The Dissertation Committee for Paige Erin Wheeler Certifies that this is the approved version of the following Dissertation:

Phonetic Corpus Studies of Enenlhet Vowels: Quality, Duration, and Phonation

Committee:

Scott Myers, Co-Supervisor

Patience Epps, Co-Supervisor

Rajka Smiljanic

Anthony C. Woodbury

John Elliott

Phonetic Corpus Studies of Enenlhet Vowels: Quality, Duration, and Phonation

by

Paige Erin Wheeler

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2024

Dedication

For Willow and Oz, who saw me through my first two degrees.

Acknowledgements

This dissertation would not exist without the help of many people, and it won't be possible to name them all here. First and foremost, thank you to all the Enenlhet and Angaité people who have contributed their voices to my work and welcomed me into their homes; this dissertation doesn't do justice to the incredible complexity of their languages. I'm exceedingly grateful also to Raina Heaton and Manolo Romero for their careful work on the Enenlhet Documentation collection. The members of the UT Austin phonetics and phonology seminar have seen many drafts of this document, and their careful feedback and comments have helped to wrangle the project into its current state. I am especially grateful to Pattie and Scott, who guided me through many iterations of this project during the uncertainty of COVID and helped me develop its final shape, and to John, Rajka, and Tony, who all read drafts of this dissertation in very short time frames.

Further support for this project has come in many forms. Ben, Erica, and Erika provided critical administrative and logistical support and made sure that everyone was paid, insured, and enrolled as needed. I am exceptionally grateful for Katie, Emily, and Ishan's assistance with data processing, which has saved me hundreds of hours of labor. I am also grateful to the various health professionals who have listened, counseled, and treated me through the stress of this degree. Phil Keoghan and the contestants of the *Amazing Race* deserve a special shout-out for soundtracking all the statistical analysis presented in this dissertation.

I'm lucky to have gone through this program with an excellent cohort – Laura, May, Dag, Katya, and Scott all deserve special thanks for their willingness to commiserate, celebrate, and demonstrate paths through (and out of) academia. I'm also *so* thankful for Kithri/Oihana/Fnilla/Thundertread/Azi/Lenn, Astra/Rúmil/Kaz/Zaidi/Leri, Namee/Cleo, Maeve/Angus, Sariel/Bortle, Dove, and Carrion/Fnipper/Bark for the many hours they have spent playing alongside me. Thanks also to the possums for their enablement, support, and cheerleading; I'm glad to be a part of the nicest place on the internet with you. Caitlin, Toli, Maggie, and Ella have also helped buoy me through this degree with (sometimes very sporadic, oops) phone calls, pet pictures, and encouragement from afar.

Finally, I'm indebted to Sierra, who talked me through numerous crises, and my parents, who at this point have listened to hundreds of hours of monologues about this project. Maggie and Jasmine have provided an invaluable sounding board and a well of unconditional support. And finally, I could not have completed this dissertation without the support of Mackenzie and Lizzy, who – through quarantine, breakdowns, losses, travel, and fieldwork – have kept me functional and mostly sane. I'm very grateful and very lucky.

Abstract

Phonetic Corpus Studies of Enenlhet Vowels: Quality, Duration, and Phonation

Paige Erin Wheeler, PhD

The University of Texas at Austin, 2024

Supervisors: Scott Myers, Patience Epps

This dissertation presents three phonetic studies of vowels in Enenlhet, the first of their kind for the language. These studies draw on the naturalistic monologues in the Enenlhet Documentation corpus. The studies aim to acoustically characterize features of Enenlhet's cross-linguistically unusual vowel inventory, which contains only three phonemic vowel qualities /a, e, o/. These studies contribute to the description of Enenlhet and to the growing body of phonetic research on Enlhet-Enenlhet languages more generally, opening the door to phonetic comparison, perception studies, and historical study of the language family.

Each of these studies focuses on a specific feature of the Enenlhet vowel inventory: vowel quality (Chapter 2), vowel duration (Chapter 3), and voice quality (Chapter 4). Chapter 2 finds greater spread in the F2 dimension compared to the F1 dimension. The F1 dimension in Enenlhet does not use formant values associated with phonemic high vowels in other languages. Chapter 3 finds small changes in duration associated with immediately pre-pausal syllables, following voiced consonants, and open syllables. It does not find evidence for fixed stress or phonemic vowel length. Chapter 4 finds that vowels with

adjacent glottal stops have a lower HNR, intensity, and H1-H2 and higher F0 than vowels adjacent to other segments. This study also finds acoustic evidence of glottalization in word-initial, onsetless syllables and in immediately pre-pausal syllables.

In addition to providing a detailed phonetic exploration of these features, these studies contribute to the field of phonetics and phonology more broadly. Together, they paint a picture of a phonology that does not prioritize distinctiveness or salience of vowels. Such a system is cross-linguistically atypical, raising questions for speech perception research. In addition to questions related to speech perception, these studies showcase a broad range of variability, highlighting the importance of corpus studies for phonetic research. I argue that both experimental studies and corpus work are critical for robust phonetic descriptions.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

South America is home to some 108 language families – recently estimated to include 53 with at least two members and 55 isolates – amounting to around 25% of the world's linguistic genetic diversity (Campbell 2012a). This massive diversity is still generally underdocumented and underdescribed, though recent work (e.g., Epps & Michael 2023) has made great strides toward accessible, high quality information. Phonetic studies of South American languages are even more rare, owing in part to the high quality of data typically sought for phonetic study and the time required to process data.

This dissertation compiles three studies, each of which is a detailed phonetic exploration of a feature of the vowel inventory of Enenlhet (Enlhet-Enenlhet, [tmf]): vowel quality, vowel duration, and voice quality. As there are no previous phonetic studies of Enenlhet, these descriptions vastly expand the close-grained phonetic information available about the language. Furthermore, these studies use a corpus of naturalistic speech, which is therefore highly ecologically valid; the differences between elicited and spontaneous speech are well-attested (see, e.g., Chelliah 2001, Chellia & de Reuse 2010). The extant corpora for the Enlhet-Enenlhet family were curated with an eye toward language documentation, especially of naturalistic speech. Therefore, the use of spontaneous speech maximalizes this work's ability to act as a basis for comparison in future phonetic studies of Enlhet-Enenlhet languages.

These three studies raise many questions for future research, both on Enenlhet and related languages and for phonetics and phonology more generally. The vowel quality study (Chapter 2) suggests that Enenlhet vowels are not maximally dispersed in the F1 x F2 space, challenging cross-linguistic generalizations about how vowel systems tend to pattern and raising questions for future perception and production work in the language.

The vowel duration study (Chapter 3) details synchronic lengthening processes which may represent starting points for the development of phonemic length in sister languages Enlhet and Enxet. The effects found in the duration study are cross-linguistically typical, but small effect sizes raise questions about potential differences in these effects between speech registers and genres. The voice quality study (Chapter 4) is a first step toward describing phonation in Enenlhet more broadly. As discussed at various points throughout this dissertation, Enenlhet apparently does not prioritize strong phonation, suggesting many avenues for future research both in this language family and for speech perception work more broadly.

The remainder of this chapter provides some background on Enenlhet and the Enlhet-Enenlhet language family; describes my methods for data processing and annotation, which all three studies have in common; and discusses the limitations of this methodology.

1. Language Context

Enenlhet (sometimes called Toba-Enenlhet or Toba-Maskoy) is one of six Enlhet-Enenlhet languages, all of which are spoken in Western Paraguay, in the South American Gran Chaco region (Fabre 2013). The language family is described as comprising one or more dialect continuua. From roughly west to east the languages are: Enlhet, Enxet, Angaité, Sanapaná, Guaná, and Enenlhet (Unruh & Kalisch 2003: 3–4). This language family was previously called Maskoy, or Lengua-Maskoy, with *Lengua* often used as a generic term for any of the varieties in the family. Nowadays, Enlhet-Enenlhet, a combination of the words meaning 'person, man' in these six languages, is used for the family (Unruh & Kalisch 2003: 2).

Like most Indigenous communities in lowland South America, the archaeological record of Enlhet-Enenlhet speakers is sparse, and little documentation exists from the early colonial period (Carvalho 1992: 457). What information does exist suggests that Enlhet-Enenlhet speakers were historically divided into many smaller groups, which may have been relatively fluid. However, during colonization, especially beginning in the late 1800s, the organization and location of the Enlhet-Enenlhet people was massively disrupted due to commercial ranching operations in their historic territory (Elliott 2021: 14–15). There was a rupture in inter-generational transmission of the language in many communities at this time, especially those which had been settled around Puerto Casado (Manelis Klein & Stark 1977). Though groups since then have once again dispersed and settled into communities in small portions of their ancestral lands, much dialect diversity was lost, and group identification, which had always been relatively fluid, became more homogeneous. Now, the major linguistic boundaries generally correspond to the major ethnic identities of speakers, though each language/ethnic category also comprises several smaller historical units with their own linguistic differences (Unruh & Kalisch 2003: 7).

Enenlhet's vitality falls in the middle of the spectrum formed by the six Enlhet-Enenlhet languages. There are approximately 1,200 speakers of Enenlhet (*Enenlhet* 2022), and according to Paraguayan census data from 2012, the ethnic population is around 2,000 people, mostly residing in the three Chaco departments: Presidente Hayes, Boquerón, and Alto Paraguay (Cartes Jara, Molinas Vega, Barrios Kück, & Barrios Sosa 2015). Inter-generational transmission continues, and the community has had some success in teaching the language in schools and developing pedagogical materials (e.g., Unruh, Kalisch, & Romero 2003). In contrast, Angaité has fewer than 800 speakers, with a substantially larger ethnic population.[1] Angaité is highly endangered, as there has been a complete rupture of inter-generational transmission and most speakers are quite elderly. Enlhet and Enxet both have larger ethnic populations and more vital language transmission. The census reports that about 50% of Enxet people still speak the language, though Elliott (2021: 22–23) notes that in many communities that percentage is much higher, and the *Catalogue of Endangered Languages* records 7,500 speakers of Enlhet, out of an ethnic population of 8,100 (*Enlhet* 2022).

Linguistic scholarship on Enenlhet is relatively sparse, but this dissertation relies heavily upon two main sources. Unruh, Kalisch, and Romero (2003), a pedagogical grammar, is currently the most complete description of the language. Heaton (2019–) is a documentary corpus providing recorded narratives; interviews; and elicited lists of flora, fauna, and traditional medicines. Many of these recordings also include orthographic transcriptions and translations to Spanish, which are time-aligned at the utterance level. Heaton's analysis of the language based on the documentary corpus is ongoing, and this project relies on her data and time-aligned transcriptions. The nonprofit group *Nengvaanemkeskama Nempayvaam Enlhet,* formerly led by Hannes Kalisch, has compiled a number of monolingual text collections in Enenlhet, but as these do not provide translations I do not rely on them here.[2]

The Enlhet-Enenlhet languages are quite similar, both lexically and morphologically, and contact between them is also constant. Many people speak multiple

[1] The *Catalogue of Endangered Languages* suggests an ethnic population of 3,694 based on census data from 2002. The 2012 census records a population of 5,992. This apparent population growth is unlikely to correspond to increasing numbers of speakers, though, as the vast majority of speakers in the 2012 census report speaking primarily Guaraní.

[2] The work of *Nengvaanemkeskama Nempayvaam Enlhet* is all available at https://enlhet.org/analysis.html. This site includes the full text of Unruh, Kalisch, & Romero (2003) as well as the text collections in Enenlhet and the works by Unruh, Kalisch and co-writers that focus on other Enlhet-Enenlhet languages.

Enlhet-Enenlhet languages, marry speakers of related languages, and interact with speakers of sister languages frequently enough to maintain at least passive bilingualism in them. Language-specific differences are attested, but the basic facts of each language’s structure (e.g., word order, morphological type, basic morpheme classes) and much of the lexicon are similar across all six languages (Unruh & Kalisch 2003: 2). Because of this contact and the grammatical similarities within the family, linguistic scholarship on related languages is an informative starting point for study of Enenlhet. Therefore, recent work from Elliott (2016, 2021), van Gysel (2017, 2022), and Wheeler (2020) on sister languages Enxet, Sanapaná, and Angaité, respectively, guide my hypotheses about Enenlhet.

2. Phonology and Morphosyntax

This section describes the aspects of Enenlhet phonology which are relevant to the phonetic studies in Chapters 2 through 4 and briefly sketches some basic facts about Enenlhet morphology and syntax.

2.1. Enenlhet Phonology

Enenlhet has a relatively small phonemic inventory. According to Unruh, Kalisch, and Romero (2003), the language contains fifteen consonants, as shown in Table 1.1. Where they differ from the IPA representation, orthographic representations appear in <> in the chart.

Table 1.1: Consonant phonemes of Enenlhet

MANNER	LABIAL	ALVEOLAR	PALATAL	VELAR	UVULAR	GLOTTAL
PLOSIVE	p	t		k (g)	q	ʔ <’>
NASAL	m	n		ŋ <ng>		
FRICATIVE		s				h
LAT. FRICATIVE		ɬ <lh>				
APPROXIMANT	w <v>	l	j <i>			

Unruh, Kalisch, and Romero (2003: 300) explicitly state that their orthography is meant to represent each Enenlhet phoneme with a separate grapheme. The IPA values of most of these graphemes are inferred via their comparisons to Spanish and Paraguayan Guaraní. They state that <p, t, k, m, n, s, l> are equivalent to the corresponding letters in Spanish, and <ng, h, ’> are used as in Guaraní. The IPA values for these Guaraní sounds are described in terms of articulatory features by Meliá, Farré, and Pérez (1995) as [ŋ, h, ʔ], respectively. Unruh, Kalisch, and Romero (2003: 301) compare <i> to the final sound in the Spanish words *rey* [rej] and *soy* [soj], so I treat it as [j]. They give examples in Enenlhet of words containing the remaining Enenlhet letters: *enenlhet* ‘man’ for <lh>, *havok* ‘brother’ for <v>, *paga* ‘mosquito’ for <g>, and *iaqtepa* ‘pumpkin, gourd’ for <q>. Based on the pronunciation of the cognate words in Angaité (see Wheeler 2020) and my (brief) pilot fieldwork with Enenlhet speakers, I infer these to be [ɬ, w, ɡ, q], respectively.

Phoneme /ɡ/ appears in parentheses in Table 1.1, as it appears in very few tokens. In fact, the example word *paga* ‘mosquito’, is the only token in Unruh, Kalisch, and Romero (2003) which contains /ɡ/; /ɡ/ does not appear in the corpus used for this study. The uvular stop /q/ is also very infrequent; it appears in 11 distinct lexical items in the corpus, always in coda position following /a/ or /o/. I found [q] to be similarly infrequent in Angaité (Wheeler 2020).

Like all Enlhet-Enenlhet languages, the vowel inventory of Enenlhet contains just three phonemes. Unruh, Kalisch, and Romero (2003) describe Enenlhet <a, e, o> as similar to the equivalent graphemes in Spanish and Guaraní. Elliott (2021) provides phonetic data and minimal pairs showing these three contrasts (as well as a phonemic length contrast) in Enxet (e.g., [**a**ɬog] ‘I will go’ vs. [**e**ɬog] ‘He will go’; [ha:p**e**ʔ] ‘It is soft’ vs. [ha:p**o**ʔ] ‘white egret'; [negmom**o**] 'collect habitually' vs. [negmom**a**] 'have, grab').[3] Likewise, Wheeler (2020) provides a brief phonetic analysis of Angaité vowels showing three distinct clusters in the F1 x F2 space which correspond to what I transcribe as /a, e, o/. Elliott (2016) and van Gysel (2022) present phonetic studies of vowel quality in Enxet and Sanapaná, respectively; they find roughly similar F1 and F2 values in both languages. Elliott (2016) shows that Enxet-Spanish bilinguals use similar F1 ranges for Enxet /e, o/ and the Spanish mid vowels. These categories occupy F1 ranges which are higher than what is used for Spanish /i, u/. Given these phonetic studies of related languages, vowel quality in Enenlhet is expected to be similar; Chapter 2 discusses vowel quality and the factors which influence it in more detail.

The Heaton (2019–) corpus does not contain perfect minimal pairs for all vowel qualities. However, there are some near-minimal pairs and minimal environments which suggest that the proposed three phonemic qualities are contrastive, at least in some positions. See (1) – (3) for examples of the relevant contrasts. [4]

(1)	a.	***e****valhok*	***a****valhok*	/e/ ~ /a/
		/**e**wałok/	/**a**wałok/	
		'be happy (1SG)'	'be happy (NONFIRST)'	

[3] It's worth noting that there are no three-way minimal pairs for these vowel contrasts in Enxet, per Elliott's (2021) description, and these vowel qualities have historically shifted very rapidly. E.g., the *a*- irrealis prefix (as in *alog* vs. *elog* above), was *o*- a few generations ago (Elliott 2024, p.c.).

[4] Examples are presented with Enlhet-Enenlhet orthography on the first line (italicized) and a proposed phonemic transcription on the second line (regular type IPA). Proposed phonemic transcriptions are based on descriptions of the Enlhet orthography (Unruh & Kalisch 1999) and the Enenlhet orthography (Unruh, Kalisch, & Romero 2003). Abbreviations are adapted from the original sources; an abbreviation list appears in Appendix A. The original language of the translation appears first in the free translation line; where necessary, translations to English were done by me. Question marks in the free translation line indicate my uncertainty about the translation, because there are no Enenlhet resources with morpheme-by-morpheme translations. When drawn from a source other than Heaton (2019–), citations are right aligned.

	b.	*tai**e**p* /taj**e**p/ ‘allá / over there’	*i**a**pa’* /j**a**paʔ/ ‘armadillo sp.’	
	c.	*malh**e**k* /mał**e**k/ ‘pozo / well (n.)’	*malh**a**k* /mał**a**k/ ‘already (?)’	
(2)	a.	*aktem**a*** /aktem**a/** ‘forma / type’	*aktem**o*** /aktem**o/** ‘because (?)’	/a/ ~ /o/
	b.	*malh**a**k* /mał**a**k/ ‘already (?)’	*evalh**o**k* /ewał**o**k/ ‘be happy (1SG)’	
	c.	*talh**a*** /tał**a/** ‘fire’	*elialh**o*** /eljał**o/** ‘my sisters’	
(3)	a.	*enneng**ko’o*** /enneŋk**oʔo/** 1PL	*hengk**e’e*** /heŋk**eʔe/** ‘queda, deja, aquí / stay, here’	/o/ ~ /e/
	b.	*hank**o**k* /hank**o**k/ ‘tener [años] / be aged’	*aptavk**e**k* /aptawk**e**k/ ‘eat (NONFIRST.MASC)’	
	c.	*k**o**lhve* /k**o**łwe/ ‘now’	*akn**o**lhkek* /akn**o**łkek/ ‘be here (NONFIRST.FEM)’	

As (3a) indicates, the orthographic vowel qualities in the VʔV sequences are also contrastive. The /a/ ~ /o/ contrast appears in some common lexical items, such as *ma’a* (some kind of past TAM marker) versus *mo’ok* ‘other’. The /e/ ~ /a/ contrast is particularly evident in the first ~ non-first-person markers, which surface in (1a). Chapter 2 investigates the vowel quality contrasts in more detail by examining F1 and F2 values of these categories.

Enenlhet has not been described to have phonemic long vowels. However, the history of vowel length contrasts in the family is somewhat contested – Unruh and Kalisch (2003) treat it as a feature of proto-Enlhet-Enenlhet, while Elliott (2021) treats it as an innovation shared between Enlhet and Enxet. These analyses agree that Enenlhet does not make a phonemic length contrast, but they make different predictions about Sanapaná and Angaité. In Sanapaná, vowel length is linked to syllable structure, with phonetically long vowels appearing only in open syllables (van Gysel p.c. 2021). Angaité, however, may have phonemic long vowels; Wheeler (2020) notes some cases of phonetically long vowels that resulted in native speaker corrections if they were produced with a shorter duration, but I did not do a comprehensive analysis. Enenlhet, as the only language in the family which has not been described as having a phonemic length contrast, provides a starting point for investigating the history of those contrasts in its sisters and may help make more robust predictions about what to expect in the contested cases, especially Angaité.

No previous work on Enlhet-Enenlhet languages makes robust generalizations about stress. For Enlhet, Powys (1929: ii) makes a general statement that stress tends to be penultimate. Wheeler (2020) observes some kind of prominence in final syllables in Angaité, though I remain ambivalent about whether to call this prominence lexical stress. Elliott (2021) provides a more robust starting point in Enxet. He measures F0 and intensity, primarily in disyllabic nouns, and finds that these two correlates of prominence do not always appear together in the same syllable, and their location varies based on syllable structure and morpheme class. Unruh, Kalisch, and Romero (2003) do not discuss stress in Enenlhet, but as work on related languages has provided some qualitative descriptions of word level prominence, lexical stress may be present in Enenlhet. Chapter 3 investigates vowel duration, aiming to confirm that Enenlhet does not have phonemic long vowels and to investigate the possibility of fixed lexical stress.

There are also no descriptions of Enenlhet syllable structure. However, the phonotactics of the language as seen in Unruh, Kalisch, and Romero (2003) provide a reasonable starting point. They provide examples of both vowel-initial and consonant-initial words as well as vowel- and consonant-final words. Words can begin with any consonant except /ʔ/, and all three vowels appear word initially. Example (4) shows words beginning with vowels, other sonorants, and obstruents.

(4) a. *engvahek*
/eŋwahek/
'nariz / nose' Unruh, Kalisch, & Romero (2003: 214)

b. *angken*
/aŋken/
'tu madre / your mother' Unruh, Kalisch, & Romero (2003: 214)

c. *oka*
/oka/
'solo, pero / only, but (?)'

d. *kakha*
/kakha/
'matá / kill (IMP)' Unruh, Kalisch, & Romero (2003: 112)

e. *tape'e*
/tapeʔe/
'chicken'

f. *meme*
/meme/
'mamá / mom'

g. *ngkelvet'a*
/ŋkelwetʔa/
'vieron, vio / saw (NONFIRST)' Unruh, Kalisch, & Romero (2003: 158)

h. *sosekhe*
/sosekhe/
'mañana / morning' Unruh, Kalisch, & Romero (2003: 136)

i. *lheia*
/ɬeja/
'vos, ella / NONFIRST.FEM.PRO' Unruh, Kalisch, & Romero (2003: 108)

j. *iemmen*
/jemmen/
'water'

Words beginning with /o/ are rare; /oka/ is the only example in my corpus. I also do not have any examples of words beginning in /ʔ/. In Enlhet, syllable onsets are required, and orthographically vowel-initial words in that language have an underlying glottal stop onset (Unruh & Kalisch 1999: 5); since the Enenlhet orthography is based on the Enlhet one, it may be that the lack of word-initial /ʔ/ is therefore an orthographic choice rather than a phonotactic restriction.

Words can also end with any vowel and any consonant except /h/. See (5) for some examples.

(5) a. *aiongkomelh*
/ajoŋkomeɬ/
'agua salada / saltwater'

b. *nenekev*
/nenekew/
'Laguna Porã (place name)'

c. *ko'o*
/koʔo/
1SG.PRO

d. *tenoq*
/tenoq/
'cat'

e. *semheng*
/semheŋ/
'dog'

f. *peia'*
/pejaʔ/
'batatas / sweet potatoes'

There are never more than two adjacent consonants in my dataset. Consonant clusters can appear medially or finally, but not word initially. Final clusters always consist of a glide followed by a glottal stop. Example (6) shows some final clusters from my corpus; refer to (5) for word-medial clusters.

