of French vowels in unemphatic stress. *Lang. and Speech* 14. 383–391.

[22] Maddieson, I. 1985. Phonetic cues to syllabification. In: Fromkin, V. (ed.) *Phonetic linguistics: Essays in honor of Peter Ladefoged*, Orlando. 203–221.

[23] Lippus, P., Asu, E., Teras, P., Tuisk, T. 2013. Quantity-related variation of duration, pitch, and vowel quality in spontaneous Estonian. *J. Phon.* 41. 17–28.

[24] Klatt, D. 1975. Vowel duration is syntactically determined in connected discourse. *J. Phon.* 3. 129–40.

[25] O'Shaughnessy, D. 1981. A study of French vowel and consonant durations. *J. Phon*. 9. 385–406.

[26] Esposito, A. 2001. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica.* 59. 197–231.

[27] Boersma, P., Weenink, D. 2019. *Praat: Doing phonetics by computer* [computer program]. http://praat.org. Version 6.04.49.

[28] Goldman, J. 2011. Easyalign: An automatic phonetic alignment tool under Praat [Praat plugin]. *Proc. Interspeech* Firenze. http://latcui.unige.ch/phonetique/easyalign.php. Version 04.2012.

[29] van Gysel, J. 2017. *Temporal predicative particles in Sanapaná and the Enlhet-Enenlhet language family (Paraguay): A descriptive and comparative study,* Leiden.

[30] van Rossum, G., Drake, F. 1995. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

[31] R core team. 2019. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing.

[32] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *J. Statistical Software* 67. 1–48.

[33] Lenth, R. V. 2022. *Estimated marginal means, aka least-squares means.* R package version 1.8.2. https://CRAN.R-project.org/package=emmeans.

[34] Begus, G. 2017. Effects of ejective stops on preceding vowel duration. *J. Acoust. Soc. Am*. 142. 2168–2184.

[35] Klatt, D. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *J. Acoust. Soc. Am.* 59.5. 1208–21. https://doi.org/10.1121/1.380986.