(6) a. *nenekev'*
/nenekewʔ/
'Laguna Porã (place name)'

b. *hai'*
/hajʔ/
'good, okay'

Word medially, /ɡ/ is not attested in consonant clusters. Most other combinations are possible, with the caveat that /ŋ/ is never the second sound in a word-medial consonant cluster, and /ʔ/ is never the first. See Appendix B for examples of consonant clusters compiled from Unruh, Kalisch, and Romero (2003). None of the examples in Unruh, Kalisch, and Romero (2003) show adjacent vowels, but Heaton's (2019–) transcriptions do have some rare cases of adjacent vowels. Without any firm evidence to the contrary, I assume that only vowels can be syllable nuclei.

Based on this distribution, I propose that the Enenlhet syllable inventory is (C)V(C). Word-finally, an additional, presumably extra-syllabic, consonant can appear. CCV sequences do not appear in my corpus, but Unruh, Kalisch, and Romero (2003) show words beginning with CCV.[5] Other Enlhet-Enenlhet languages have similar syllable inventories. As noted above, Enlhet has only CV(C) syllables. For Angaité, Wheeler (2020) suggests (C)V(C)(C) syllables. I found that complex codas are very rare (only one case of a [jk] coda in my Angaité dataset), and onsetless syllables only occur word-medially, since I treated creaky voice on all orthographically vowel-initial words as being indicative of a [ʔ] onset. I did not find examples of complex onsets in Angaité. Elliott (2021: 89, 150) notes a strong preference for closed syllables in Enxet, describing morphophonological processes involving [ʔ] that maintain or produce syllable codas whenever possible, though open syllables do appear.

As hinted at by the discussion of syllable structure, glottal stop in Enlhet-Enenlhet languages behaves differently from other consonants. Elliott (2021) includes an extended discussion of glottal stop in Enxet, showing that it participates in several unusual morphophonological processes, including surfacing as [k], [ŋ], [j], and [w], depending on the surrounding segmental context. He also notes a series of processes where a Vʔ rhyme is added to verbal constructions to avoid sequences that would result in surface deletion of semantically critical consonants (Elliott 2021: 101–102). Wheeler (2020) describes a highly variable realization of glottal stop in Angaité, ranging from aperiodic voicing of the adjacent vowels to decreased intensity and pitch changes. I also noted some alternations

[5] Many of Unruh, Kalisch, & Romero's (2003) examples of word-initial consonant clusters are related to forms from Heaton (2019–) which have an initial vowel. E.g., *ptengiak* [pteŋjak] in Unruh, Kalisch, & Romero (2003: 76) vs. *aptengiak* [apteŋjak] in Heaton (2019–). Probably, the initial consonant clusters in Unruh, Kalisch, & Romero (2003) refer to a more surface-level transcription, as word-initial vowels are often deleted.

with other consonants, most often /k/ (e.g., *tase'* [taseʔ] ~ *tasek* [tasek], 'good'). In Enenlhet, also, the distribution of glottal stop is different from other consonants. It is the only consonant which does not appear word-initially and the only one which participates in word-final complex codas. Word-medially, it is the only consonant (except /g/, which is marginal) that cannot begin a consonant cluster.

In Enenlhet, the most common location for /ʔ/ is between two vowels of identical qualities (henceforth: VʔV). Unruh, Kalisch, and Romero (2003), based on their orthographic choices, implicitly treat sequences of a glottal stop and a vowel (Vʔ, ʔV, VʔV) as sequences of two/three phonemes. However, the VʔV cases have been analyzed in at least two ways in other languages. Like Unruh, Kalisch, and Romero (2003), Elliott (2021) treats these cases in Enxet as a sequence of two identical vowels separated by a phonemic glottal stop. In the case of /aʔa/, the glottal stop is often deleted, resulting in a phonetic long vowel (Elliott 2021: 97). Gomes (2013: 107), in contrast, describes these instances in Sanapaná as cases of epenthesis, where glottal stop is inserted to avoid long vowels created by suffixation. I am skeptical of this analysis, as he also includes some monomorphemic tokens as examples of this phenomenon, such as *ko'o* /koʔo/ 1SG.PRO (Gomes 2013: 107). In these instances it is not clear whether the glottal stop should be seen as underlying, epenthetic, or as a suprasegmental feature of some kind. Wheeler (2020) describes some cases of long vowels in Angaité which appear to have two distinct portions – a modal voiced first part followed by a creaky voiced second part. At least some of the cases identified are words whose cognate in Enenlhet has a sequence of VʔV, so I suspect they are related phenomena. Given the unusual distribution of /ʔ/ in Enenlhet, and the uncertain analysis of the VʔV cases, Chapter 4 focuses on voice quality in vowels adjacent to glottal stop, with a particular focus on the timing of voice quality changes in Vʔ and ʔV cases compared to the VʔV ones.

In addition to the unique distribution of glottal stop compared to other consonants, Enenlhet exhibits a wide range of other (apparently) suprasegmental features which all broadly fall under the umbrella of “phonation”. Suprasegmental prosody has not been clearly described in any Enlhet-Enenlhet language, so what follows is a relatively qualitative description based my observation of the Heaton (2019–) corpus.

Final syllables are often devoiced or only weakly phonated. Sometimes final devoicing spreads over several segments, or even multiple syllables. For example, see Figure 1.1, which illustrates final devoicing.

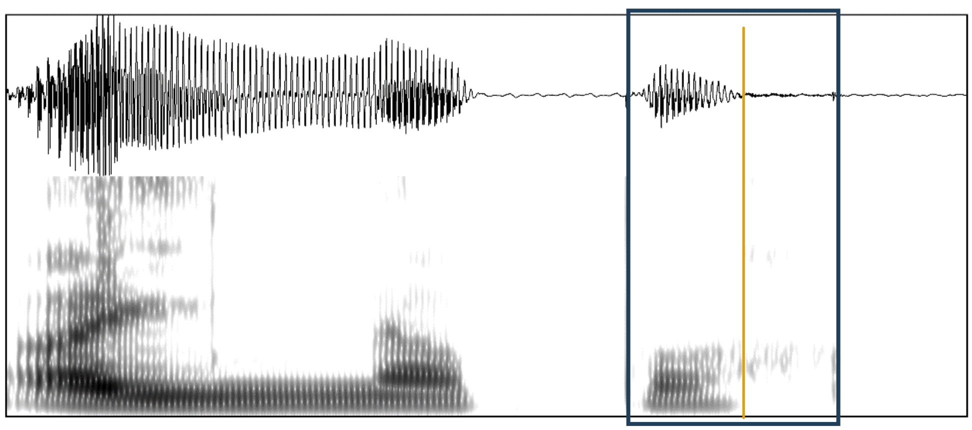

Figure 1.1. Pre-pausal *ainapa* [ajnapa], with final devoicing [ER; 92.989725s]

In Figure 1.1, the black box marks the final /a/ in the word, and the orange line marks the cessation of periodic voicing. F2 for this vowel continues very weakly for approximately the same duration as the preceding voicing.

In addition to pervasive final devoicing, many speakers in the corpus have a relatively frequent use of non-modal voicing, even in environments that do not contain a glottal segment. For example, see Figure 1.2.

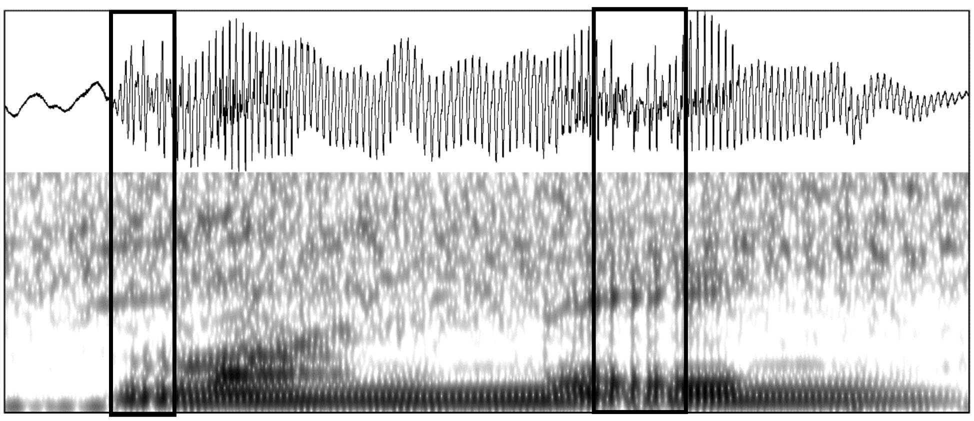

Figure 1.2. Non-modal voicing in *maneng* [maneŋ] [LF; 195.760679s]

The first box in Figure 1.2 marks non-modal phonation in the initial /m/ of this token, and the second box marks an interval of very slow, low-intensity, irregular glottal pulses during the final vowel, which has no adjacent glottal segment. Some speakers do this more than others, exhibiting what I very loosely call a more non-modal "baseline" phonation. Other speakers rarely use non-modal voicing, even adjacent to orthographic glottal stop. Chapter 4 discusses non-modal phonation in more detail.

Speakers also tend to speak very quietly in general. This impression may be due, in part, to the fact that speakers were recorded using lapel microphones rather than a headset mic, or because I did automated noise reduction on the recordings to mitigate background noise (See Section 4). Even so, intensity is quite low across the board. Highest intensities often appear after obstruents, as in Figure 1.3.

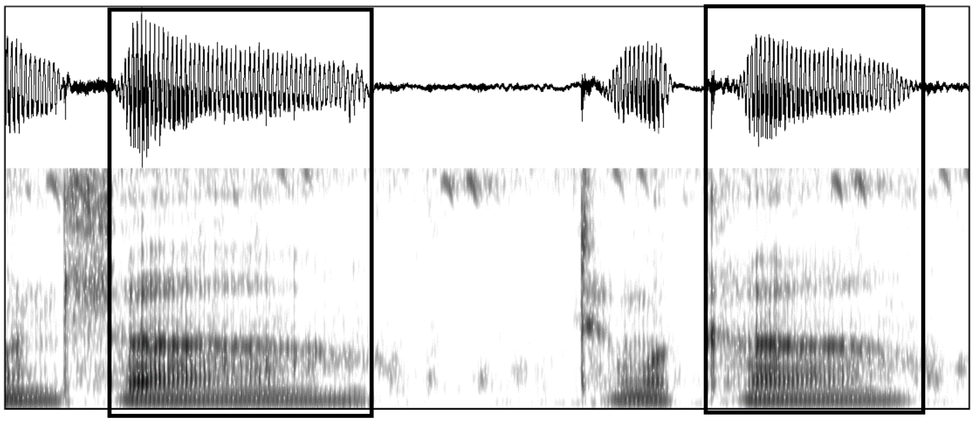

Figure 1.3. Excerpt from *vanlha temakha* [MR; 65.186340s] showing intensity peaks after obstruents

The left box in Figure 1.3 shows the final vowel in *vanlha* [wanɬa], and the box to the right indicates the final vowel of *temakha* [temakha]. In both cases, there is a sharp increase in intensity compared to the intensity of the vowel before the obstruent and then a steady decrease in intensity across the vowel.

2.2. Enenlhet Morphosyntax

Unless otherwise noted, the following information about Enenlhet morphosyntax is drawn from Unruh, Kalisch, and Romero's (2003) description. Morphology in Enlhet-Enenlhet languages is primarily suffixing. The only prefixes are person/possessive markers. Person marking prefixes occur on every verb; there are several paradigms, all of which are made up of portmanteau morphemes that simultaneously indicate person (FIRST vs NONFIRST), number (SINGULAR vs. PLURAL), and TAM categories. The language uses a direct-inverse system, as shown in Table 1.2 (Unruh, Kalisch, & Romero 2003: 283).

Table 1.2: Enenlhet person markers[6]

PERSON	FUTURE / IMPERATIVE / NEGATION	PRIMATIVE / SECONDATIVE	SUBJUNCTIVE / INFINITIVE	NOUN/ADJECTIVE
1SG.DIR	a(ŋw)-	(a)s(k)-		e-
1SG.INDIR	he(j)- me(j)- (in non-negated FUTURE only)			–
1PL.DIR	aN-	neN-		eN-
1PL.INDIR	heN- meN- (in non-negated FUTURE only)			–
NONFIRST.FEM	(ŋ)ko- / (ŋ)ka-	(a)N(k)-	a- / (a)k-	a- / (a)N-
NONFIRST.MASC	e(nj)-	(a)p(k)-		(a)p-

There are numerous allomorphs for each morpheme, and Unruh, Kalisch, and Romero (2003) do not propose underlying forms. The nasals (indicated with capital N) generally assimilate to the place of articulation of the following consonant. If there is none, they tend to surface as [n]. Initial vowels can often be deleted, except in the noun/adjective first-person forms, where the initial vowel is needed to distinguish from the NONFIRST FEMININE marker. Notably, the FIRST-PERSON INDIRECT markers are the same across all the TAM values shown in Table 1.2. In addition to occurring on verbs, these prefixes mark the possessors of possessed nouns. They also appear on nouns when acting as the primary predicate of a clause (as in copula clauses, which can consist of a single noun marked with a prefix).

Nouns do not take suffixes. Verbs, on the other hand, can be marked with a wide variety of suffixes marking subordination, associated motion (e.g., VENITIVE, APPROXIMATIVE, COMPLEXIVE), TAM categories (e.g., REPETITIVE), benefactive, and causative, among other categories. The surface allomorph of each suffix depends on the stem to which it attaches and its relationship to other clauses, resulting in large paradigms of allomorphs for each inflectional category and each verb stem. Some of these inflectional

[6] PRIMATIVE refers to when the verb appears first in the clause and SECONDATIVE is used when the verb is elsewhere (Unruh, Kalisch, & Romero 2003: 91). The markers are the same, but other syntactic alternations make these constructions distinct.

categories combine in portmanteaus which are not easily separable. For example, Unruh, Kalisch, and Romero (2003: 293, 299) show -*ekh* marking REPETITIVE in past tense subjunctive verbs, and -*eskes* marking REPETITIVE BENEFACTIVE in past tense subjunctive.

In addition to prefixes and suffixes, Enenlhet has a set of particles which both mark TAM features and emphasize the main predicate of a clause. Van Gysel (2017) calls these "temporal predicative particles" and describes them as enclitics in Sanapaná. He notes three different clitics, each with phonologically-conditioned allomorphs. In Enenlhet, the cognate forms of these clitics are =*alhta* [=aɬta] PREHODIERNAL PAST, =*lhkek* [=ɬkek] HODIERNAL PAST, and =*hata* [=hata] FUTURE. In Heaton (2019–), the most frequently-occurring form of =*alhta* is =*lhta* [=ɬta]. The most frequent form of =*lkek* in my corpus omits the final /k/. =*hata* is the least frequent, and it always appears in its full form. These particles are written as separate words in Enenlhet orthography, but often the phonetic form depends on the previous word (e.g., =*alhta* usually appears as =*lhta* when the preceding word ends with a vowel and as =*alhta* when the preceding word ends with a consonant)

Basic word order in Enlhet-Enenlhet languages is difficult to determine because the robust cross-referencing of verbal arguments on verbs means that independent noun phrases are often omitted. However, Kalisch (2019) indicates that the most basic order is VOS. Example (7) shows this order in Enlhet.

(7)	*ang-ya'pa-s-kas-kek*	*lhaak*	*sa'kok*	*meeme*
	/aŋ-jaʔpa-s-kas-kek	ɬa:k	saʔkok	me:me/
	F-bathe-CAUS-CAUS-PRIM	recent	daughter	mother

'Mamá bañó a la niña / Mother bathed the girl.'

Kalisch (2019: 144)

Alternate word orders can appear due to information structural constraints (focus, topicalization, etc.). For example, (8) shows VSO order in Enenlhet.

(8)	*ng-hana-khek*	*meme*	*seppo*
	/ŋ-hana-khek	meme	sep:o/
	F-cook-REP.REAL	mother	manioc

‘Mamá cocinó mandioca / Mother cooked manioc (again/often).’

Unruh, Kalisch, & Romero (2003: 74)

3. DISSERTATION STRUCTURE

This dissertation presents three phonetic studies of Enenlhet vowels. Chapter 2 discusses Enenlhet vowel quality. As the Enenlhet vowel phoneme inventory has been described, it is cross-linguistically unusual for two reasons. First, it is somewhat unusually small – three vowel inventories are fairly common in South America but not particularly common the world over. Campbell (2012b: 266) overviews descriptions of a number of South American languages with similar inventories (Amuesha (Arawakan) /e, o, a/, Selk’nam, Tehuelche (Chonan) /e, o, a/, Qawasar (Qawasaran) /ə, o, a/ (data from Moran, McCloy, & Wright 2019). Perhaps more importantly, there exists a widespread generalization that three-vowel inventories usually (or perhaps always) maximally disperse across the F1 x F2 space (see Vaux & Samuels 2015 for an overview). Previous phonetic studies of other Enlhet-Enenlhet languages (e.g., Elliott 2016, van Gysel 2022) have shown that Enxet and Sanapaná do not maximalize the F1 space, with the non-low vowels (/e, o/) having higher F1 values than the high vowels in Spanish, which has a five-vowel system (/i, e, a, o, u/). As Campbell (2012b) indicates, similar inventories (with no vowels occupying the lowest slice of the F1 space) have been described for other South American languages. These descriptions make empirical description of Enenlhet vowels (and other systems like them) particularly pressing, as they provide counter-examples to the cross-linguistically attested pattern.

Chapter 3 discusses vowel duration. As noted in Section 2.1, descriptions of phonemic vowel length in Enlhet-Enenlhet languages agree that Enenlhet does not have such a contrast. The first goal of this study is to provide quantitative evidence for this descriptive claim. To do this, Chapter 3 uses Linear Mixed Effects models to account for other factors affecting vowel duration and probes to see if there are systematic errors in the model for some lexical items. The second topic of Chapter 3 is lexical stress. Stress is cross-linguistically common, and it is correlated with increased duration in stressed syllables. Chapter 3 specifically focuses on a fixed stress position, identifiable as an independent effect of a vowel’s position within the word. This investigation of fixed lexical stress also provides a starting point for further investigations of Enlhet-Enenlhet prominence, as it rules out numerous possible stress positions.

Chapter 4 focuses on voice quality. South American languages famously use a variety of suprasegmental features, such as voice quality, vowel duration, tone systems, and nasalization, as contrastive features (refer, e.g., to Epps & Michael 2023 for descriptions of many languages which employ these features). As indicated in Section 2.1, Enlhet-Enenlhet languages are no exception. Glottal stop in Enenlhet has a different distribution than other obstruent consonants, and descriptions of related languages Enxet and Angaité indicate that it participates in unusual phonological processes. Furthermore, the VʔV sequences <a’a, e’e, o’o> have been analyzed in different ways. The study of voice quality focuses on characterizing the voice quality changes associated with an adjacent /ʔ/ and investigating the timing of these changes to see if the VʔV context is different from the ʔV or Vʔ cases.

All three studies rely on the same corpus of semi-spontaneous Enenlhet speech. A corpus study is particularly useful for these topics for a number of reasons. First, pragmatically, corpus data was the only data available to me when I began the project, due to the COVID-19 pandemic. The Enenlhet corpus was selected because it was the only corpus of Enlhet-Enenlhet language data whose primary curator was not already using the data for phonetic studies.

However, more importantly, corpus data is robust in a number of ways that make it richly informative on these three topics. The difference between elicited data and more naturalistic speech is well-attested (e.g., Félix-Brasdefer 2007; DiCanio et al. 2015). The Enenlhet corpus contains speakers from different ages and genders with a variety of baseline speaking rates and voice source characteristics. Vowels in this corpus appear in a wide variety of consonantal contexts, in a broad range of lexical items. The many factors that can influence vowel quality, voice quality, and duration have not been studied in Enenlhet, so the diverse contexts that appear in the corpus allow exploratory investigations that account for many of these possible influences. These corpus-based studies can help formulate more specific hypotheses for future experimental investigations. Furthermore, naturalistic speech data is the most broadly available type of data for many of the languages in the area (due to language documentation projects that focus primarily on documenting spontaneous discourse rather than controlled elicitation), so these studies will be relatively replicable for surrounding languages. Since the Gran Chaco is a region of long-standing and high intensity language contact (see, e.g., Campbell 2013), comparison between Enenlhet and other unrelated languages is an important step forward in investigating the linguistic history of the region.

4. Methods

This section describes only the methods which are relevant for all three studies presented in Chapters 2–4. It includes the details of the corpus, data preparation, data annotation, and some general notes on statistical methods.

4.1. Corpus

This dissertation uses a corpus of naturalistic Enenlhet speech. Recordings are monologic narratives or interviews which were recorded as a part of a documentary project directed by Raina Heaton (U. of Oklahoma). The recordings are from eight adult speakers of the language from around Pozo Amarillo, Paraguay. Four speakers were women and four were men. Only the male speakers stated their age on their recordings; these speakers range in age from 35 to 77.

Recordings were selected from the Heaton (2019–) archive deposit. The deposit is organized into folders, with each recording instance or theme given a separate folder. The corpus currently contains 19 narratives, as well as one conversation and two interviews about traditional food. In addition to these more naturalistic recordings, it contains elicitation sessions including terms for flora and fauna, weaving, and ethnobotany, and grammatical elicitations. The folders for these elicitation sessions also include written notes transcribing the session or presenting the elicitation prompts. Heaton also includes as a separate folder scans of her original field notes from each field visit. In addition to these more structured materials, the corpus also contains recordings of three totally unstructured events: a broadcast from a local radio station in Enenlhet, a church service in the village of Tobatí, and a wedding in Tobatí.

Each recording in the collection includes a title in Spanish and English as well as a short description of the content or context in which it was made and the speakers involved.

Many of the recordings also include time-aligned transcriptions in ELAN (ELAN 2022) which include segmentation at the utterance level. Each utterance is transcribed in Enenlhet orthography and translated into Spanish. Transcription for each speaker appears on a separate tier of the ELAN file. These transcriptions and translations were created by native speakers of Enenlhet, primarily Manolo Romero (speaker MR in the corpus), in conjunction with Heaton. The recordings that I selected for inclusion in this study are those narratives or interviews which had time-aligned transcriptions and translations at the time that the study corpus was made (approx. fall 2022). After downloading the selected recordings and their associated files from the archive, I exported the ELAN transcription files to Praat (Boersma & Weenink 2023) as TextGrid files. The runtime of the selected recordings is about 3.5 hours.

These recordings are of a quality typical to language documentation work; they were recorded with a Zoom h4n recorder at 24 bit/96kHz resolution. Speakers wore lavalier mics with an XLR connection while recording. At various points, these recordings contain wind, electrical static, cooking noise, animals, or occasional overlapping speech. I reduced background noise by selecting an ostensibly silent sample of each recording and applying the noise reduction function in Audacity (Audacity Team 2022); settings were adjusted until the selection was silent, or close to it. This process reduced constant background noise caused by factors like electrical static but did not eliminate inconsistent interruptions, such as birds or gusts of wind.

4.2. Forced Alignment

In order to extract acoustic information about the vowels, each vowel had to be identified and segmented. To speed up segmentation, a first pass segmentation was conducted using untrained forced alignment with EasyAlign (Goldman 2011), which operates as a Praat plugin. There are no forced aligners trained on Enenlhet, and training one was determined to be too time-consuming to be realistic, so EasyAlign's Spanish (with seseo) model was used instead. I selected the Spanish model because Spanish (with seseo) is the contact language used in the Enenlhet recordings and because Spanish contains most of the sounds of Enenlhet, including relatively close matches for the Enenlhet vowels. EasyAlign uses orthographic transcriptions to generate a proposed phonetic transcription, and since Enenlhet orthography is loosely based on the Spanish orthography, this match was expected to facilitate the process. The time-aligned Enenlhet transcriptions, with segmentation at the utterance level, were the input to the forced aligner, which then generated TextGrids with separate tiers for the time-aligned Spanish translation, utterance-level orthographic transcription, utterance-level phonetic transcription, individual words, syllables, and segments.

EasyAlign uses the eLite (Beaufort & Ruelle 2006) grapheme-to-phoneme module to produce a proposed phonetic transcription of each utterance in the SAMPA alphabet. This transcription is generated in a new tier of the TextGrid with boundaries that match the original utterance boundaries. As in previous studies of untrained forced alignment (e.g., Coto-Solano & Solórzano 2017; Coto-Solano, Nicholas, & Wray 2018), I manually modified this proposed phonetic transcription to facilitate accurate segmentation. Sounds which occur in Enenlhet but not in Spanish (i.e., /q, ɬ, h/) had to be substituted, because EasyAlign accepts only symbols corresponding to phones in the trained language. I made substitutions to match each Enenlhet sound with a symbol that was judged to be its closest Spanish match: /q/ > /k/, /ɬ/ > /s/, /h/ > /x/. This substitution neutralized the difference between /ɬ/ ~ /s/ and /q/ ~ /k/, but the actual identities of these consonants were retrievable from the orthography and sound files. The minimal orthographic differences between Spanish and Enenlhet were also corrected at this stage. Enenlhet <i> /j/ and <v> /w/ were interpreted as their values in Spanish orthography (i.e., /i, β/) and had to be corrected to the appropriate SAMPA symbol for the Enenlhet audio. Adjacent identical vowels (indicated with a doubled letter, e.g., <aa>) and glottal stops <’> were left in place even though neither appear in Spanish orthography.

Of the thirteen .wav files in the corpus, EasyAlign automatically created a proposed phonetic transcription of five. In the other cases, EasyAlign for unknown reasons generated the tier for the phonetic transcription with the boundaries in place to match the utterance level boundaries in the original transcription, but this new tier was entirely blank. The recordings which were not phoneticized by EasyAlign were done manually by me.

During segmentation, the aligner produces three additional tiers – word, syllable, and segment – using the HTK Toolkit. The aligner relies on the phonetic transcription tier to determine which sounds to search for in audio recording. It first estimates each word boundary and then places segment boundaries within each word. Syllable boundaries are assigned last based on segment boundaries and a set of sonority-based syllabification rules that prefer CV(C) syllables (these auto-generated syllable structures were discarded). The relevant tier for this study is the segment tier and, secondarily, the word tier. See Figure 1.4 for an example of a typical force-aligned utterance.

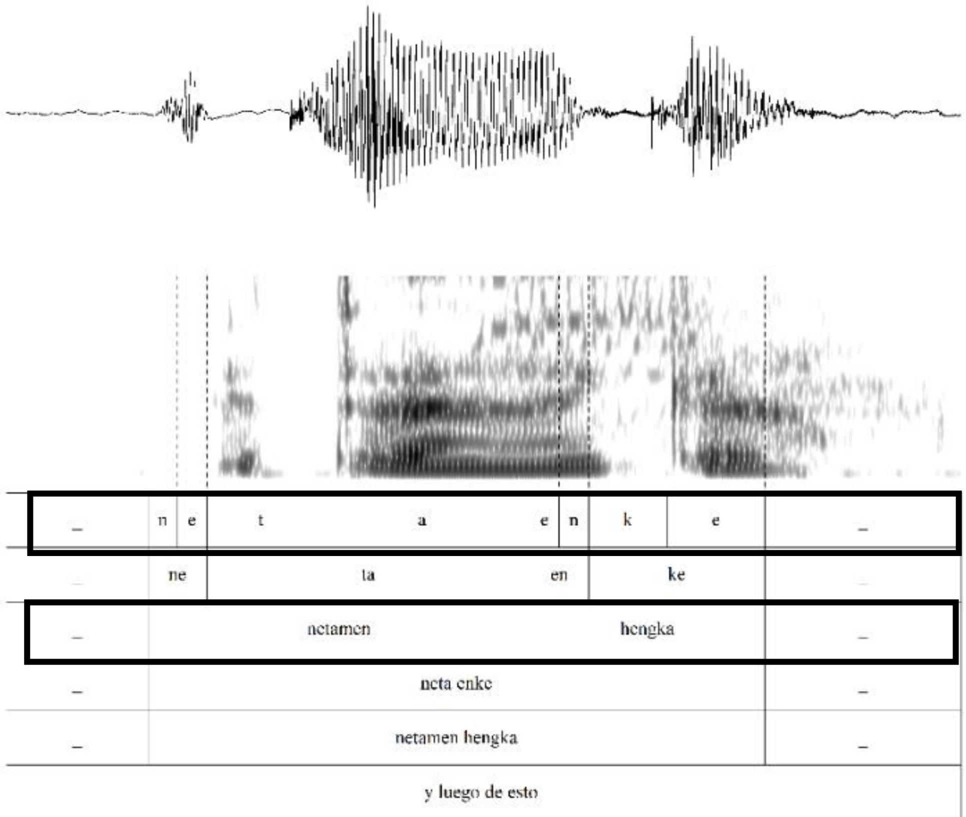

Figure 1.4. Example uncorrected forced-alignment of the utterance *netamen hengka* 'y luego esto / and then this' [CA; 1006.851s]

Figure 1.4 shows all the tiers generated by EasyAlign, from bottom: Spanish utterance translation (from Heaton 2019– translation), Enenlhet utterance orthographic transcription (from Heaton 2019– transcription), Enenlhet utterance phonetic transcription (SAMPA alphabet), word segmentation (Enenlhet orthography), syllable segmentation, phoneme-level segmentation. The black boxes mark the tiers relevant to these studies.

4. 3. Boundary Correction

Once aligned, the boundaries of each vowel were checked for accuracy. Since one of the factors of interest was duration, boundary placement needed to be both reasonable and consistent across all tokens. Qualitative examination of the automatically-produced boundaries indicated that the forced aligner performed inconsistently. Common errors included missing the beginning of voicing in an utterance; missing the end of formants in an utterance, especially with devoiced final syllables; and inconsistently marking the beginning of a vowel following a plosive. For example, see Figure 1.5, which shows two different boundary placements for a vowel adjacent to a plosive.

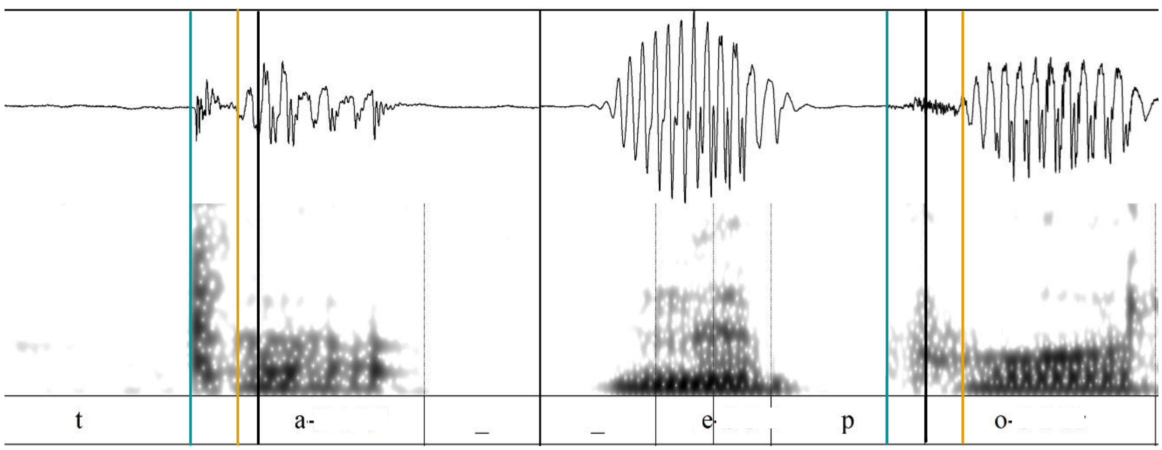

Figure 1.5. Incorrectly marked plosive boundaries [MRR_nentoma1; 2116.44 & 2740.70s]

In Figure 1.5, the automatically-placed boundaries are marked with black lines. The orange lines mark the start of periodic voicing (see waveform), and the blue lines mark the release of the stop burst (see waveform and spectrogram). On the left side of Figure 1.5, the beginning of the vowel following /t/ was automatically marked after the start of vowel voicing, which can be seen both in the voicing bar in the spectrogram and in the waveform. On the right, the boundary of the vowel after /p/ is marked in the middle of the stop burst, before periodic voicing begins.

Since EasyAlign did not place segment boundaries consistently, vowel boundaries were manually corrected. All boundary corrections were made using a Praat zoom window of one second to ensure consistency. A set of segmentation criteria were developed to account for each environment in which a vowel occurs. Enenlhet speakers often devoice pre-pausal vowels, so vowel offset at the end of an utterance was marked at the end of continuous F2 or the end of voicing, whichever was later. Utterance-initial vowels were marked as beginning at the start of periodic voicing.

After plosives, vowel onset was marked at the start of the stop burst. Since all oral stops in the corpus are voiceless, vowel offset preceding plosives was marked at the offset of voicing. Adjacent to fricatives, vowel boundaries were marked at the onset and offset of frication, which usually corresponded to the onset and offset of voicing. When it did not, aperiodicity in the waveform was prioritized unless background noise made the boundary between periodicity and frication unclear, in which case voicing was used. In some cases, a preceding /ɬ/ included a small spike in the waveform before the start of voicing, which looked similar to a closure release. Where this appeared, the vowel was marked as beginning at this spike, analogous to boundary placement after plosives.

Nasals and laterals in Enenlhet are characterized by an abrupt drop in amplitude. Nasals are additionally characterized by antiformants, and laterals, by a relatively level F2 which is broader and lower in amplitude than the vowel. Vowels after nasals/laterals were marked as beginning when the amplitude began to increase, and vowels before nasals/vowels were marked as ending when the amplitude decrease ended. When amplitude was ambiguous, antiformants or the F2 decrease were used as secondary cues to vowel boundaries; in these cases, boundaries were marked when antiformants began/ended, or when F2 stopped decreasing (/Vl/) or began increasing (/lV/). See Figure 1.6 for an example of a nasal before a vowel and Figure 1.7 for an example of laterals on either side of a vowel.

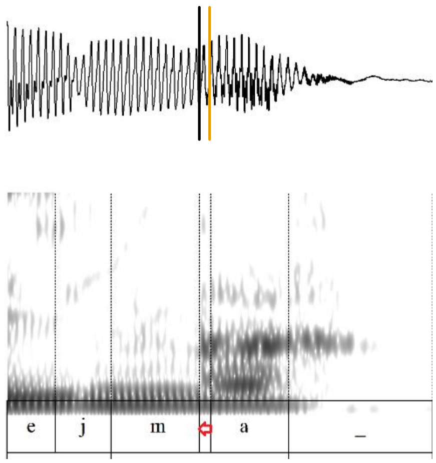

Figure 1.6. Boundary correction of prevocalic /m/ [CA; 1009.751s]

In Figure 1.6, the black line in the waveform marks the beginning of the amplitude increase, which is the location of the corrected boundary. The orange line marks the location of the automatically placed boundary, which falls two periods later.

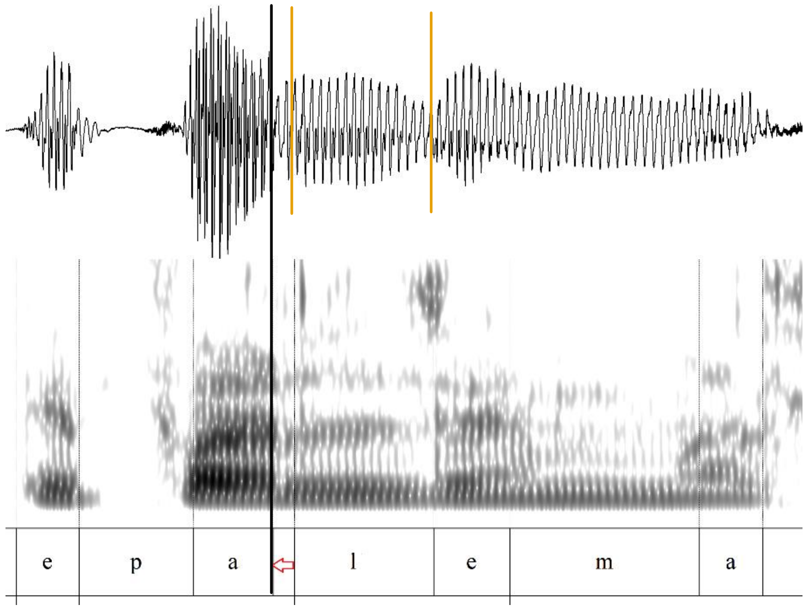

Figure 1.7. Boundary correction of postvocalic lateral; following vowel (/e/) boundary not in need of correction [CA; 1026.914s]

In Figure 1.7, the original boundaries of the /l/ are marked with orange lines. Here, the boundary of /a/ before /l/ was moved leftward to correspond to the end of the amplitude decrease into the /l/ and the abrupt decrease in F2. The new boundary is marked with the black line on both the waveform and the spectrogram. The boundary of the /l/ before /e/ was not moved; the orange line corresponds to the beginning of the amplitude increase into the vowel and the abrupt increase in F2.

Vowels adjacent to glides were marked using the same criteria as nasals and laterals. A vowel onset following a glide was marked at the beginning of the amplitude rise, or at the beginning of the F1 rise if the amplitude was unclear. Vowel offsets preceding glides were marked at the end of the decrease in amplitude.

Finally, when two vowels of different qualities were adjacent save for an orthographic glottal stop with no visible stop closure in the waveform/spectrogram, the vowels were segmented only when a dip in amplitude appeared on the waveform, and the boundary between them was placed at the lowest amplitude. When there was no dip in amplitude, the vowels were left unsegmented and subsequently eliminated from the corpus. If two vowels were adjacent with no intervening orthographic consonant, they were not segmented, and these tokens were removed.

4.4. Annotation

The utterance boundaries from the original transcription were always treated as utterance boundaries. In addition, after vowel boundaries had been corrected, any gap of greater than 250 ms between the end of one vowel and the beginning of the next was treated as an utterance boundary. A gap of 250ms was selected as a threshold large enough to avoid erroneously treating VCC.CV sequences as utterance breaks.

Though Enenlhet is not described as having phonemic vowel length, sequences of identical vowel qualities separated by a glottal stop (e.g., /oʔo/) were treated as single units rather than as two short vowels. As discussed in Section 2, Gomes (2013: 107) describes this pattern as a morphophonological process of vowel lengthening in Sanapaná. Since these cases are relatively frequent, difficult to segment, and usually morpheme internal, they were coded distinctly and analyzed as a separate category. These cases are treated more thoroughly in Chapter 4.

In some cases, actual productions appeared to vary from what was transcribed. For example, some fricatives transcribed as <lh> /ɬ/ sounded like [s], and post-vocalic nasals were often highly reduced or apparently deleted. In these cases, the orthographic transcription was used to compute syllable structure and adjacent sounds. This was done for two reasons. First, since these transcriptions were done with the assistance of native speakers, and the orthography aims for a 1:1 grapheme:phoneme correspondence, I assume that they have some kind of psychological (i.e., phonemic) reality for at least some speakers. The transcribed consonants may be relevant to acoustic productions even when not identifiable as a distinct segment (consider e.g., nasal deletion where the nasalization appears on the vowel, or realizations of glottal stop as creaky voiced vowels). Further, purely pragmatically, I am not a speaker of the language and could not consult with one, so I was not prepared to make judgments about what was "really" present when the transcription appeared to differ from what I heard.

Information about annotation categories relevant to each study is presented in the corresponding chapters. Once annotated, the interval labels for each vowel were extracted via Praat script, and syllable structure was computed using a Python script (Rossum & Drake 1995) that syllabified consonants as onsets whenever possible and avoided tautosyllabic consonant clusters.

4.5. Exclusion Criteria

During manual boundary correction, speech disfluencies and unintelligible tokens were removed. These tokens were identified based on the transcription, where Heaton provided markers of hesitation (e.g., <*kel- kelasma*> where the dash indicates a restart, or utterances marked '[unintelligible]'). In some cases, tokens were transcribed, but the recording itself was too noisy to allow for accurate segmentation, in which case they were not included.

Also excluded were Spanish borrowings or code-switches (e.g., /letse/ from Spanish *leche* 'milk', /sikleta/ from Spanish *bicicleta* 'bicycle', or /watka/ from Spanish *vaka* 'cow'). Enenlhet morphemes which were attached to borrowed morphemes were likewise excluded. Spanish loan words were excluded to avoid skewing the vowel quality and vowel duration studies. Since Spanish includes vowel qualities that Enenlhet does not (/i, u/) and nothing is known about Enenlhet stress (if present), Spanish loan words, which may or may not have been phonologically adapted to the Enenlhet vowel inventory and stress pattern(s), would have added an additional confounding factor.

Once the corpus was annotated, acoustic measures were collected using Praat scripts. The scripts also returned the word in which each individual segment appeared on the word tier.[7] The final corpus contains 15,555 vowels and 2,033 distinct lexical items. See each chapter for details on study-specific exclusion criteria and outlier identification.

The tokens are not evenly divided between speakers; Table 1.3 shows the breakdown by speaker. Table 1.4 shows the asymmetries in the vowel qualities; /a/ is by far the most frequent, followed by /e/, and then /o/, which is much lower frequency.

Table 1.3: Corpus statistics by speaker

SPEAKER	SPEAKER GENDER	SPEAKER AGE	*n* TOKENS	*n* LEXICAL ITEMS
CA	M	72	8389	1238
ER	F	unknown	905	204
LF	F	unknown	403	103
LM	F	unknown	681	157
MM	M	35	695	172
MR	M	39	506	134
MRR	M	77	3383	415
TF	F	unknown	799	211

[7] Praat script adapted by me from the sample script provided by Riggs (2016).

Table 1.4: Corpus statistics by vowel quality

VOWEL QUALITY	*n* TOKENS	*n* LEXICAL ITEMS
/a/	8067	1725
/e/	5632	1496
/o/	1854	555
TOTAL	15,555	2033[8]

4.6 Analysis

All analyses in this dissertation were done in R (R Core Team 2023). Linear Mixed Effects models were generated with the *lme4* package (Bates, Mächler, Boker & Walker 2015). Post hoc pairwise comparisons were completed with the *emmeans* package (Lenth, Bolker, Buerkner, Giné-Vázquez, Herve, Jung, Love, Miguez, Riebl & Singman 2022). In Chapter 2, robust distances were calculated with the *rrcov* package (Hardin & Rocke 2005). All data visualizations were done with the *ggplot2* package (Wickham 2016).

5. LIMITATIONS

The main limitation of the these studies is the orthographic transcription. As is noted at various points throughout this document, I relied on the provided orthographic transcription to determine word boundaries, adjacent consonants, and vowel quality. I took this approach for two reasons, one pragmatic and one principled. Pragmatically, the orthographic transcriptions were what was available at the time the studies were conducted, so I was constrained to work with them. However, I also chose to rely on the orthographic transcription instead of adjusting based on my perception because this transcription was created by a native speaker with experience in linguistic consultation. I therefore assumed

[8] Note that this is not the sum of this column, as speakers overlapped in the lexical items they produced. The total in this column represents the total number of distinct lexical items in the entire corpus, across all speakers.

that the orthographic transcription on some level represents this speaker's linguistic intuitions about the language.

Relying on the orthographic transcription does impose some limitations. The transcription represents a native speaker's judgements about the language, but native speakers often vary. We observe this variability in each of the studies included here. Therefore, the transcriptions may not accurately represent each speaker's individual variety. This is a downside that can be mitigated by close consultation with a variety of native speakers in future studies.

I relied on the word boundaries (spaces) in the orthographic transcriptions to determine word boundaries in all cases except one: the case of temporal predicative particles. Since the particles *=lhta, =hata, =lhkek* rely on the preceding word for their phonological shape, I treated them as part of the preceding word, although Enenlhet orthography writes them as separate words. However, cross-linguistically clitics often fall outside the stress domain. To mitigate this uncertainty, in Chapter 3, I ran the duration models on the dataset both with and without the words that ended in clitics.

However, even setting aside the case of clitics, determining word boundaries is a non-trivial exercise. As Tallman (2020) notes, grammatical and phonological words do not always align, and these categories are not necessarily cross-linguistically comparable. Unruh and Kalisch (1999) also discuss the difficulty in determining word boundaries for the purposes of developing an orthography of Enlhet. Tallman advocates for robustly testing constituency on a variety of levels to develop a language-specific definition of a "word". Because I was working only with orthographic transcriptions and not with native speakers, such testing was not possible in my case.

Haplology also occurred frequently in the corpus; the orthographic transcriptions sometimes include syllables which simply do not appear in the phonetic signal. In these cases, I used the orthography to identify which syllables were present, and I treated these as "the word". For example, *teiapmakha* /tejapmakha/ produced as /tejapmak/ (CA 81.865s) was treated as a three-syllable word with a final syllable /mak/. Again, I did this primarily for pragmatic reasons. I assumed that most effects related to position (e.g., final lengthening) would affect the vowel that was actually produced in that position, regardless of whether the orthography indicated that it was phonemically underlying. However, especially in the case of stress, the patterns of haplology themselves may provide valuable information about the stress domain and syllable prominence.

I also relied on the orthographic transcription to determine the adjacent consonants. This is limiting particularly in the case of glottal stop. Enenlhet glottal stops, especially VʔV intervals, have widely variable acoustic realizations. Glottal stop is acoustically variable cross-linguistically, so this is not necessarily surprising. However, since the orthographic transcription was relatively consistent across speakers (because it was primarily created by one speaker), the orthography may blur genuine inter-speaker differences if some speakers produce glottal stops in places where others do not.

It's also worth pointing out that, while in an ideal world we would have robust transcriptions which draw on the native speaker intuitions of a range of speakers, the limitations inherent in relying on the orthographic transcription of Enenlhet are the same ones that all linguists working with orthographic transcriptions must grapple with. In a corpus of English, for example, speakers are likely to map many different vowel qualities onto the same spellings. Some studies compensate for this by limiting their studies to tokens from a specific demographic, but some simply accept a high level of between-speaker variation and orthographic imprecision as intrinsic to working with speech corpora. The Enenlhet situation is a little different because orthographic choices in this language are primarily directed by one speaker, rather than conventionalized over a long period of time, and because not enough is known about language variation to fully identify the relevant factors in the corpus. However, overall, the limitations of this study are familiar to corpus studies more generally.

Chapter 2: Vowel Quality

1. INTRODUCTION

As noted in Chapter 1, Enenlhet has a vowel inventory with just three proposed phonemic vowel qualities, /a, e, o/. This inventory is the same as the one described for Sanapaná in van Gysel (2022), and the same as the one for Enxet described by Elliott (2021), except for Enxet's phonemic length distinction. Wheeler (2020) also tentatively proposes the same inventory for Angaité, modulo uncertainty about phonemic length. The goal of this study is to provide phonetic evidence which explores Unruh, Kalisch, and Romero's (2003) qualitative description of the language and to compare the formant ranges found for each Enenlhet vowel category to the ranges present in related languages and in languages with five-vowel (/i, e, a, o, u/) inventories.

Three vowels are close to the minimum number observed in the world's languages, but these systems are common in the Americas. For example, several other South American languages have been described to have four (Tacana) (Guillaume, forth.) or three vowels (Aymara, Quechua) (Adelaar & Muysken 2004). Small vowel inventories are also frequent in Australia, but are rare in Africa, Eurasia, the Pacific Islands, or New Guinea (Maddieson 2013). In contrast to their prevalence in the Americas and Australia, in the WALS sample of vowel inventories, only 93 languages (~16.5%) have inventories of 2–4 vowels (Maddieson 2013).

As noted in Chapter 1, Enenlhet (and related languages) is also unusual in the qualities included in its inventory. There is a cross-linguistic generalization that vowels spread apart within the F1 x F2 space, which predicts quality inventories like /i, u, a/ for three-vowel systems. A wide range of languages have systems like this, including Arabic, Cherokee, Greenlandic, Haida, and Quichua (Bani-Yasin & Owens 1987; Maddieson

2013; Lass 1984; Schwartz, Boë, Vallée & Abry 1997; Guion 2003, respectively). There are other languages reported to have three-vowel systems with no high vowels, including some in South America (e.g., Yanesha/Amuesha, Lass 1984), but they are much less common, and not all of these descriptions are robustly supported by phonetic evidence.

There have been several attempts to theoretically account for the cross-linguistic observation that vowels appear to prefer to be spread out in the articulatory space. Liljencrants and Lindblom's (1972) Dispersion Theory (DT) proposes that vowel qualities in an inventory of a given size will spread to as much as necessary in the F1 x F2 space to maintain the phonetic differences needed to make a phonological contrast. DT predicts that inventories with a larger number of contrastive vowel qualities will utilize a larger F1 x F2 space than one with a smaller number of categories. Subsequent studies have found some support this prediction. For example, Bradlow (1993, 1995) finds that English's vowel space, with 11 contrastive monophthongs, is larger than both Spanish and Greek, two five-vowel systems (however, cf. Disner 1983). DT was later revised to argue that vowel systems aim to maintain "sufficient" contrast between all categories, in an effort to account for the fact that the original model over-produced high vowels (Lindblom 1986). We might suppose that sufficient F1 x F2 dispersion in three vowel inventories maps to /i, u, a/, given the frequency with which it is observed.

Quantal Theory (QT) (Stevens 1972, 1989) provides another potential explanation for the prevalence of the /i, u, a/ inventory among three-vowel systems. QT proposes that the extreme points of the F1 x F2 space are areas of high stability in articulation and perception, which may account for their high cross-linguistic frequency. Given that Enenlhet contains only three vowels, QT suggests that they should be drawn to these areas of high stability. However, most studies fail to find any significant differences in acoustic variability between /i, u, a/ and other vowels (Disner 1983, Flege 1989, Bradlow 1993, 1995). Although neither DT nor QT fully account for all the observed vowel inventories of the world's languages, these theories capture the observation that we see patterns of quality dispersion and frequently recurring vowel qualities such that relatively robust predictions about which vowel qualities will be present in an inventory of a given size are possible, as demonstrated by Lindblom (1986).

As noted in Chapter 1, the difference from cross-linguistic tendencies in the Enlhet-Enenlhet languages' implementation of the three-vowel inventory appears to be a genuine phonetic fact rather than an issue of category labeling. Elliott (2016) finds that F1 and F2 of Enxet /e, o/ are comparable to F1 and F2 of Spanish /e, o/, as well as the formant values of phonemic mid vowels from other languages with 5-vowel systems in the Becker-Kristal (2010) survey. He also finds that Enxet-Spanish bilingual speakers *do* produce [i, u] tokens akin to Spanish monolinguals' productions of phonemic high vowels when speaking in Spanish and therefore provides convincing evidence that Enxet /e, o/ are more similar to Spanish /e, o/ than /i, u/. Van Gysel (2022) looks at vowel quality in Sanapaná and finds a range of formant values, with older speakers producing more peripheral tokens than younger ones. However, Sanapaná /e, o/ still are relatively lower than Spanish /i, u/ as produced by Sanapaná-Spanish bilinguals.

This study investigates the F1 x F2 space in the Enenlhet vowel inventory so that they can be compared with previous work on Enenlhet's sister languages and so that this system can be accounted for in cross-linguistic generalizations and theories about vowel inventories. To do so, I investigate midpoint F1 and F2 for each of the three phonemes while accounting for vowel duration, adjacent consonants, and surrounding vowels, which are observed cross-linguistically to affect the realization of F1 and F2.

2. METHODS

Refer to Chapter 1 to see specifics of the corpus, participants, vowel segmentation and boundary adjustment. The methods and results for F1 and F2 are presented separately. Section 2.1 discusses methods for the F1 study, and Section 2.2 presents the methods for the F2 study. The primary interest in this study is the formant ranges associated with each of the apparent Enenlhet phonemes. Therefore, all other factors in this analysis were included to rule out their effects on formants; they are really of interest only insofar as they interact with vowel quality.

2.1. F1 Annotation and Classification

2.1.1. Vowel Quality

Vowel quality was recorded for each token based on the Enenlhet orthographic transcription of the vowels, which was taken directly from Heaton (2019–). Spellings are mostly consistent across words. Since the same native speaker collaborated on developing this orthography and on transcribing these Enenlhet data, I assume that these orthographic choices represent some categorical difference. However, as noted in Chapter 1, Section 5, the orthography may not fully represent the relevant contrasts for *every* speaker. Since the range of formant values that fall within each category is the primary interest of this study, I exclusively relied on the orthographic transcription to classify vowel quality, even when the auditory impression (to me) was something different. F1 typically corresponds to vowel height, so vowels were either classified as low (/a/), or non-low (/e, o/) for the F1 study. Non-low vowels are expected to have lower F1 values than the low vowel. The VOWEL HEIGHT factor contains two levels: non-low (containing orthographic <e, e’e, o, o’o>) and low (containing orthographic <a, a’a>).

2.1.2. Duration

Previous research suggests that longer vowels will be more peripheral than shorter vowels of the same quality, perhaps because the additional duration allows the articulators to more fully achieve the specified articulatory target (e.g., Abramson & Ren 1990; Gendrot & Adda-Decker 2007; Lippus, Asu, Teras, & Tuisk 2013). In terms of formants, this generalization predicts a lower F1 for longer tokens of /e/ and /o/ and higher F1 for longer tokens of /a/. DURATION was recorded in milliseconds for each vowel, and these measures were scaled to treat the mean duration as zero; negative duration values indicate a shorter-than-average vowel, and positive duration values indicate a longer-than-average vowel.

2.1.3. Adjacent Consonants

Formant values are also expected to vary based on the adjacent consonants; much of the literature on the interaction between consonants and vowel quality shows relatively predictable patterns of place-based assimilation. The consonant inventory of Enenlhet is repeated for convenience in Table 2.1.

Table 2.1: Enenlhet consonant inventory

MANNER	LABIAL	ALVEOLAR	PALATAL	VELAR	UVULAR	GLOTTAL
PLOSIVE	p	t		k (g)	q	ʔ
NASAL	m	n		ŋ		
FRICATIVE		s				h
LAT. FRICATIVE		ɬ				
APPROXIMANT	w	l	j			

Each vowel was marked with its preceding and following consonant. The adjacent consonants were computed using a Python (Rossum & Drake 1995) script that used the Enenlhet orthographic transcription to determine the adjacent sounds. This script ignored word boundaries, so a vowel was only marked as having no preceding or following sound when it appeared utterance-initially or finally, respectively. Though I relied on the orthographic transcription for determining adjacent sounds, apparent deletions, especially of word-final syllables, occurred frequently in the corpus (see Chapter 1 for further discussion of the reasoning behind this decision and its limitations).

The primary effects on F1 were expected to be due to nasals, glides, and uvulars. Therefore, six factors were generated that marked the preceding and following consonants as either nasals (or not), glides (or not), and uvulars (or not). Table 2.2 shows these variables.

Table 2.2: Variables indicating adjacent consonants for F1 study

EFFECT	VALUES	BASELINE
PRECEDING NASAL	nasal, not nasal	not nasal
FOLLOWING NASAL	nasal, not nasal	not nasal
PRECEDING /q/	/q/, not /q/	not /q/
FOLLOWING /q/	/q/, not /q/	not /q/
PRECEDING GLIDE	glide, not glide	not glide
FOLLOWING GLIDE	glide, not glide	not glide

Adjacent to nasals, the F1 spectral peak tends to be lower in frequency and amplitude and broader compared to F1 in oral vowels (Esposito 2002). F1 is therefore predicted to be lower for vowels adjacent to nasals. This lowering results in varying perceptual effects depending on the original height of the vowel – apparently "lowering" high and mid front vowels and "raising" low vowels and mid back vowels (see, e.g., Speeter Beddor, Krakow, & Goldstein (1986) for an overview). The PRECEDING NASAL and FOLLOWING NASAL factors account for the effect of nasals on F1.

Adjacent glides were also expected to result in lower F1 values. Since glides are articulated with more constriction than vowels, they are expected to have a lower F1 (Maddieson & Emmorey 1985). Recall that increasing F1 was, in fact, used as a secondary criterion for determining the boundary between a glide and a vowel. The PRECEDING GLIDE and FOLLOWING GLIDE factors were used to account for the expected effects on F1 of glides.

Uvular sounds cross-linguistically result in vowel lowering, which translates to higher F1 values; for example, see Arabic and Quechua (Bani-Yasin & Owens 1987; Holliday & Martin 2017). This effect also appears in the conditioning of allophones of /ə/ in Kaqchikel; the lower allophone [a] appears adjacent to uvulars, compared to [ɨ] which appears with coronals (Bennett 2018). /q/ is the only uvular sound in Enenlhet; it is predicted to result in a higher F1 for adjacent non-low vowels. A separate set of two binary variables classified each preceding/following consonant as either uvular or non-uvular. The PRECEDING /q/ and FOLLOWING /q/ factors accounted for the effects of uvulars on either side of the vowel.

2.2. F2 Annotation and Classification

2.2.1. Vowel Quality

All three vowels were expected to differ in F2, which corresponds to anteriority, so the vowel quality factor used in the F2 study had three levels: front (/e/), central (/a/), and back (/o/). Front /e/ is expected to have the highest F2 values, and back /o/ is expected to have the lowest. Vowel qualities were recorded based on the orthographic transcriptions provided by Heaton. The VOWEL QUALITY factor contained three levels: /e/ (containing orthographic <e, e'e>), /o/ (containing orthographic <o, o'o>), and /a/ (containing orthographic <a, a'a>).

2.2.2. Duration

Vowel duration also affects F2 values in some languages, resulting in more peripheral productions in longer tokens (e.g., Gendrot & Adda-Decker 2007; Lippus et al.

2013). For Enenlhet, therefore, F2 is predicted to be higher for longer tokens of /e/ and lower for longer tokens of /o/. Duration was measured and scaled in the same way for the F2 study as described for F1 in Section 2.1.2, and the DURATION factor was also included in this model.

2.2.3. Surrounding Vowels

Some degree of vowel-to-vowel coarticulation is also expected. Benguerel and Cowan (1974) find that the French upper lip rounding gesture can begin up to six segments before a rounded vowel so long as the intervening segments are unspecified for roundedness (i.e., not labials or [ʃ]). Similarly, Martin and Bunnell (1982) find anticipatory coarticulation in nonce CV_1CV_2 sequences produced by English speakers, with F1 and F2 of V_1 raised before /i/ and lowered before /a/. Speetor Beddor, Harnsberger, and Lindemann (2002) report vowel-to-vowel coarticulation in Shona as well as in English.

Unlike the consonant effects, which tend to be most prominent at vowel edges, Benguerel and Cowan (1974) find that vowel-to-vowel coarticulation extends over multiple segments, suggesting that in $CV_1CV_2CV_3$ sequences, the identity of V_1 and V_3 may affect formant values of the intervening $V_{2,}$ and that effects of this coarticulation may be present at the midpoint. The primary effect of vowel-to-vowel coarticulation expected in Enenlhet is that of lip rounding, with adjacent /o/ resulting in lower F2 values for /e/ and /a/, or adjacent /e/ resulting in higher F2 values for /a/ and /o/.

Vowels in the adjacent syllables were computed with the same Python script that noted the preceding and following consonant. The script used the orthographic transcription of the word and noted the quality of the vowel in each preceding and following syllable. The VʔV sequences were expected to differ from plain Vs primarily in *voice* quality, not vowel quality, so they were treated the same as the plain Vs for the purposes of determining the quality of the adjacent vowel (/aʔa/ = /a/, /eʔe/ = /e/, /oʔo/ = /o/). Table 2.3 shows the two variables for the quality of adjacent vowels.

Table 2.3: Variables indicating adjacent vowels for F2 study

EFFECT	VALUES	BASELINE
PRECEDING VOWEL	/a/, /e/, /o/, none	/a/
FOLLOWING VOWEL	/a/, /e/, /o/, none	/a/

The PRECEDING VOWEL factor indicated the quality of the vowel in the syllable before the target, and the FOLLOWING VOWEL factor was used to account for the effect of the vowel quality in the next syllable.

2.2.4. Adjacent Consonants

Various studies also report effects on F2 due to the place of articulation of adjacent consonants. The adjacent consonants for each vowel were identified as in the F1 study, described in Section 2.1.3. Then, PRECEDING PLACE and FOLLOWING PLACE factors were generated; these factors coded each place of articulation as a separate level. If there was no adjacent consonant (utterance-initial, onsetless syllables, utterance-final open syllables), this was marked as a separate level of the factor. Since labial consonants were expected to have big effects on F2, ALVEOLAR was used as the baseline. These factors are shown in Table 2.4.

Table 2.4: Variables indicating adjacent consonants for F2 study

FACTOR	POSSIBLE VALUES	BASELINE
PRECEDING PLACE	labial, alveolar, palatal, velar, uvular, glottal, no preceding C	alveolar
FOLLOWING PLACE	labial, alveolar, palatal, velar, uvular, glottal, no following C	alveolar

Stevens and House (1963) and Hillenbrand, Clark, and Nearey (2001) find that, in American English, labial consonants result in a lower F2 in following front vowels. Stevens and House (1963) find that in American English alveolars also result in a higher F2 in following back vowels, and Hillenbrand, Clark, and Nearey (2001) find higher F2 values for back vowels after velars in English. In Turkish, Korkmaz and Boyacı (2018) also find a higher F2 for /e/ after /t/, and a higher F2 for /a/ after velars. Therefore, Enenlhet labials were expected to result in lower F2 for /e/ and perhaps /a/, /j/ was expected to result in higher F2 for /o/, and velars were expected to result in higher F2 for /o/ and maybe /a/. Since, as noted in Section 2.1.3, uvulars are sometimes associated with backing as well as lowering, /q/ was also expected to result in a lower F2 for /e/ and maybe /a/.

2.3. Measurement

F1 and F2 values were extracted from Praat for each vowel at its midpoint. The optimal formant settings for each speaker were determined by examining at least 20 tokens from each speaker and manually adjusting the formant settings until good tracking was achieved in most cases. Optimal settings for each speaker are presented in Table 2.5. Other settings for the formant object were left at Praat's default levels (time step 0.0, window length 0.025 s, pre-emphasis from 50.0 Hz).

Table 2.5: Formant settings for each speaker in the corpus

SPEAKER	GENDER	MAX RANGE	# FORMANTS
CA	M	5000 Hz	4
ER	F	5500 Hz	5
LF	F	5500 Hz	5
LM	F	5500 Hz	4
MM	M	5000 Hz	4
MR	M	5000 Hz	4
MRR	M	5000 Hz	5
TF	F	5000 Hz	4

The midpoint measurement is anticipated to be the point at which the least amount of coarticulation with adjacent consonants is present. However, because stop bursts were included in the vowel intervals, in some cases the midpoint of the vowel was placed quite close to the onset of voicing, so the measurements may reflect more coarticulation with preceding plosives than with other preceding consonants.

2.4. Exclusion Criteria

Following Hernandez, Perry, and Tucker's (2023) study of Mexican Spanish, which also used a corpus of naturalistic speech, outliers were statistically calculated using robust distances (Rousseeuw, 1985; Hardin & Rocke 2005; Hubert, Debruyne & Rousseeuw 2017). Robust distances utilize the Minimum Covariance Determiner (MCD) to estimate the mean value of a multivariate data matrix that minimizes the spread of the data. For this study, robust distances were calculated individually for each vowel quality within each speaker. The *rrcov* package (Hardin & Rocke 2005) was used to calculate the MCD of the data matrix containing just the F1 and F2 measures. Tokens outside the 97.5% confidence interval of the χ^2 test were eliminated, resulting in removing 3,021 tokens (19% of the corpus). The final corpus for analysis contained 12,663 vowels: 6,607 /a/, 4,660 /e/, and 1,396 /o/. Only two speakers had fewer than 50 tokens for any given vowel (MR had 41 /o/ tokens, and LF had 28 /o/ tokens), and for all vowels except /o/, all speakers had over 100 tokens. Therefore, though a substantial number of vowel tokens were excluded as probable tracking errors, the corpus was still judged to be robust enough to support a statistical analysis.

3. Results

Figure 2.1 shows F1 and F2 values by vowel category for all speakers. These values are normalized using the Bark scale, which is a vowel-intrinsic method of normalizing for between-speaker differences in vowel quality. Bark normalization slightly stretches differences in the F1 dimension; this distortion is consistent with human perception, which is more sensitive to differences in F1 compared to F2. The main reason for the use of Bark normalization to represent the data here is that it was also used in van Gysel (2022), so the figures in this study should be relatively comparable with that study of Enenlhet's sister Sanapaná. Normalized values are used only for visually representing the data; the statistical analysis use the raw formant values (in Hz).

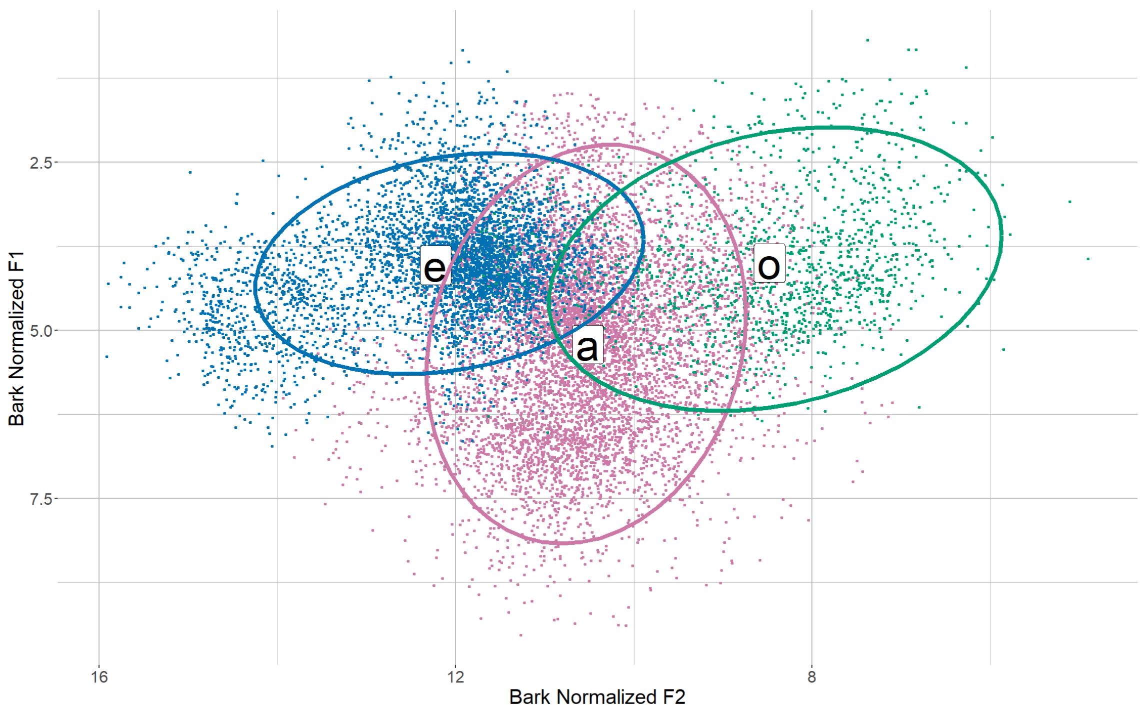

Figure 2.1. Bark-normalized F1 and F2 values by vowel category for all tokens

Figure 2.1 shows all the tokens included in this study. It is colored based on the annotated vowel quality, with ellipses approximating the 95% confidence interval around the mean formant values for each category. The labels are placed on the mean values for each category. This plot shows that each annotated vowel category occupies a somewhat distinct area of the F1 x F2 space, with substantial overlap in the center. The ellipses for /e/ and /o/ show a large spread in the F2 dimension, while the ellipse for /a/ shows a broad spread in F1. As expected, /e/ tokens take up the highest F2 values, and /o/ tokens make up the lower end of the F2 range. The bottom half of the ellipse for /a/ shows the highest F1 values; /e, o/, and /a/ all overlap in the lower F1 range.

Since the study includes only eight speakers, vowels were qualitatively examined separately for each speaker. Figure 2.2. shows the vowels, still Bark-normalized, by speaker.

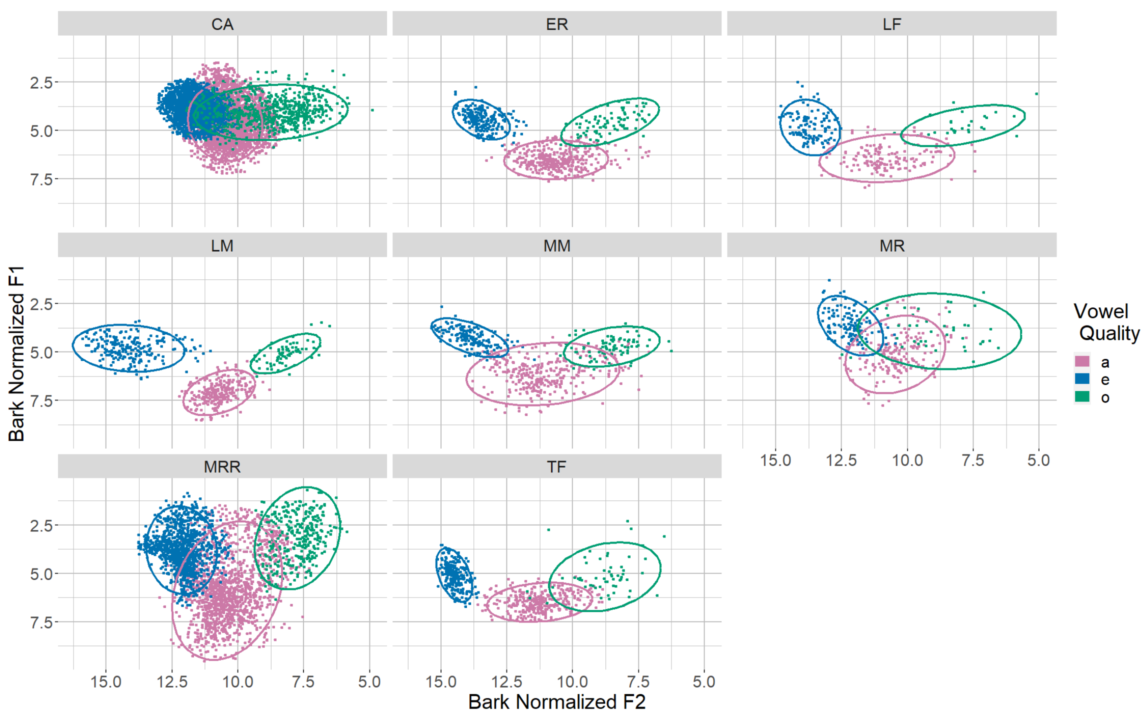

Figure 2.2. Bark-normalized vowels divided by speaker

Most speakers show the same trends seen in Figure 2.1, with the vowels spread primarily across the F2 space, without much variation in height. However, MRR appears to have a greater F1 range, though there is still substantial overlap between /e, o/, and /a/ in the lower end of this range. Speakers LM, MM, TF, ER, and LF have relatively lower /e/ and /o/ tokens, centering around or below 5.0 Bark on the F1 dimension. CA, MRR, and MR have slightly higher vowels. Speaker CA also shows a much more compact inventory than other speakers, with lower F2 values for /e/ and much less separation between /e/ and /a/ in this dimension. This difference is particularly useful to keep in mind, as CA accounts for roughly half the tokens in the corpus.

Note that the speakers who show the densest, and most overlapping vowel qualities in Figure 2.2 are those who account for the largest portion of the corpus (CA and MRR). It may be that the other speakers look more dispersed in the F2 dimension and less dispersed in the F1 dimension because they have fewer overall tokens. With more exemplars from the other six speakers, the inventories might look more similar. It could also be that the category boundaries present in MRR's speech (the speaker who assisted with the transcriptions) are not the same as for other speakers, and that some of the variability is due to the transcriber's bias. More controlled, balanced samples are needed to investigate the former possibility, and more consultation (and perception studies) are needed to address the latter.

3.1. Model Construction

Linear Mixed Effects models for F1 and F2 were built using a stepwise procedure to determine which of the proposed fixed effects are relevant to Enenlhet formant measures. The alpha level was set to 0.01 to minimize the chance of false positive results in the highly variable dataset that resulted due to the nature of the corpus.

3.1.1. F1 Model

The initial model for F1 included just a random intercept for SPEAKER and a fixed effect of VOWEL HEIGHT.[9] As expected, this model indicates that /e/ and /o/ are significantly different from /a/ (baseline). To determine which other factors and interactions to include in the model, additional preliminary models were constructed. Each preliminary model contained a random intercept for SPEAKER, a fixed effect of VOWEL HEIGHT, a fixed effect for the other possible factor, and an interaction between VOWEL HEIGHT and that factor. If there was a significant main effect or interaction, that factor was then included in the maximal model. Interactions were not included in the maximal model unless they were significant in the preliminary models.

For example, the preliminary model for duration included VOWEL HEIGHT and VOWEL DURATION as fixed effects and a two-way interaction between VOWEL DURATION*VOWEL HEIGHT. The main effect and the interaction were significant, so both were included in the overall model. The preliminary model for adjacent glides contained VOWEL HEIGHT, PRECEDING GLIDE, and FOLLOWING GLIDE as fixed effects, and two-way interactions between VOWEL HEIGHT*PRECEDING GLIDE and VOWEL HEIGHT* FOLLOWING GLIDE. The main effects were significant but the interactions were not, so only the main effects were included in the overall model.

The only factor which did not show a significant main effect or interaction was the PRECEDING UVULAR factor. Therefore, the overall model for F1 contained random intercepts for SPEAKER and LEXICAL ITEM and fixed effects for VOWEL HEIGHT, DURATION,

[9] LEXICAL ITEM was not included as a random intercept in the preliminary models because the models failed to converge when it was included. I double checked that the relevant F1 distinction was LOW ~ NON-LOW by first building a minimal model with a three-way VOWEL QUALITY factor and conducting *post hoc* pairwise comparisons for this model. Pairwise comparisons showed that /e/ and /o/ both differed from /a/ but not from each other, confirming that the relevant distinction is LOW (/a/) vs. NON-LOW (/e/, /o/).

PRECEDING GLIDE, FOLLOWING GLIDE, PRECEDING NASAL, FOLLOWING NASAL, and FOLLOWING UVULAR. It also contained interactions between VOWEL HEIGHT*PRECEDING NASAL, VOWEL HEIGHT*FOLLOWING NASAL, VOWEL HEIGHT*DURATION.

3.1.2. F2 Model

Construction of the F2 model was very similar to the F1 model. The minimal model containing just a random effect for SPEAKER and a fixed effect for VOWEL QUALITY (the three-way variable) showed that VOWEL QUALITY was significant. Post hoc pairwise comparisons showed the expected three-way contrast between /e/, /a/, and /o/ F2 values ($p<0.0001$). Subsequent models of F2 retained this three-way quality variable. As in the F1 analysis, subsequent preliminary models were constructed to test if DURATION, PRECEDING PLACE, FOLLOWING PLACE, PRECEDING VOWEL, and FOLLOWING VOWEL had an effect on F2.

DURATION did not show a significant main effect on F2, but it did show significant interactions with VOWEL QUALITY. Both PRECEDING PLACE and FOLLOWING PLACE for adjacent consonants showed significant main effects and interactions with VOWEL QUALITY. PRECEDING VOWEL and FOLLOWING VOWEL also showed significant main effects and interactions. Therefore, the final model included VOWEL QUALITY, DURATION, PRECEDING PLACE, FOLLOWING PLACE, PRECEDING VOWEL, and FOLLOWING VOWEL as fixed effects as well as two-way interactions between each of these effects and VOWEL QUALITY. This model failed to converge when random intercepts for both SPEAKER and LEXICAL ITEM were included, so only one random intercept, for SPEAKER, was used.

3.2. Results of F1 Model

Table 2.5 shows the results of the final model of F1. Interactions involving /q/ were rank-deficient so they were automatically dropped from the results summary by *lmer*.

Significant effects and interactions are shaded grey. Since the model was rank-deficient, it didn't provide an estimate for a following nasal or its interaction with vowel height.

Table 2.5: Results of F1 Model

```
F1 = (1|speaker) + (1|lexicalItem) + vowelHeight + duration + precedingGlide + followingGlide + precedingNasal + followingNasal + following/q/ + vowelHeight*duration + vowelHeight*precedingNasal + vowelHeight*followingNasal
```

FACTOR	ESTIMATE	s	d.f.	t	p
(INTERCEPT)	607.01	27.83	7.07	21.82	<0.001
VOWEL HEIGHT: NON LOW	-133.94	2.63	11207.81	-51.01	<0.001
DURATION	13.76	1.53	12644.6	9.02	<0.001
VOWEL HEIGHT: NON-LOW* DURATION	-21.55	2.4	12571.41	-8.97	<0.001
PRECEDING GLIDE	-19.45	3.15	11421.69	-6.17	<0.001
FOLLOWING GLIDE	-9.44	3.99	11789.18	-2.36	0.02
PRECEDING NASAL	9.69	4.03	9770.2	2.41	0.02
FOLLOWING /q/	48.01	11.28	11944.44	4.26	<0.001
VOWEL HEIGHT: NON-LOW * PRECEDING NASAL	-20.08	5.41	10573.49	-3.71	<0.001

As expected, the non-low vowels show a significantly lower F1 than /a/, which is estimated to have an F1 of around 607 Hz (the model intercept). The difference between /a/ and the non-low vowels is estimated to be about 134 Hz.

There is a significant main effect of DURATION, estimating that F1 increases by about 14 Hz per ms increase in duration. VOWEL HEIGHT also interacts with this effect; for the non-low vowels, this increase is estimated to be about 22 Hz less than for /a/ (i.e., a decrease of about 8 Hz per ms). Figure 2.3 shows the interaction between DURATION and VOWEL HEIGHT. Figure 2.3 presents log-transformed duration; the actual statistical modeling was done with the raw data scaled around zero, but since speech segment duration data tend to fall into lognormal distributions (see discussion in Rosen 2005), the log-transformed data were used for data visualization.

Figure 2.3. Relationship between DURATION and F1

For low vowel /a/, on the left, the slope of the trend line is positive, reflecting the significant main effect in the model. For non-low vowels, the slope of the line is basically flat. Recall that longer vowels were expected to be more peripheral; in terms of F1, this means higher F1 for longer /a/, and lower F1 for longer /e/ and /o/. The model (and Figure 2.3) suggests that /a/ does indeed become lower as duration increases, but the non-low vowels remain mostly unaffected. However, Figure 2.3 also shows that, though the effect of DURATION is significant, there is a large amount of variation for both low and non-low vowels, and the values don't fit a linear trend line particularly well. A non-linear model might better represent the relationship between these two variables.

Table 2.5 also shows a significant main effect of PRECEDING GLIDE on F1; F1 is estimated to be about 19 Hz lower when the vowel is preceded by a glide. This effect is as expected; glides have a lower F1 than vowels do, since they are articulated with greater constriction. Figure 2.4 shows this effect, which does not interact with VOWEL HEIGHT.

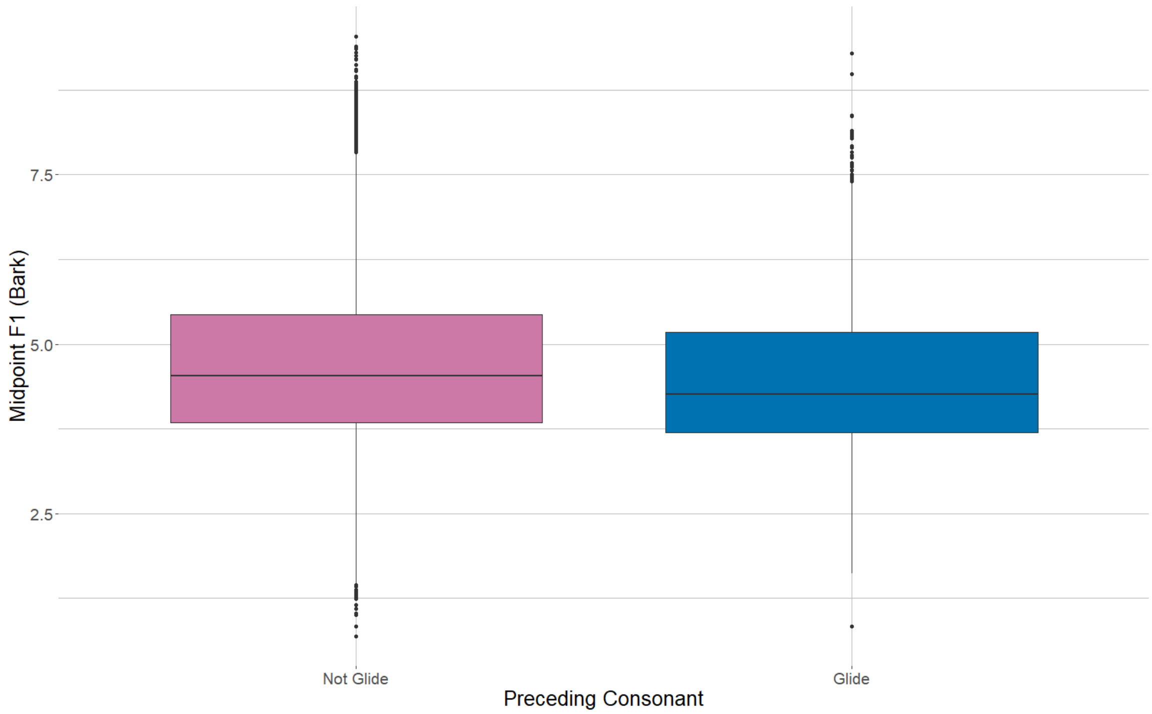

Figure 2.4. Relationship between PRECEDING GLIDE and F1

Vowels preceded by a glide have a mean F1 of 459.46 Hz, compared to vowels not preceded by a glide, which have a mean F0 of 481.46 Hz. Note that the "not preceded by glide" category includes vowels which were preceded by other consonants and vowels in utterance-initial, onsetless syllables.

Table 2.5 suggests one other main effect, that of FOLLOWING /q/. Following /q/ are estimated to raise F1 by about 48 Hz. This effect is also as expected; higher F1 corresponds to a lower vowel, which is the predicted effect of a uvular. Figure 2.5 shows this relationship.

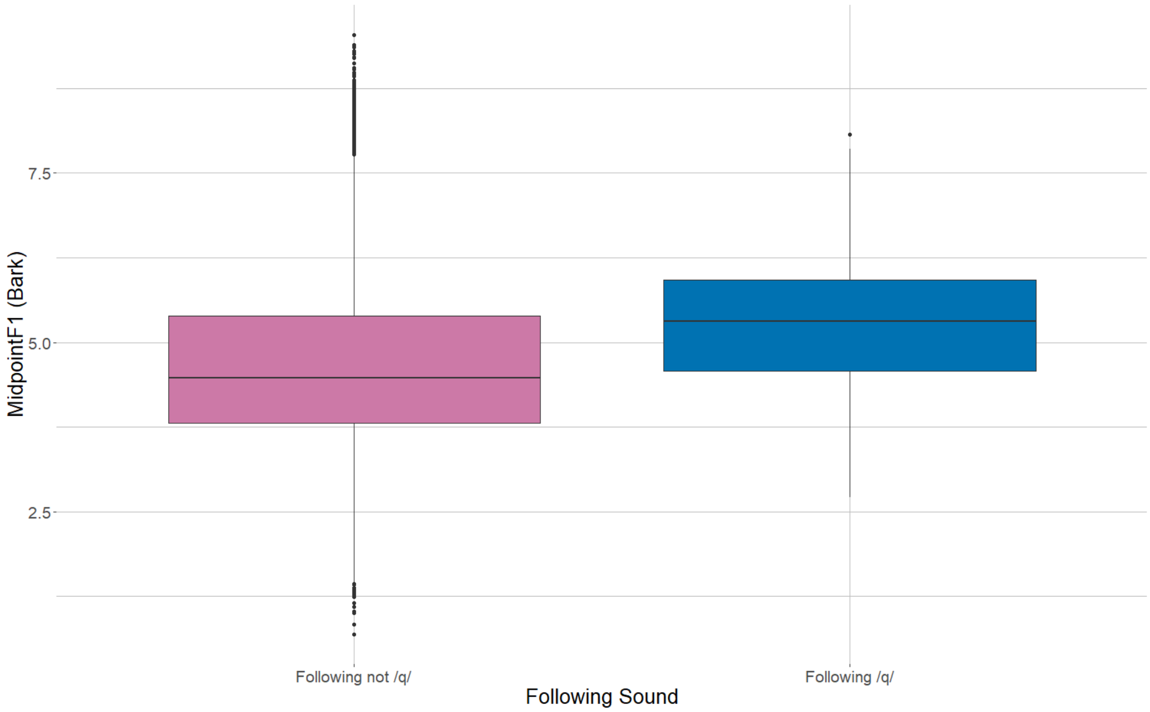

Figure 2.5. Relationship between FOLLOWING /q/ and F1

Vowels followed by /q/, shown on the right in Figure 2.5, have a mean F1 of 554.74 Hz, compared to just 477.41 Hz when they are not followed by a uvular. Note that, parallel to PRECEDING GLIDE, the category of "following not /q/" contains vowels followed by other consonants as well as vowels in open, utterance-final syllables.

Finally, the model shows a significant interaction between PRECEDING NASAL and VOWEL HEIGHT. Preceding nasals do not have a significant main effect in the model, though note in Table 2.5 that if we selected a less conservative alpha level (e.g., $p<0.05$), this effect would be considered significant. However, the interaction suggests that when a non-low vowel is preceded by a nasal, F1 is about 20 Hz lower than it would be if it were an /a/ preceded by a nasal. Figure 2.6 shows the interaction.

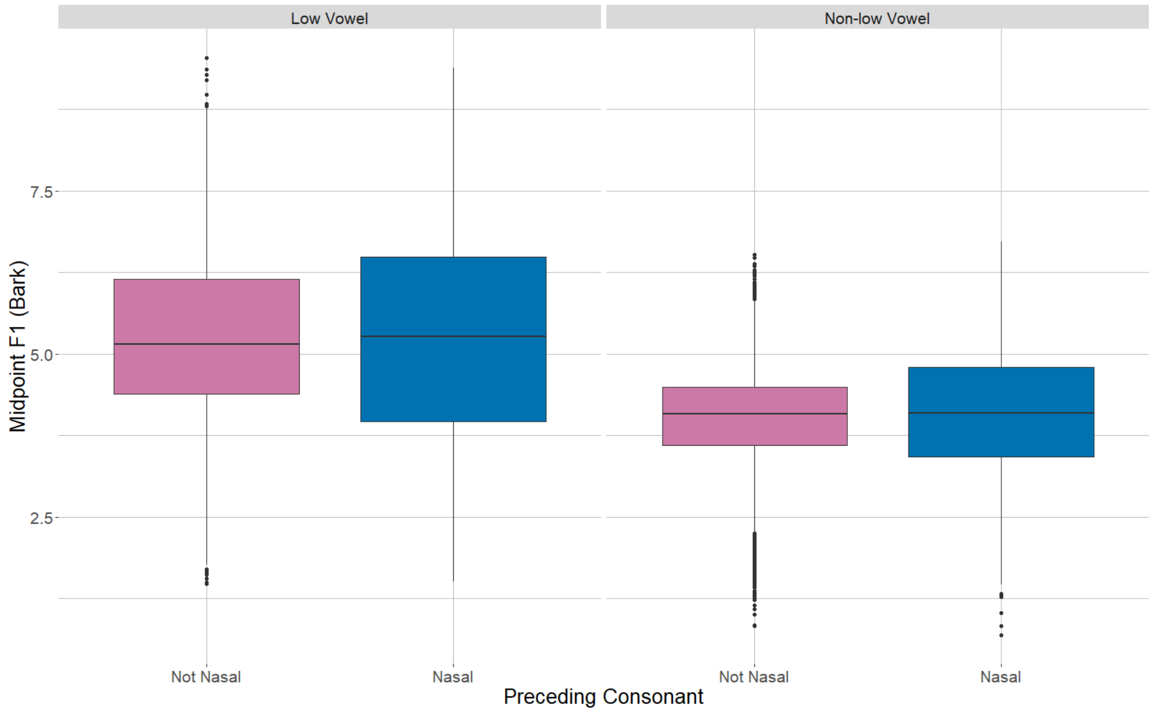

Figure 2.6. Relationship between PRECEDING NASAL and VOWEL HEIGHT for F1

In Figure 2.6, the interaction is primarily visible in comparing the variation between the categories on the left and right sides of the figure. The mean F1 for low vowels preceded by a nasal is 551.75 Hz, compared to 543.20 Hz when preceded by something else. This is a difference of about 8 Hz. The mean F1 for a non-low vowel preceded by a nasal is 411.67 Hz, compared to 402.34 Hz when preceded by some other sound, also a difference of about 8 Hz. In both cases, vowels preceded by a nasal have a slightly higher mean F1. However, the non-low vowels show much less variation than the low vowels do; this must be why the model in Table 2.5 identifies a significant interaction between vowel quality and a preceding nasal.

3.3. Results of F2 Model

Table 2.6 shows the results of the F2 model. Significant results in Table 2.6 are shaded grey.

Table 2.6: Results for model of adjacent sounds and F2

```
F2 = (1|speaker) + vowelQuality + duration + precedingPlace + followingPlace +
precedingVowel + followingVowel + vowelQuality*duration + vowelQuality*precedingPlace +
vowelQuality*followingPlace + vowelQuality*precedingVowel + vowelQuality*followingVowel
```

FACTOR	ESTIMATE	s	d.f.	t	p
(INTERCEPT)	1546.27	55.05	7.29	28.09	<0.001
VOWEL QUALITY: /e/	370.45	11.59	10009.18	31.95	<0.001
VOWEL QUALITY: /o/	-335.42	18.76	10009.25	-17.88	<0.001
DURATION	-21.53	4.45	10009.83	-4.83	<0.001
VOWEL QUALITY: /e/ * DURATION	83.49	7.14	10009.27	11.69	<0.001
VOWEL QUALITY: /o/ * DURATION	-41.51	8.48	10009.23	-4.9	<0.001
PRECEDING PLACE: LABIAL	-160.86	7.2	10009.1	-22.35	<0.001
VOWEL QUALITY: /e/ * PRECEDING PLACE: LABIAL	97.08	11.13	10009.08	8.72	<0.001
VOWEL QUALITY: /o/ * PRECEDING PLACE: LABIAL	9.02	18.51	10009.17	0.49	0.63
PRECEDING PLACE: PALATAL	149.81	10.83	10009.11	13.83	<0.001
VOWEL QUALITY: /e/ * PRECEDING PLACE: PALATAL	-31.01	14.87	10009.11	-2.09	0.04
VOWEL QUALITY: /o/ * PRECEDING PLACE: PALATAL	166.35	45.16	10009.06	3.68	<0.001
PRECEDING PLACE: VELAR	-10.19	9.18	10009.26	-1.11	0.27
VOWEL QUALITY: /e/ * PRECEDING PLACE: VELAR	38.39	12.79	10009.24	3	<0.01
VOWEL QUALITY: /o/ * PRECEDING PLACE: VELAR	-17.94	19.67	10009.07	-0.91	0.36
PRECEDING PLACE OF ARTICULATION: UVULAR	-37.11	48.98	10009.16	-0.76	0.45
PRECEDING PLACE: GLOTTAL	-27.24	9.99	10009.13	-2.73	<0.01
VOWEL QUALITY: /e/ * PRECEDING GLOTTAL	37.81	17.21	10009.12	2.2	0.03
VOWEL QUALITY: /o/ * PRECEDING GLOTTAL	134.25	30.36	10009.12	4.42	<0.001
FOLLOWING PLACE: LABIAL	-48.03	8.8	10009.03	-5.46	<0.001
VOWEL QUALITY: /e/ * FOLLOWING PLACE: LABIAL	2.65	12.36	10009.06	0.22	0.83
VOWEL QUALITY: /o/ * FOLLOWING PLACE: LABIAL	-11.16	18.34	10009.28	-0.61	0.54
FOLLOWING PLACE: PALATAL	143.75	10.87	10009.1	13.23	<0.001
VOWEL QUALITY: /e/ * FOLLOWING PLACE: PALATAL	-113.93	18.15	10009.07	-6.28	<0.001
VOWEL QUALITY: /o/ * FOLLOWING PLACE: PALATAL	32.08	48.24	10009.05	0.67	0.51
FOLLOWING PLACE: VELAR	-18.03	7.57	10009.19	-2.38	0.02
VOWEL QUALITY: /e/ * FOLLOWING PLACE: VELAR	68.22	11.17	10009.06	6.11	<0.001
VOWEL QUALITY: /o/ * FOLLOWING PLACE: VELAR	-26.82	16.67	10009.17	-1.61	0.11

Table 2.6: *(cont.)*

FACTOR	ESTIMATE	s	d.f.	t	p
FOLLOWING PLACE: UVULAR	-68.7	20.37	10009.09	-3.37	<0.001
VOWEL QUALITY: /e/ * FOLLOWING PLACE: UVULAR	-52.93	54.81	10009.05	-0.97	0.33
FOLLOWING PLACE: GLOTTAL	-2.75	13.7	10009.07	-0.2	0.84
VOWEL QUALITY: /e/ * FOLLOWING PLACE: GLOTTAL	-39.93	22.08	10009.08	-1.81	0.07
VOWEL QUALITY: /o/ * FOLLOWING PLACE: GLOTTAL	-85.4	48.16	10009.17	-1.77	0.08
PRECEDING VOWEL: /e/	16.23	7.01	10009.11	2.32	0.02
VOWEL QUALITY: /e/ * PRECEDING VOWEL: /e/	14.99	10.09	10009.11	1.49	0.14
VOWEL QUALITY: /o/ * PRECEDING VOWEL: /e/	-4.15	16.6	10009.09	-0.25	0.8
PRECEDING VOWEL: /o/	-32.4	10.67	10009.16	-3.04	<0.01
VOWEL QUALITY: /e/ * PRECEDING VOWEL: /o/	29.47	17.92	10009.24	1.64	0.1
VOWEL QUALITY: /o/ * PRECEDING VOWEL: /o/	23.63	22.65	10009.08	1.04	0.3
PRECEDING VOWEL: NONE	7.98	8.19	10009.12	0.97	0.33
VOWEL QUALITY: /e/ * PRECEDING VOWEL: NONE	23.63	11.64	10009.06	2.03	0.04
VOWEL QUALITY: /o/* PRECEDING VOWEL: NONE	16.91	17.54	10009.12	0.96	0.34
FOLLOWING VOWEL: /e/	8.14	6.88	10009.05	1.18	0.24
VOWEL QUALITY: /e/ * FOLLOWING VOWEL: /e/	10.64	10.02	10009.08	1.06	0.29
VOWEL QUALITY: /o/ * FOLLOWING VOWEL: /e/	8.85	17.09	10009.1	0.52	0.6
FOLLOWING VOWEL: /o/	-21.05	9.51	10009.12	-2.21	0.03
VOWEL QUALITY: /e/ * FOLLOWING VOWEL: /o/	36.36	13.84	10009.15	2.63	<0.01
VOWEL QUALITY: /o/ * FOLLOWING VOWEL: /o/	-9.88	22.13	10009.1	-0.45	0.66
FOLLOWING VOWEL: NONE	15.37	9.5	10009.28	1.62	0.11
VOWEL QUALITY: /e/ * FOLLOWING VOWEL: NONE	0.43	14.34	10009.13	0.03	0.98
VOWEL QUALITY: /o/ * FOLLOWING VOWEL: NONE	8.5	21.67	10009.06	0.39	0.69

Table 2.6 shows a significant main effect of each vowel quality. Vowel /a/, the baseline, is estimated to have an F2 around 1546 Hz (the intercept). Front vowel /e/ is estimated to have an F2 around 370 Hz higher than that. Front vowels are expected to have higher F2 than central or back vowels, because the constriction at the front of the mouth shortens the resonant space for F2. Back vowel /o/ is estimated to have an F2 about 335 Hz lower than /a/. Back vowels are expected to have lower F2s than central vowels, as the constriction of at the back of the mouth, paired with lip rounding, results in a longer resonant space for

F2. Post hoc pairwise comparisons with *emmeans* show that all three categories differ from one another in F2 ($p<0.0001$).

Table 2.6 suggests a significant effect of DURATION – vowels with a greater duration tend to have a lower F2 on average. However, there is also a significant interaction with VOWEL QUALITY; for /e/, longer tokens have an F2 which is greater than the overall effect by about 83 Hz per ms of duration, while /o/ has an F2 which is about 42 Hz lower than the main effect per ms of increased duration. These effects are consistent with increased peripheralization. Vowel /e/ becomes more front (higher F2) and /o/ becomes more back, or more rounded, resulting in a lower F2. Figure 2.7 shows the relationship between DURATION and F2 for each level of VOWEL QUALITY.

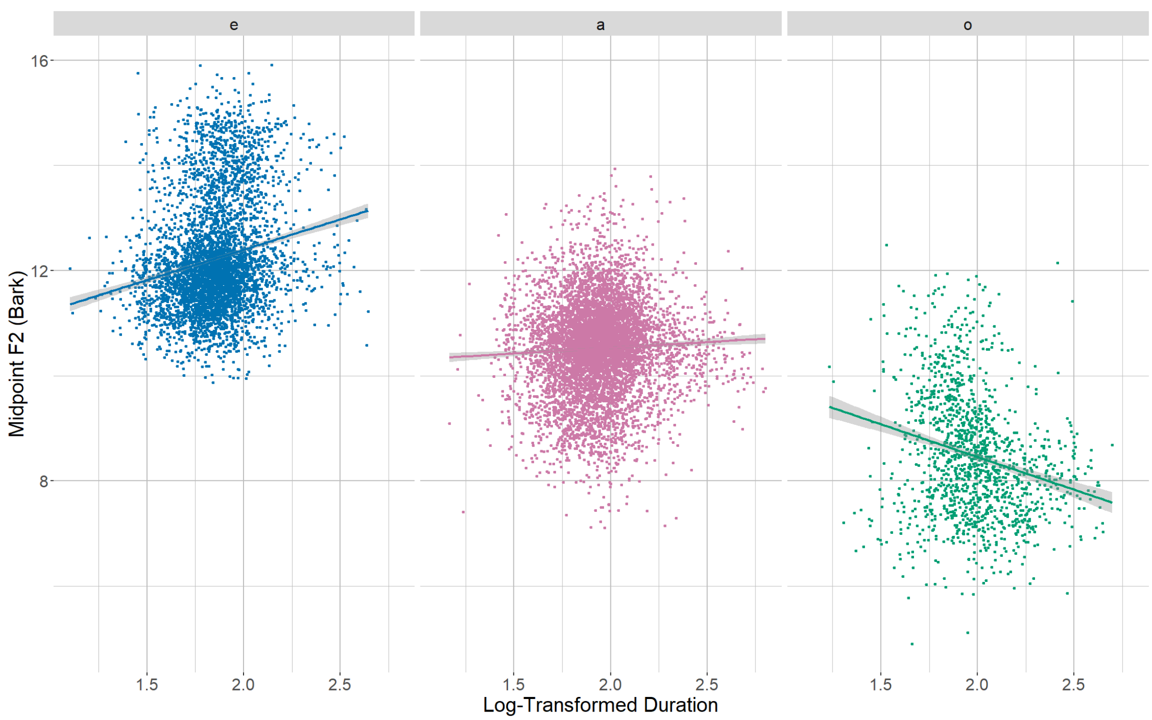

Figure 2.7. Relationship between DURATION and F2 for each VOWEL QUALITY

As with the F1 values, the trend lines show the changes in F2 as suggested by the model. For /e/, F2 increases along with duration, for /a/ the trend line is basically flat, and for /o/, F2 decreases with duration. Like the F1 plot, these values don't fit a linear trend line particularly well.

The model also shows a number of significant main effects and/or interactions of the adjacent sounds. Figure 2.8 shows the mean F2 at each level of PRECEDING PLACE, divided by VOWEL QUALITY.

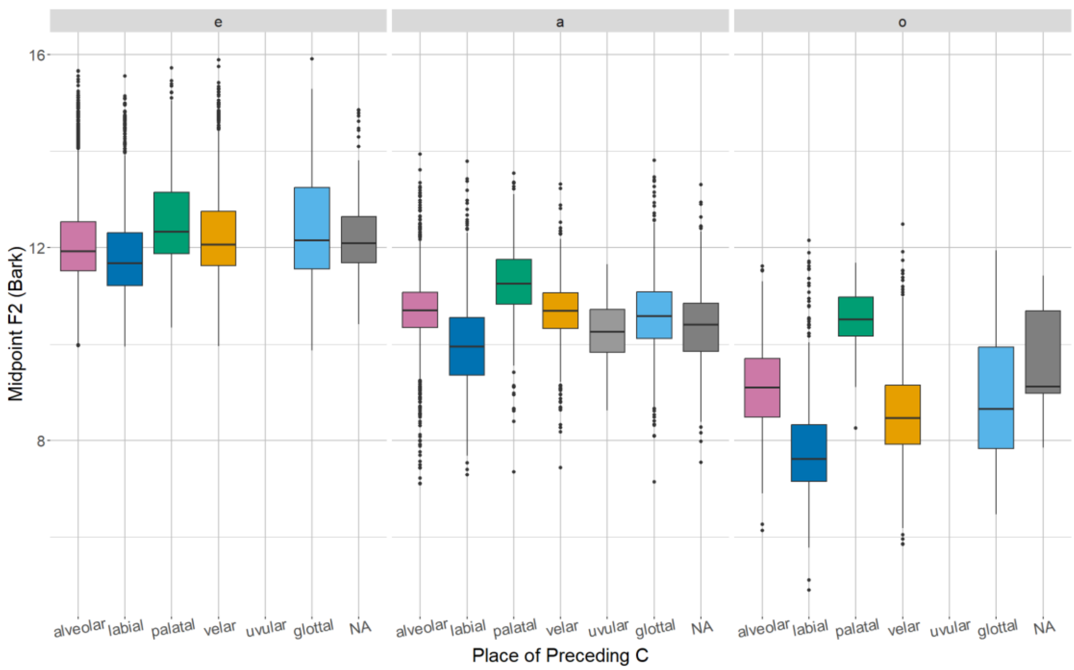

Figure 2.8. Effect of PRECEDING PLACE on F2, divided by VOWEL QUALITY

A preceding labial is estimated to result in an F2 that is approximately 161 Hz lower than when the vowel is preceded by an alveolar. This factor also significantly interacts with VOWEL QUALITY, suggesting that the F2 lowering is approximately 97 Hz less when the vowel is /e/. The difference between the blue and pink boxes in Figure 2.8 illustrates this effect and interaction. The mean F2 for an /a/ preceded by a labial is 1274.44 Hz, compared to 1419.01 Hz when the /a/ is preceded by an alveolar (difference ~ 145Hz). In contrast, the mean F2 for an /e/ preceded by a labial is 1728.62 Hz, compared to 1789.30 Hz when preceded by an alveolar (difference ~ 61 Hz). This effect of preceding labials is in the expected direction, though the fact that /e/ is less affected by labials than /a/ is surprising, given that previous work (e.g., Hillenbrand, Clark, & Nearey (2001) on American English) finds that labial sounds affect F2 primarily in front vowels.

Preceding palatals also showed a significant main effect; vowels when preceded by /j/ have an F2 around 150 Hz higher than when preceded by other sounds. There is a significant interaction with VOWEL QUALITY here as well; for /o/, the effect is around 166 Hz greater than for /a/. This effect can be seen by comparing the green and pink boxes in the middle and right-hand sections of Figure 2.8. The mean F2 for /a/ preceded by a palatal is 1547.47 Hz (vs. 1419.01 Hz for preceding alveolars). Back vowel /o/ has a mean F2 of 1362.27 Hz when preceded by a /j/, compared to 1110.07 Hz when preceded by an alveolar. This is a difference of ~ 252 Hz, versus just 128 Hz for /a/.

Preceding velars do not show a significant main effect, but there is a significant interaction. Front vowel /e/ preceded by a velar has a higher F2 than /e/ preceded by other consonants. The mean F2 of /e/ after a velar is 1838.84 Hz, compared to 1789.39 Hz after an alveolar. Compare the orange and pink graphs in the left pane of Figure 2.8 to the ones in the middle pane. This effect was not predicted; previous work has found an effect of velars on F2 for non-front vowels /a/ and /o/, but no effect on /e/.

Preceding glottals show a significant main effect which suggests that F2 is about 27 Hz lower when a vowel follows a glottal. There is also a significant interaction with /o/. For /o/, the effect of a preceding glottal on F2 is an increase in F2 of about 134 Hz. Compare the light blue and pink means in the right most pane in Figure 2.8. The mean F2 of /o/ when it follows a glottal is 1084.87 Hz. This effect is unexpected; glottals do not require any particular lip or tongue configuration, so they are not expected to affect formants. One possible explanation is that for some reason in Enenlhet speakers produce glottal sounds with spread lips, which would account for the higher F2. Alternatively, Praat may have done a worse job of tracking pitch and formants when vowels are adjacent to glottals due to glottalization and aperiodicity (see Chapter 4), in which case this effect might be due to measurement errors.

In addition to the effects of preceding consonants, there are also significant effects of some levels of FOLLOWING PLACE in the model. Figure 2.9 shows the relationship between FOLLOWING PLACE and F2 for each level of VOWEL QUALITY.

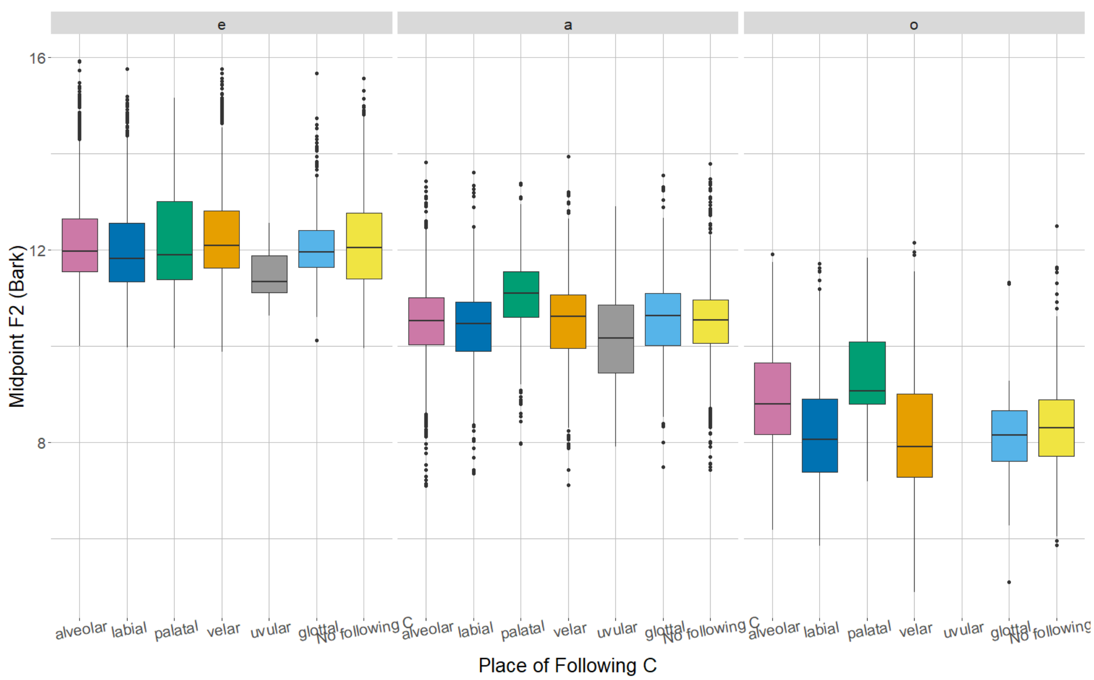

Figure 2.9. Effect of FOLLOWING PLACE on F2, divided by VOWEL QUALITY

Following labials resulted in a significant main effect indicating lower F2 for vowels followed by labials compared to other consonants. The size of this effect is estimated by the model in Table 2.6 to be around 48 Hz. Unlike the effect of a preceding labial, the effect of a following labial consonant does not interact with VOWEL QUALITY. Figure 2.10 shows this effect for all vowels not divided by quality.

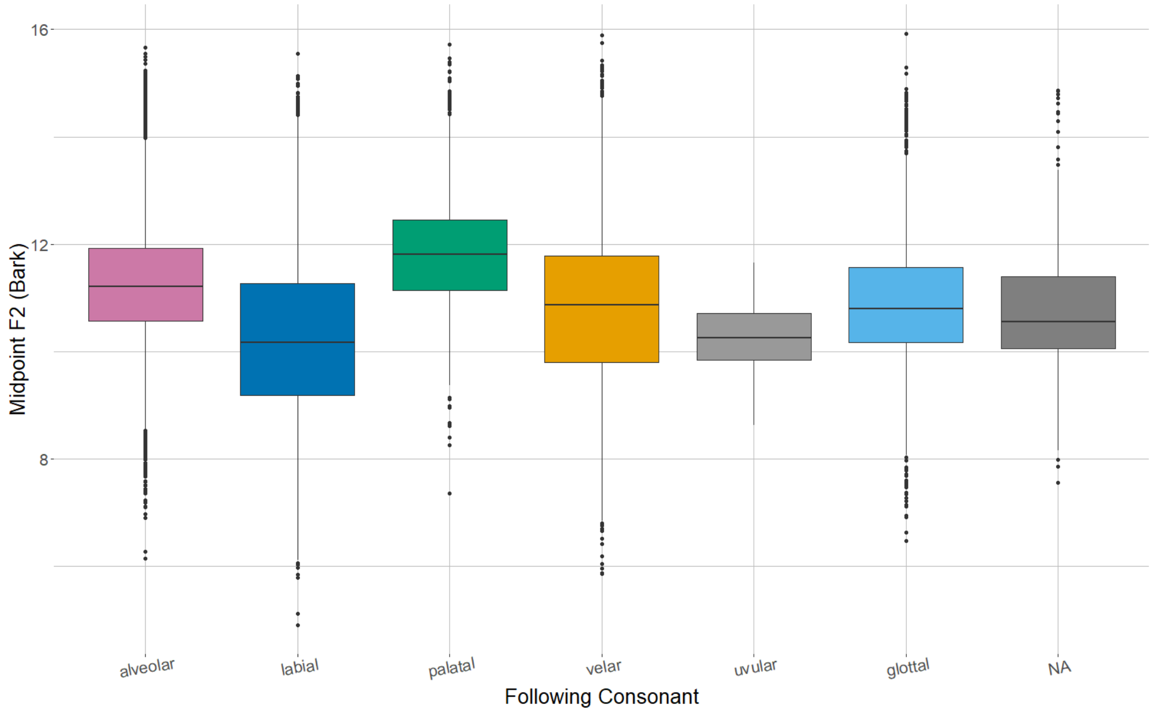

Figure 2.10. Effect of FOLLOWING PLACE on F2 for all vowels

In Figure 2.10, the effect of a following labial can be seen by comparing the mean F2 for vowels followed by labials (dark blue, 1460.32 Hz) to the mean F2 for vowels followed by alveolars (pink, 1541.70 Hz). The means differ by around 80 Hz.

Following palatals raise F2 by an estimated 144 Hz, according to Table 2.6. There is also a significant interaction with VOWEL QUALITY; the effect is estimated to be around 114 Hz less when the vowel is /e/. The mean F2 for /e/ followed by /j/ is 1775.38 Hz (vs.

1791.07 Hz when followed by an alveolar), and the mean F2 for /a/ followed by /j/ is 1492.84 Hz (vs. 1388.12 Hz when followed by an alveolar). Figure 2.9 shows the effect of the following /j/ on /e/ compared to /a/ and /o/ – compare the green (palatal) to the pink (baseline, alveolar). The direction of this effect is as expected; palatalization results in higher F2 values. It is also unsurprising that this effect is smaller for /e/ than it is for /a/ and /o/; /e/ already has the highest F2 values of the inventory, so it has less "space" for F2 to increase.

Following uvulars also have a significant effect on F2 which does not interact with VOWEL QUALITY. The model estimates that F2 is 69 Hz lower when the vowel is followed by a uvular. Figure 2.10 shows the effect of a following uvular (light grey, mean 1361.56 Hz), compared to a following alveolar (pink, mean 1541.70 Hz). This effect is also as predicted; uvulars often trigger backing, which results in lower F2. It is surprising that this apparently occurs for all vowel qualities rather than affecting primarily the front vowels.

The final set of significant effects and interactions in the model relates to the adjacent vowels. Preceding /o/ results in lower F2 values for all vowels; this effect does not interact with VOWEL QUALITY. This effect is as expected based on previous literature. As has already been seen, /o/ has the lowest F2 of the three Enenlhet phonemes, and changes in lip configuration are quite slow (Benguerel & Cowan 1974), so lowered F2 adjacent to /o/ is likely due to the fact that /o/ is rounded. In contrast, following /o/ is not a significant main effect in the model, but it does significantly interact with /e/; /e/ followed by /o/ have an F2 about 36 Hz greater than other vowels followed by /o/. The mean F2 for /a/ followed by /a/, the baseline value for both factors, is 1370.01 Hz. The mean F2 for /a/ followed by /o/ is 1374.68 Hz (~ 4 Hz higher than when /a/ follows /a/ – this difference is not significant). In contrast, the mean F2 for /e/ followed by /o/ is 1816.49 Hz, versus

1777.46 Hz when it is followed by /a/, a difference of ~ 39 Hz. As indicated above, PRECEDING VOWEL has the expected effect. However, the effect of FOLLOWING VOWEL was not anticipated. Following /o/ was expected to result in a lower F2 due to lip rounding, and instead a following rounded vowel results in *higher* F2 values for unrounded /e/.

4. DISCUSSION

This study provides empirical evidence which supports the qualitative description of Enenlhet presented in Unruh, Kalisch, and Romero (2003) and Heaton (2019–). Enenlhet makes two phonetic distinctions in height (F1), which I have called a low ~ non-low contrast; the non-low vowels /e/ and /o/ have lower F1 values than the low vowel /a/. Enenlhet also makes three phonetic distinctions in anteriority (F2); /e/ has the highest F2 values and /o/ the lowest. It also shows that the Enenlhet vowel space is quite similar to Enxet (Elliott 2016), Sanapaná (van Gysel 2022), and Angaité (Wheeler 2020). All four languages show more variation in the F2 dimension compared to F1 and relatively high F1 values for the non-low vowels. The F1 values in these data also suggest a higher /a/ vowel than is found in many 5-vowel systems, supporting my qualitative observation that many tokens of /a/ are quite schwa-like.

Cross-linguistic comparison is useful in contextualizing these formant values. Though each category encompasses a range of values in the F1 x F2 space, and, as seen here, these formants are affected by both adjacent consonants and vowels, there are some cross-linguistic trends in the ranges associated with different vowel categories. Table 2.7 shows the F1 and F2 ranges reported for low, mid, and high vowel phonemes in a variety of other languages.

Table 2.7: Average F1 and F2 values for point vowels in other languages.

SOURCE	LANGUAGE	VOWEL	F1 (HZ)	F2 (HZ)
Öhman 1966	Swedish	ɑ		1000
Öhman 1966	English	a		1160–2340
Abramson & Ren 1990	Thai	a	810–880	1505–1580
	OVERALL RANGE	a	810–880	1000–2340
Öhman 1966	English	u		830-960
Abramson & Ren 1990	Thai	u		750
Bradlow 1993	American English	u	325	1238
Bradlow 1993	Madrid Spanish	u	322	992
Bradlow 1993	Greek	u	339	879
	OVERALL RANGE	u	325–340	750–1240
Öhman 1966	English	i		2260-2340
Abramson & Ren 1990	Thai	i	250	
Bradlow 1993	American English	i	268	2393
Bradlow 1993	Madrid Spanish	i	286	2147
Bradlow 1993	Greek	i	310	2040
	OVERALL RANGE	i	250–310	2040–2390
Abramson & Ren 1990	Thai	e	390-440	1985-2100
Bradlow 1993	American English	e	430	2200
Bradlow 1993	Madrid Spanish	e	458	1814
Bradlow 1993	Greek	e	474	1641
	OVERALL RANGE	e	390–475	1640–2100
Abramson & Ren 1990	Thai	o	455–490	455–490
Bradlow 1993	American English	o	382	1160
Bradlow 1993	Madrid Spanish	o	460	1019
Bradlow 1993	Greek	o	476	864
	OVERALL RANGE	o	380–490	865–1160

These languages are by no means a representative sample of the world's languages. Nevertheless, they suggest that phonemic high vowels (represented here by /i/ and /u/), tend to have F1 values between 250 and 350 Hz, and phonemic mid vowels (represented here by /e/ and /o/) tend to have F1 values between 380 and 490 Hz. The mean F1 value for the non-low vowels in Enenlhet is estimated to be around 460 Hz. In other words, the formant range used by Enenlhet non-low vowels is more similar to the F1 range occupied by phonemic mid vowels in languages with 5 (or more) vowels, and it is much higher than the F1 range typically occupied by phonemic high vowels in other languages. The F2 ranges occupied by Enenlhet vowels, in contrast, are comparable to the F2 ranges reported for phonemic front, central, and back vowels in other languages. These results are consistent with Elliott's (2016) findings for Enxet and van Gysel's (2022) results from Sanapaná. While Enenlhet (and sisters) make a familiar low ~ non-low contrast, the non-low vowels are *phonetically* quite "mid", rather than occupying the F1 space that phonemic high vowels usually fall into (and which bilingual Enxet and Sanapaná speakers produce for high vowels /i/ and /u/ when speaking Spanish).

This inventory presents a major question for future perceptual research. Liljencrants and Lindblom's (1972) first attempt at modeling cross-linguistic vowel inventories over-generated distinctions in the F2 dimension, particularly when F1 was low. Lindblom (1986) adjusts the model to prefer F1 over F2 for generating new vowel contrasts and finds a better match with observed vowel inventories. He suggests that F1 is preferred for contrasts because F1 is more reliably perceptible due to its higher intensity. If this is the case, Enenlhet flouts expectations by relying primarily on the *less* salient acoustic dimension to make the relevant phonemic contrasts. In Enenlhet, the average F1 values only show a range of about 140 Hz, while F2 spans a range of over 800 Hz. Future perceptual research is needed to determine which additional acoustic cues speakers use to aid in category distinction. A Support Vector Machine Learning Classifier significantly improved at classifying Enenlhet vowels when duration and intensity at the vowel midpoint were added to the model.[10] This improvement suggests that speakers of Enenlhet may rely quite heavily on secondary spectral cues in vowel identification.

In addition to the formant values associated with each proposed phoneme, which surfaced mostly as predicted, many of the effects of adjacent consonants on F1 and F2 were also consistent with observations from other languages. In the F1 dimension, preceding

[10] Thanks to Fernando Llanos for coding and explaining this classifier to help explore my formant data.

glides (/j, w/) resulted in lower F1 values, indicating vowel raising. A following /q/ resulted in higher F1 values, indicating vowel lowering. These effects are as expected based on previous studies. However, Enenlhet showed a surprising number of global effects; that is, factors which affected both non-low and low vowels similarly. In most languages, since non-low vowels and low vowels differ in F1, adjacent consonants are often found to have opposite effects on these two categories, or to affect one and not the other. The effect of adjacent nasals did interact with vowel height; non-low vowels showed lower F1 after nasals, and /a/ showed a higher F1 in the same position. Nasals were expected to lower F1, so the fact that the opposite occurred for /a/ is surprising. In Enenlhet, apparently, a preceding nasal results in greater dispersion in the F1 dimension, with a higher articulation of non-low vowels, and slightly lower productions of /a/.

In the F2 dimension, most of the largest effects were consistent with previous research. Adjacent labials resulted in lower F2 values, and adjacent /j/ resulted in higher F2 values for non-front vowels and slightly lower F2 values for /e/. These results are as expected based on previous work. However, the effect of velars on /e/ and the effect of glottal sounds on /o/ were both unexpected. I don’t have an explanation for why velars raise F2 for /e/. The effect of adjacent glottals on /o/ is very odd, and I suspect it has something to do with the effect of glottals in Enenlhet on F0 (see Chapter 4), or with measurement errors caused by aperiodicity around glottal consonants. The only effect of adjacent vowels on F2 was due to adjacent /o/. When /o/ precedes another vowel, it lowers F2, which is the expected effect. However, when /o/ follows an /e/, it apparently raises F2, which is unexpected.

5. CONCLUSION

This study has provided the first empirical description of Enenlhet vowel quality and the first large-scale study of vowel quality in any Enlhet-Enenlhet language. It provides robust acoustic descriptions of Enenlhet vowel categories and shows that the proposed phonemic non-low vowels are have higher F1 values than peripheral vowels in other three-vowel inventories. This study supports the use of phonetic labels [e] and [o] for the non-low vowels to capture this fact. Because of this relatively small spread in the F1 dimension, Enlhet-Enenlhet languages have a highly typologically unusual vowel inventory that makes more use of the F2 space than F1.

What this study cannot do is definitively describe the phonemic contrasts for all speakers. The transcriptions reflect the orthographic, and perhaps phonemic, intuitions of one speaker, and thus the study suggests that this speaker (MRR) likely has three phonemic vowel qualities. Other speakers show a similar spread of tokens; they use roughly the same acoustic space, and therefore *may* also have similar phonemic inventories. The fact that a machine learning classifier was able to achieve fairly accurate categorization of the tokens mostly *a priori* also suggests that the tokens roughly cluster around three different centers of gravity in the acoustic space. However, perception studies and additional consultation with a range of native speakers is needed to assess whether there is inter-speaker variation (or dialect variation) in the actual phonemic contrasts speakers use.

In addition to describing the phonetic ranges associated with each orthographic vowel quality, this study has explored the effects of duration and adjacent sounds on both F1 and F2. Many of the effects shown here are familiar from other languages, particularly the effects on F2 of labial and palatal sounds and the effects of glides and uvulars on F1.

This look at Enenlhet provides further evidence for these types of coarticulation as robust cross-linguistic patterns rather than regional or language-family-specific habits.

In addition to closer exploration of speaker variation, this study presents several questions for future perceptual work. First and foremost, perception studies are needed to closely define the boundaries between phonemes for a range of speakers, especially in the F2 dimension where all three categories occupy a somewhat distinct space. Perception studies should also investigate which additional acoustic cues speakers use for vowel identification, given the minimal use of the perceptually salient F1 dimension. Adjacent nasals were shown to raise F1 in Enenlhet in contrast to the effects reported for many other languages; future work is also needed to determine whether, and how, speakers use this change in F1 to maintain the two-way height contrast in the language in environments which may be more perceptually ambiguous (such as in the presence of reduced amplitude and antiformants adjacent to nasals).

Chapter 3: Vowel Duration

1. Introduction

This study provides a detailed phonetic description of vowel duration in Enenlhet with an eye toward identifying lexical stress and determining if the language makes a phonemic vowel length distinction. As noted in Chapter 1, there has not been a firm generalization about stress placement in any Enlhet-Enenlhet language, and in Enenlhet, the only mention of word *or* phrase level prominence is a brief statement related to information structure that "the word before the verb receives a certain accent, it is emphasized" (Unruh, Kalisch, & Romero 2003: 110).[11] For related languages, Elliott (2021) provides phonetic data showing that word-level prominence varies based on both syllable structure and morpheme type. Wheeler (2020) describes final pitch rises and final lengthening in Angaité, but these data were elicited in isolation, conflating word- and phrase-level prominence. Given this sporadic account of word-level prominence in related languages, and the fact that stress is cross-linguistically frequent, an exploratory analysis to identify it is worthwhile.

The other primary goal of this study is to investigate phonemic length in Enenlhet. Sister languages Enxet and Enlhet make phonemic vowel length contrasts, and Wheeler (2020) also suggests on the basis of speaker corrections of vowel length that there are long vowels in Angaité. Van Gysel (p.c. 2021), on the other hand, analyzes phonetic long vowels in Sanapaná as the result of syllable structure rather than a phonemic contrast. As discussed in Chapter 1, the history of phonemic length contrasts in Enlhet-Enenlhet languages is still under discussion, but two previous analyses agree that Enenlhet does not have a phonemic length contrast. However, they make different predictions for the other languages in the

[11] Original Spanish: "…la palabra delante del verbo recibe cierto acento, es enfatizada." (translation mine)

family (Angaité, Sanapaná, Guaná). The goal of this description is to empirically confirm previous descriptions of Enenlhet as *not* having a phonemic length contrast. Doing so requires ruling out other factors affecting duration, which also supports future historical research. If phonemic length is an innovation in Enxet and Enenlhet, the synchronic lengthening processes in Enenlhet may suggest historical starting points for the development of phonemic length in its sisters.

The following sections provide the methods specific to this study, including a rationale for each effect included in the model (Section 2); the results of the statistical analysis (Section 3); and a brief discussion (Section 4).

2. METHODS

The corpus and participants for this study are described in detail in Chapter 1, Section 3. Refer there as well for segmentation criteria and general exclusion criteria.

2.1. Annotation and Classification

This section provides a rationale for each of the factors that was included in the statistical analysis and details how tokens were classified with respect to these factors.

2.1.1. Position Within Phrase

Pre-boundary lengthening is widely attested, beginning as far back as Oller (1973), who finds final lengthening in English in sentences of various types (declarative, interrogative, and imperative) and in different final syllable structures, stress conditions, and boundary strengths. This effect has also been attested in Dutch (Cambier-Langeveld 1997), Finnish (Nakai, Kunnari, Turk, Suomi, & Ylitalo 2009), Hebrew (Berkovits 1994), Hungarian (Hockey & Faygal 1999), Japanese (Ueyama 1997), and Swedish (Fant, Nord,

& Kruckenberg 1987). Given its prevalence, pre-pausal lengthening is likely a universal feature of language, perhaps required by universal cognitive processes or motor constraints (see Byrd & Saltzman 2003).

The domain and magnitude of pre-boundary lengthening vary. For example, Nakai et al. (2009) report that the Finnish half-long vowel category exhibits a ceiling on pre-pausal lengthening. They also find that it operates progressively across several syllables, while Turk and Shattuck-Hufnagel (2007) find that final-lengthening in American English is not progressive. Cambier-Langeveld (1997) reports that final lengthening in Dutch is limited to the immediately pre-boundary syllable, though lengthening is progressive within the segments of that syllable, and the effect is modulated by syllable structure.

The research on pre-boundary lengthening makes strong predictions for Enenlhet. It suggests pre-pausal vowels will experience lengthening. This lengthening is most likely to affect word-final vowels in pre-boundary words, but it may progress backward across multiple syllables, with vowels closer to the boundary lengthened more. If Enenlhet has a phonemic vowel length contrast, pre-boundary lengthening may exhibit caps on lengthening to maintain the phonemic length contrast.

Phrase boundaries were identified using the criteria described in Chapter 1. Briefly, any utterance boundaries in the Heaton (2019–) transcriptions were treated as utterance boundaries and any vowel ending more than 250 ms before the start of the next one was also treated as pre-pausal. This is a relatively conservative criterion adopted to avoid misclassifying VCC.CV sequences as utterance breaks, since vowel boundaries were the only ones adjusted and extracted. A PRE-PAUSAL WORD factor marked whether each word was pre-pausal or not. This factor was not expected to have a significant main effect on its own, but previous work suggests that it may interact with a vowel's position within a word (with final vowels in pre-pausal words lengthening more than penults, etc.). Each vowel was also marked as being in a pre-pausal syllable (final vowel in pre-pausal word) or not (all other vowels). PRE-PAUSAL SYLLABLE *was* expected to have a significant effect in the model, as this is the position most likely to be affected by pre-pausal lengthening. Section 3.1 provides further discussion of these two factors and the process that was used to determine which to include in the overall duration model.

2.1.2. Position Within Word

Duration has been robustly shown to correlate with stress in a wide variety of languages, including: Balti Tibetan (Caplow 2016), Castilian Spanish and Central Catalan (Ortega-Llebaria & Prieto 2011); Dutch (Nooteboom 1972); English, German, and Spanish (Delattre 1966); Jordanian Arabic (de Jong & Zawaydeh 1999); and Swedish (Lindblom & Rapp 1973). There are also language-specific interactions between syllable structure and the duration changes associated with stress. In Swedish, vowels are shorter when they are followed by more consonants, and, to a lesser extent, by more syllables within a word (Lindblom & Rapp 1973). In contrast, Sinhala vowels are actually *longer* in closed syllables (Letterman 1994). Lengthening due to stress may be affected by a vowel's position within a word or by syllable structure. Enenlhet has both open (C)V and closed (C)VC syllables, and Elliott's (2021) examination of stress in Enxet suggests that other possible cues to stress (pitch and intensity) are affected by these factors.

Fixed lexical stress in Enenlhet, if present, is predicted to appear as an effect of a vowel's position within a word once other factors have been accounted for. Each vowel's position within the word was marked in two ways. The WORD POSITION (RIGHT) factor marked position from the right edge of the word (treating final syllables as "1") and the

WORD POSITION (LEFT) factor marked it from the left (treating initial syllables as "1"). Since words varied in length, marking from both directions provided a consistent anchor point for stress regardless of whether it attends to the right or left edge of the word. Neither of these factors ended up being useful in the overall duration model; see Section 3.1 for description of preliminary testing using these factors to determine if there is an independent effect of a vowel's position within the word.

2.1.3. Following Consonant Voicing

While not universal (cf. Mitleb 1984 on Jordanian Arabic), vowels tend to be longer before voiced consonants than voiceless ones. This effect is particularly notable within consonant classes (i.e., voiced vs. voiceless stops or fricatives). The consonant voicing effect is very robustly documented in English, where the alternation is perceptible and phonologized (House & Fairbanks 1953; or see e.g., Sanker 2020, for an overview). Fintoft (1961) reports a similar effect in Norwegian, and Chen's (1970) cross-linguistic study of Russian, French, English, and Korean finds the same pattern in all four languages. The size of the effect varies across languages, with some authors suggesting that the smallest effects are due to articulatory constraints (e.g, Georgian; Begus 2017). Some authors have framed the consonant voicing effect as a tradeoff between consonant and vowel duration, as some languages show an inverse relationship between the two (e.g., Bye, Saguelin, & Toivonen (2009) for Inari Saami). Since many studies find that duration effects are not strictly additive, the effect of a following voiceless consonant may interact with that of pre-pausal position or with syllable structure.

The extant research on the relationship between vowel duration and following consonant voicing predicts that vowels in Enenlhet will be longer before voiced consonants and shorter before voiceless ones. However, the Enlhet-Enenlhet voicing distinction (with the exception of /g/) maps onto the obstruent-sonorant distinction, and most research has investigated the effect in pairs of consonants differing only on voicing. Enenlhet therefore presents a relatively unexplored case, as voicing is not a particularly relevant cue to consonant identification. An effect of consonant voicing in Enenlhet would point toward universal constraints as the cause rather than a language-specific phonological rule.

The identity of the following consonant was recorded for each vowel in the corpus using the methodology described in Chapter 2. A FOLLOWING VOICELESS CONSONANT factor was then computed from this annotation, and each vowel was marked as either preceding a voiceless consonant or not preceding a voiceless consonant. The category of "not preceding a voiceless consonant" encompassed cases where the following sound was a vowel; where the vowel appeared in a pre-pausal, open syllable and therefore had no following consonant; or where the following consonant was voiced.

2.1.4. Vowel Quality

Vowel qualities also have different inherent durations, mostly correlated with vowel height, which is demonstrated by both synchronic and diachronic studies. Klatt (1976) finds that low vowels in American English have a longer inherent duration than high vowels, as does Lehiste (1970). However, these differences are fragile; duration differences due to vowel quality in French are mostly leveled out at phrase boundaries (Fletcher 1991). Though inherent duration is often dependent on vowel height, there are some languages in which vowels differ in their inherent durations based on some other feature (e.g., frontness, as in Sinhala; Letterman 1994). Because languages can vary in the parameters that determine a vowel's inherent duration, I began with a three-way VOWEL

QUALITY factor, which had the same levels as the one used in the F2 study in Chapter 2. That is, orthographic <a, a'a> were treated as quality /a/, <e, e'e> were treated as quality /e/, and <o, o'o> were treated as quality /o/. I expected to find that VOWEL QUALITY: /a/ had the longest inherent duration, and that /e/ and /o/ did not differ, as this is the distinction that is reported for most languages. See Section 3.1 for further discussion of how this variable was used in the final duration model.

2.1.5. Phonemic Length

Inherent duration differences sometimes lead to the development of a length contrast. There is a synchronic alternation between vowel lowering to /a/ and vowel length in Quechua which indicates that the inherent duration has been reinterpreted in some cases as a phonemic length distinction (Adelaar 1984). The reverse of this process has occurred in Gallo-Italian (Saunders 1978), with some historic phonemic length contrasts being replaced by quality alternations. In the Americas, the Chichicastenango variety of K'iche' has undergone a similar shift from a phonemic long-short distinction (as in other varieties of K'iche') to a set of quality distinctions (see, e.g., Wood 2020).

As noted in Chapter 1, the status of phonemic vowel length is uncertain across the Enlhet-Enenlhet family. Because all authors agree that Enenlhet does not contain a phonemic length distinction, this study did not anticipate finding one. However, given the historical link between vowel quality and phonemic length, vowel quality may be an important part of the historical trajectory of phonemic length. Enenlhet may represent a starting point from which other languages developed phonemic length (if it is an innovation), or it may show remains of a length system, if the system is original and has been lost in Enenlhet. Phonemic length would surface in this analysis as systematic errors associated with specific lexical items, once other factors have been accounted for.

Though a phonemic length distinction is not expected in Enenlhet, there are two categories of vowels in the analysis that were expected to have different lengths. Recall from Chapter 1 that vowels of identical quality separated by a glottal stop were not segmented. That is, orthographic <a'a, e'e> and <o'o> sequences were each segmented as single intervals. Chapter 4 provides a further discussion of these sequences. Therefore, a VOWEL LENGTH factor in which each vowel was categorized either as "plain V" (singleton <a, e, o>) or "VʔV" (<a'a>, <e'e> or <o'o>) was included. I expected the VʔV cases to be longer than the plain V ones. Either these have a greater inherent length for some reason (phonemic length, voice quality changes), or they are sequences of multiple segments; both options suggest longer VʔVs. Though I remained ambivalent about whether to call the VʔV cases "long vowels" I called this distinction VOWEL LENGTH to disambiguate it from VOWEL QUALITY which refers to the letter used to write the vowel, mapping on to the vowel's F1 x F2 values.

2.1.6. Syllable Structure

Cross-linguistic literature, both synchronic and diachronic, suggests that vowels in open syllables are longer than vowels in closed syllables. Diachronically, open syllable lengthening is responsible for the development of phonemic length contrasts in some languages (e.g., Middle French; Loporcaro 2015). The same pattern is attested in synchronic studies. Maddieson (1985) finds phonologized vowel shortening in Arabic, Estonian, Hausa, Hindi, and Ulithian, and a similar effect on the phonetic level in Dogri, Finnish, Icelandic, Italian, Norwegian, Rembarrnga, Shilha, Sinhala, Tamil, and Telugu.

Likewise, Benguerel (1971) reports that French stressed (phrase-final) vowels lengthen more in open syllables than closed ones; Buckley (1991) finds closed-syllable shortening in Kashaya; and Younes (1995) reports closed-syllable shortening in Cairene, Palestinian, and Modern Standard Arabic. The cross-linguistic variability in the syllable structure effect suggests that it is likely to interact with the effect of the voicing of the following consonant and with pre-pausal lengthening.

Within the Enlhet-Enenlhet family, open syllables in Sanapaná trigger realization of a phonetically long vowel (van Gysel p.c. 2021). Though there is variability across languages, most of the extant literature suggests that vowels tend to be longer in open syllables compared to closed ones, which is therefore the expected effect in Enenlhet.

Syllable structure was calculated using a Python script which parsed each individual word into syllables and then located the vowel within that structure. This script used the phonotactic rules discussed in Chapter 1 to develop syllable parsings. It classified a consonant as an onset whenever possible, but it avoided tautosyllabic consonant clusters. So, for example, it produced V.CV, CV.CV, and CVC.CV parsings. Complex "onsets" or "codas" were only permitted word finally, where the second consonant is /ʔ/. For the OPEN SYLLABLE factor, each vowel was labelled as either belonging to an open/light syllable (V, CV) or a closed/heavy syllable (VC, CVC(C)).

2.2. Measurement

Once vowel boundaries had been adjusted (see Chapter 1) and vowels had been annotated based on the effects that are predicted to affect vowel duration, vowel durations in milliseconds and the lexical item in which each vowel occurred were extracted using a Praat script.

2.4. Exclusion Criteria

Vowels were divided by speaker and quality, and Z-scores were calculated for each token within these subsets. Tokens with a Z-score whose absolute value was greater than three were eliminated. This Z-score corresponds to the 99.8% confidence interval of a Normal distribution; that is, it eliminates only very extreme outliers. This process resulted in removing 318 tokens, or 2% of the corpus. These outliers were removed to avoid including overly long vowels resulting from speech hesitations or vowels drawn out for narrative effect.

3. RESULTS

Once outliers had been removed, the remaining corpus contained 15,365 tokens, distributed as seen in Table 3.1.

Table 3.1: Enenlhet vowel duration corpus

SPEAKER	N LEXICAL ITEMS	N TOKENS
CA	1230	8228
ER	204	879
LF	102	396
LM	154	668
MM	170	671
MR	121	480
MRR	411	3293
TF	202	750
TOTAL		15,365

The speakers can roughly be divided into three groups: small number of tokens (LF, MR, <500 tokens), medium number of tokens (ER, LM, MM, TF, 500–900 tokens), and large number of tokens (CA, MRR, >3000 tokens).

Not all vowel qualities are represented equally within the corpus; Table 3.2 shows the number of tokens of each vowel quality.

Table 3.2: Number of tokens of each vowel quality

QUALITY	N LEXICAL ITEMS	N TOKENS
/e/	1,498	5,557
/a/	1,721	7,943
/o/	552	1,865

As was observed in Chapter 2, /a/ outnumbers both /e/ and /o/, and /o/ is by far the least frequent of the three qualities. It appears both in fewer tokens and in only about a third as many distinct lexical items as /e/ and /a/.

As in Chapter 2, alpha for the statistical analysis was set to 0.01 to reduce the chance of false positives, given the large variability of the data and the many factors that affect duration. Section 3.1 presents the stepwise process of model building, and Section 3.2 presents the results of the overall model. Section 3.3 discusses inter-speaker variation. As in Chapter 2, duration values were log-transformed in data visualizations, but the un-transformed values were used for the statistical models.

3.1 Model Construction

As in Chapter 2, the overall model was built stepwise. All preliminary models contained random intercepts for SPEAKER and LEXICAL ITEM and one other factor. A factor was only included in the overall model if it had a significant main effect in the preliminary model ($p<0.01$). The VOWEL QUALITY, WORD POSITION (RIGHT), WORD POSITION (LEFT), and PRE-PAUSAL WORD factors were re-coded based on the results of these preliminary models, as discussed below. Discussion of the procedure used to determine which interactions and random effects to include in the overall model appears after discussion of each of the fixed effects.

The preliminary model containing only VOWEL LENGTH as a fixed effect shows this effect to be significant, so it was included in the maximal model.

The preliminary model to investigate VOWEL QUALITY showed a significant effect of both /a/ and /o/, suggesting that both are different from /e/ (the baseline). Post hoc pairwise comparisons show that /e/ is significantly shorter than both /o/ and /a/ ($p<0.0001$), but /o/ and /a/ do not differ ($p=0.96$). These pairwise comparisons provide language-internal motivation for a distinction between front and non-front vowels, which is not the cross-linguistically anticipated effect of vowel quality on vowel duration. However, because /a/ and /o/ do not differ in duration, VOWEL QUALITY was recoded into a binary FRONTNESS factor: /e/ was marked as FRONT and /o, a/ were marked as NON-FRONT. FRONTNESS had a significant effect in a preliminary model containing only only this fixed effect, so FRONTNESS was included in the maximal model.

Determining which word and utterance position variables to include in the model was more involved. PRE-PAUSAL WORD on its own was not really expected to have a significant effect on vowel duration, except insofar as it interacts with a vowel's position within a word. Figure 3.1 shows the interaction between WORD POSITION (RIGHT) and PRE-PAUSAL WORD.

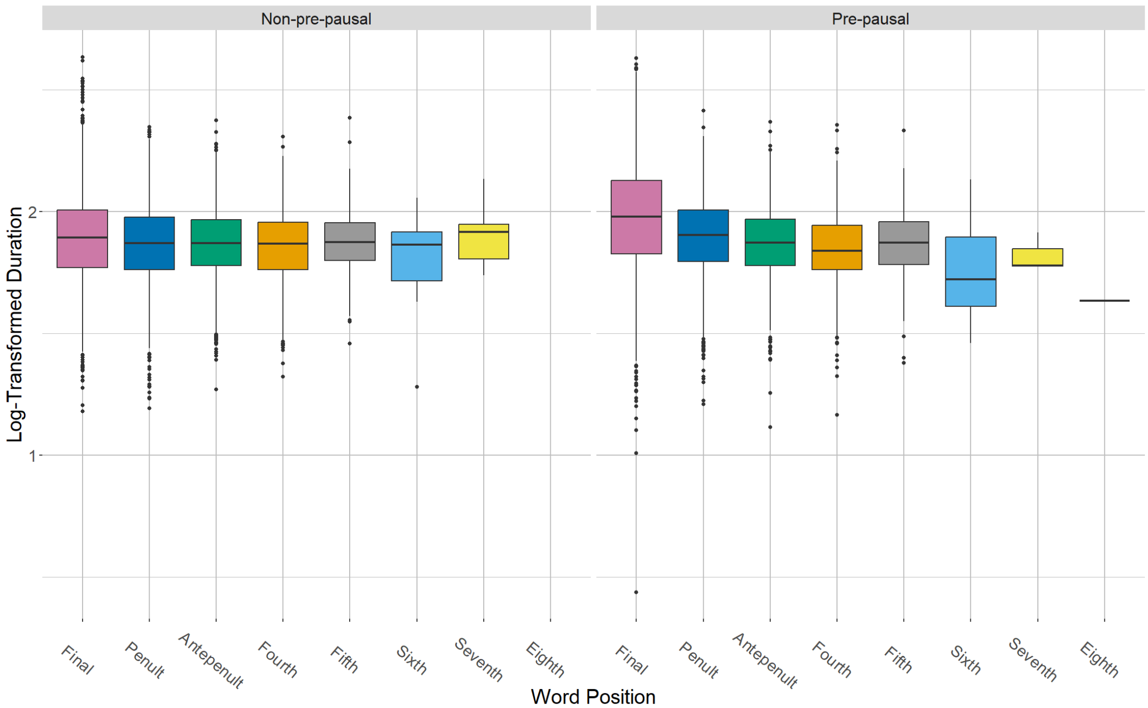

Figure 3.1. Relationship between WORD POSITION (RIGHT) and PRE-PAUSAL WORD

Figure 3.1 suggests that, in non-pre-pausal words (on the left), the mean durations across all positions are roughly the same. In contrast, the final vowel (pink) in pre-pausal words (on the right) is substantially longer than the vowels in other syllables.

Because Figure 3.1 suggests that WORD POSITION (RIGHT) interacts with PRE-PAUSAL WORD, a preliminary model was run with these two factors as two fixed effects, and a two-way interaction between them. As expected, this model showed a significant interaction between the two factors. Subset models were run to investigate the interaction. One model contained only non-pre-pausal words and the other contained only pre-pausal words. Both contained just WORD POSITION (RIGHT) as a fixed effect. For the model with pre-pausal words, pairwise comparisons show that the only significant differences are between final syllables and other syllables ($p < 0.001$).

In the model for non-pre-pausal words, pairwise comparisons show significant differences between final syllables and syllables 2–4, but not between other positions. This *could* indicate some kind of word-final lengthening effect. However, I think it's more likely that this is due to some words being mis-classified as non-pre-pausal; recall that my criterion for whether a word was pre-pausal or not was very conservative and treated words as non-pre-pausal unless the following gap was at least 250ms long. If the effect of WORD POSITION (RIGHT) were due to lexical stress, we would expect the stressed syllable to be longer than syllables in all other positions and probably for the difference between stressed and unstressed syllables to be bigger than what the model estimates (~ 5 ms).

If the effect of a vowel's position within the word is not due to stress, then the main goal is to account for the majority of the variability so that it can be ruled out down the road when looking for phonemic length. Figure 3.1 suggests that the variability due to position is primarily due to the difference between final vowels in pre-pausal words and vowels in other positions. Therefore, a model was run using just PRE-PAUSAL SYLLABLE as a fixed effect, which was significant.

It is also possible that stress attends to a vowel's position in a word as marked from the left edge, rather than the right. The same set of models was run using WORD POSITION (LEFT) instead of WORD POSITION (RIGHT). However, the models show only significant effects that are consistent with the previously-identified effect of PRE-PAUSAL SYLLABLE.

Finally, heavy syllables sometimes attract stress, regardless of their location. To eliminate the possibility of quantity-sensitive effects on stress, the same set of analyses as just described were run again on the subset of tokens from words containing only light (C)V syllables (n=2,911). If lexical stress is quantity sensitive but otherwise fixed, it should appear in the default location in words with only light syllables where there is no heavy syllable to attract it. However, the tokens from words with only light syllables did not show any independent effect of a vowel's position within the word.

Because the most prominent positional effect is that of PRE-PAUSAL SYLLABLE, the WORD POSITION (RIGHT), WORD POSITION (LEFT), and PRE-PAUSAL WORD factors were not included in subsequent models. Instead, the PRE-PAUSAL SYLLABLE factor, which was significant in its individual preliminary model, was used to capture the effect of pre-pausal lengthening.

The preliminary model to investigate syllable structure contained only the OPEN SYLLABLE factor, which was significant and therefore retained in the overall model.

Finally, a preliminary model with the FOLLOWING VOICELESS CONSONANT factor showed that this factor also had a significant effect and therefore should be included in the overall model.

As noted in Section 2, previous work suggests that the main interactions between the factors in this model will occur between the variables *not* related to vowel quality. That is, the effect of OPEN SYLLABLE may be modulated by FOLLOWING VOICELESS CONSONANT or PRE-PAUSAL SYLLABLE, and the effect of FOLLOWING VOICELESS CONSONANT may also interact with PRE-PAUSAL SYLLABLE. Therefore, these three factors, and all possible two-way interactions between them, were included together in a preliminary model. The interactions between PRE-PAUSAL SYLLABLE and both other factors were significant ($p<0.01$), but the interaction between FOLLOWING VOICELESS CONSONANT and OPEN SYLLABLE was not, so it was not included in the maximal model.

The model resulting from this stepwise process included FRONTNESS, VOWEL LENGTH, OPEN SYLLABLE, FOLLOWING VOICELESS CONSONANT, and PRE-PAUSAL SYLLABLE as fixed effects, and two-way interactions between PRE-PAUSAL SYLLABLE*OPEN SYLLABLE and PRE-PAUSAL SYLLABLE*FOLLOWING VOICELESS CONSONANT. The model did not contain any other interactions, as I did not have specific predictions based on cross-linguistic patterns about how FRONTNESS or VOWEL LENGTH might interact with the positional factors.

To determine if the random effects contribute substantively to the model, two additional nested models were created. One model removed the random intercept for SPEAKER and one removed the random intercept for LEXICAL ITEM. These models were then compared using a Likelihood Ratio Test using *lrtest* (Hothorn et al. 2022). Chi-squared tests comparing each nested model to the maximal model show that the random effects significantly contribute to the model ($p<0.01$). Therefore, the random intercepts for SPEAKER and LEXICAL ITEM were retained in the overall model.

Table 3.3 shows the set of factors that were ultimately used in the vowel duration model, along with their possible levels and baseline values.

Table 3.3: Factors and levels for final vowel duration model

TYPE	FACTOR	POSSIBLE LEVELS	BASELINE
RANDOM INTERCEPTS	SPEAKER	CA, ER, LF, LM, MM, MR, MRR, TF	–
	LEXICAL ITEM	–	–
FIXED EFFECTS	FRONTNESS	front, non-front	front
	VOWEL LENGTH	V, VʔV	V
	PRE-PAUSAL SYLLABLE	pre-pausal syllable, other	other
	OPEN SYLLABLE	open, closed	closed
	FOLLOWING VOICELESS CONSONANT	voiceless, other	other
INTERACTIONS	PRE-PAUSAL SYLLABLE*FOLLOWING VOICELESS CONSONANT		
	PRE-PAUSAL SYLLABLE*OPEN SYLLABLE		

3.2. Maximal Duration Model

Table 3.4 shows the results of the maximal model of duration. Significant results are shaded grey.

Table 3.4: Results from maximal model of duration.

```
Duration = (1|speaker) + (1|lexicalItem) + frontness + vowelLength + pre-pausalSyllable +
openSyllable + followingVoicelessConsonant + pre-pausalSyllable*openSyllable + pre-
pausalSyllable*followingVoicelessConsonant
```

FACTOR	ESTIMATE	s	d.f.	t	p
(INTERCEPT)	81.90	3.27	7.67	25.02	<0.001
VOWEL LENGTH: VʔV	47.69	2.34	6082.89	20.38	<0.001
FRONTNESS: NON-FRONT VOWEL	11.93	0.66	13948.54	18.03	<0.001
PRE-PAUSAL SYLLABLE	14.56	1.40	15082.87	10.40	<0.001
OPEN SYLLABLE	2.14	0.69	14745.98	3.10	<0.01
FOLLOWING VOICELESS CONSONANT	-9.70	0.67	15104.08	-14.38	<0.001
PRE-PAUSAL SYLLABLE * OPEN SYLLABLE	8.27	1.57	14.986.05	5.26	<0.001
PRE-PAUSAL SYLL * FOLLOWING VOICELESS CONS	-3.88	1.54	15326.85	-2.52	0.01

Figure 3.2 shows the difference between the plain V and VʔV tokens. Plain Vs, on the left, are shorter than the VʔV category on the right.

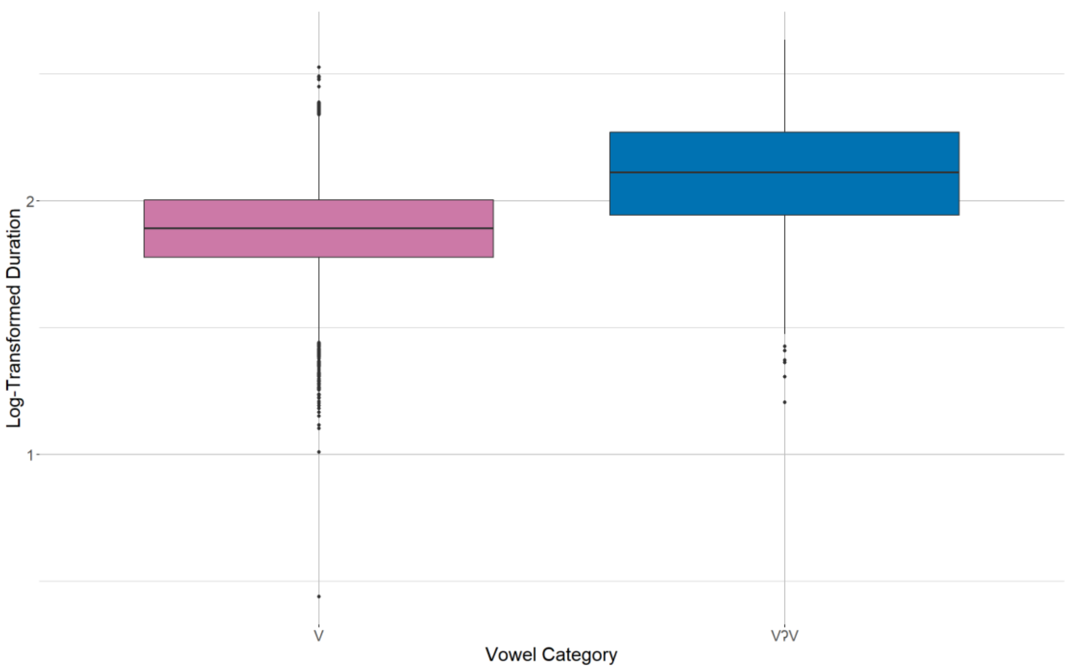

Figure 3.2. Duration differences due to VOWEL LENGTH

This difference is around 60 ms; the V tokens have a mean duration of 83.78 ms, and the VʔVs have a mean of 144.64 ms. The VʔV cases are, on average, not quite double the length of the plain Vs. The fact that the average duration of the VʔVs is not at least twice the duration of the plain Vs might suggest that they are not actually sequences of three segments. An analysis treating them as one segment with a greater inherent duration (perhaps due to the time needed to realize voice quality changes) would explain this. However, durational facts alone are not enough to make a definitive analysis.

FRONTNESS also has a significant effect on duration, though recall that a front-non-front contrast is not the effect of vowel quality that was initially anticipated. Inherent duration divides the three vowel phonemes into two categories: front versus non-front. The front vowel (/e/) is shorter on average. Figure 3.3 shows these two categories, averaged across all other variables.

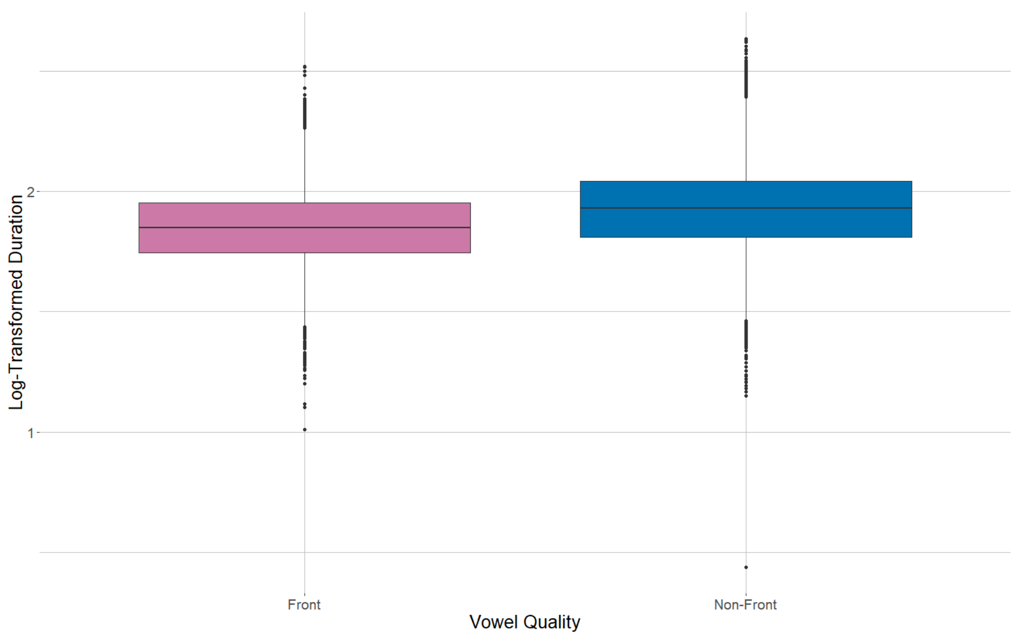

Figure 3.3. Duration differences due to FRONTNESS

Front vowel /e/ has a mean duration of 75.81 ms, compared to 92.82 ms for the non-front vowels /a/ and /o/. Since, as shown in Chapter 2, /e/ and /o/ are the same height, it is surprising that they apparently differ inherently in duration, which usually depends on vowel height.

Figure 3.4 shows the effect of PRE-PAUSAL SYLLABLE.

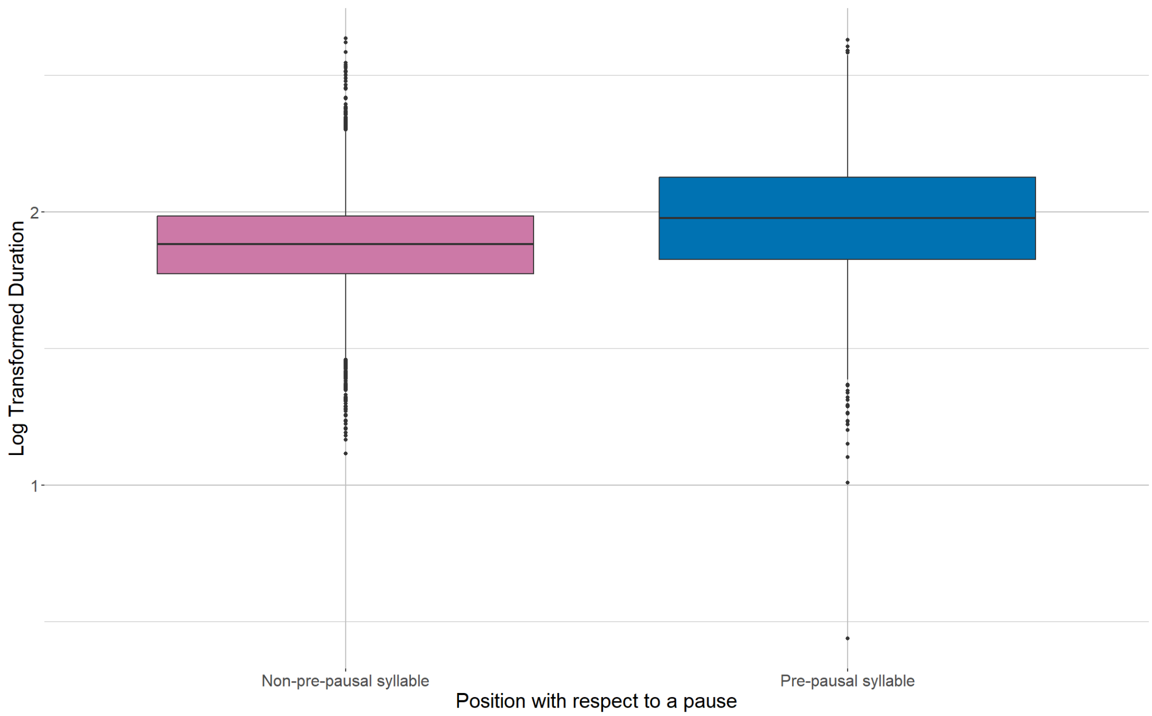

Figure 3.4. Duration differences due to PRE-PAUSAL SYLLABLE

In Figure 3.4, the pre-pausal syllables, in blue, have a greater mean (106.82 ms), which corresponds to the significant effect in Table 3.4. Non-pre-pausal vowels have a mean of 80.58 ms. This difference, about 26 ms, is smaller than the difference between the plain Vs and the VʔVs but roughly equivalent to the difference in duration between front and non-front vowels.

Table 3.4 also shows a significant effect of syllable structure. It indicates that vowels in open syllables are about 2.1 ms longer than vowels in closed syllables. Obviously, this difference is very small. There is also a significant interaction between this factor and PRE-PAUSAL SYLLABLE; vowels in open syllables are about 8 ms longer when that syllable is pre-pausal, compared to when it is non-pre-pausal. Figure 3.5 shows this interaction. The orange guidelines compare the mean duration of vowels in closed syllables to open ones, both pre-pausally and non-pre-pausally.

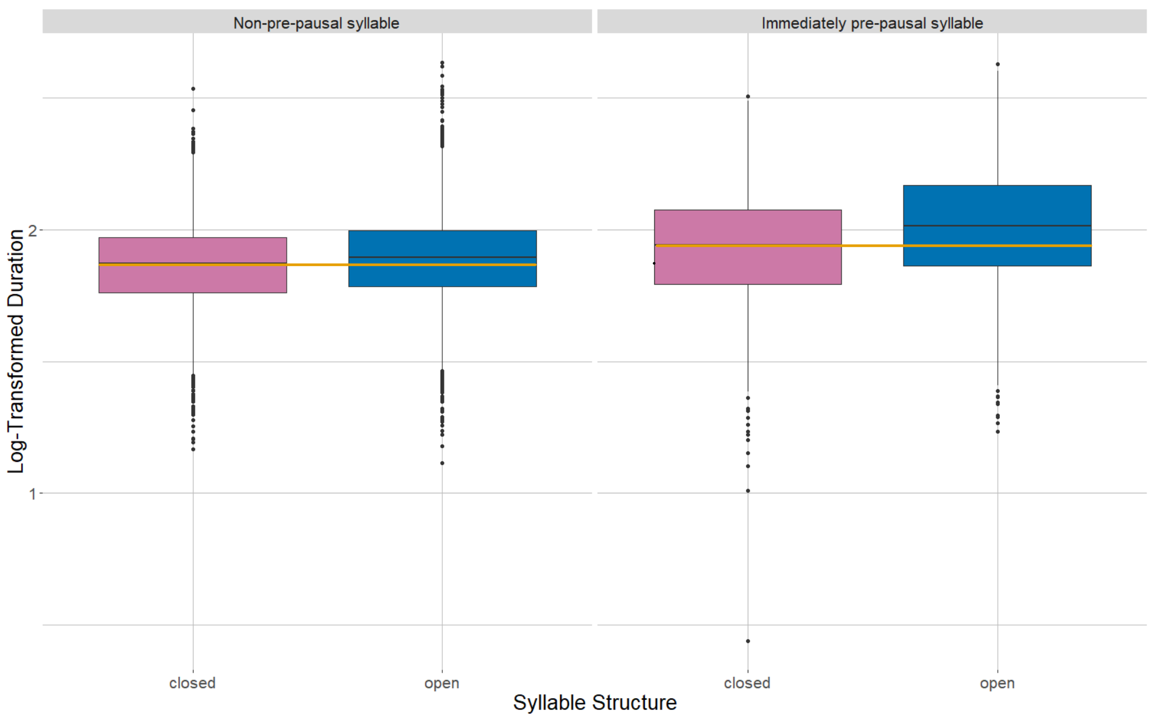

Figure 3.5. Vowel duration in open and closed syllables, grouped by PRE-PAUSAL SYLLABLE

In both pre-pausal and non-pre-pausal syllables, Figure 3.5 shows that vowels in open syllables are longer; the mean for vowels represented by the blue boxes (open syllables) is greater than for the pink (closed syllables). However, Figure 3.5 also indicates that the difference between open and closed syllables is greater pre-pausally. The mean duration of vowels in open, non-pre-pausal syllables is 83.45 ms, compared to 78.20 ms in closed syllables (difference of about 5 ms). In pre-pausal syllables, by contrast, vowels in open syllables have a mean duration of 117.14 ms, compared to just 95.96 ms in closed syllables (difference of about 21 ms).

The final effect shown in Table 3.4 is that of the voicing of the following consonant. The model suggests that vowels followed by voiceless consonants are about 10 ms shorter than vowels not followed by voiceless consonants. Figure 3.6 shows this effect.

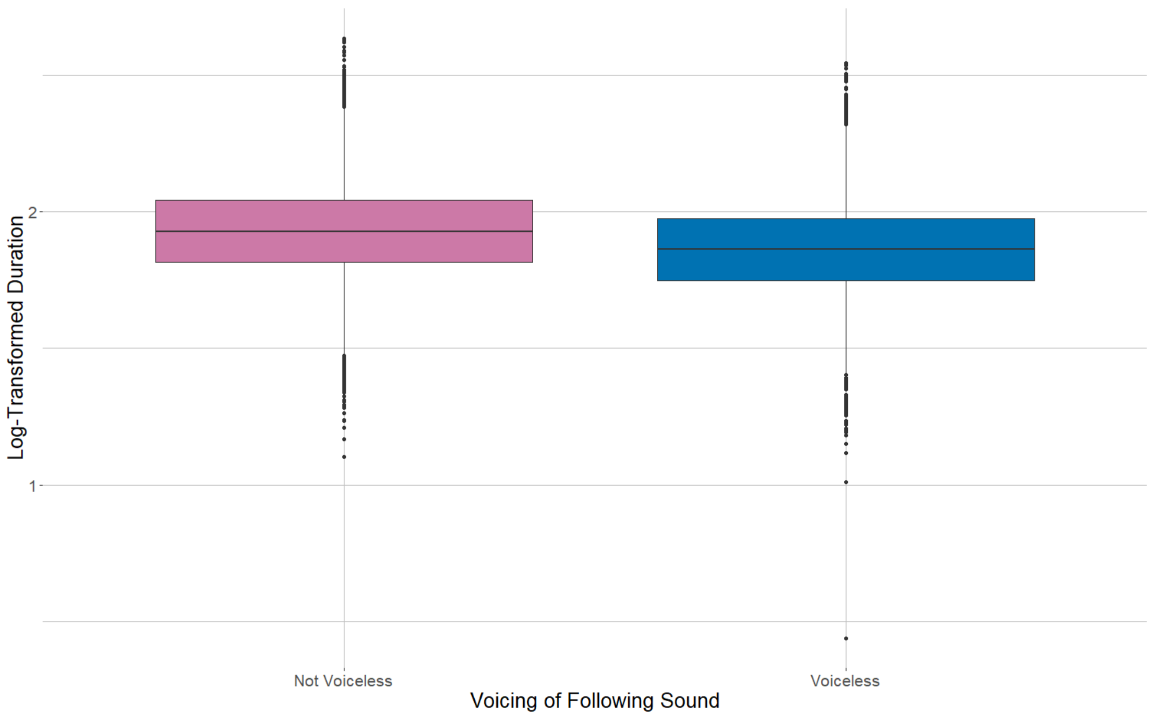

Figure 3.6. Duration differences due to FOLLOWING VOICELESS CONSONANT

The vowels followed by voiceless consonants appear in blue; their mean duration is 79.21 ms, compared to a mean duration of 93.05 ms for other vowels (followed by a voiced consonant, immediately before a pause, or followed by a vowel).

3.3. Individual Variation

The mixed effects model failed to converge when random slopes were added to account for speaker-level differences in these effects. However, each speaker's data were examined individually. The individual models had the same structure as the model in Table 3.4, with the random intercept for speaker removed. A summary of the results for each speaker appears in Table 3.5; significant results are shaded grey.

Table 3.5: Individual model results for each speaker

```
Duration = (1|lexicalItem) + frontness + vowelLength+ pre-PausalSyllable + openSyllable +
followingVoicelessConsonant + pre-PausalSyllable*openSyllable + pre-
pausalSyllable*followingVoicelessConsonant
```

SPEAKER	VʔV	NON-FRONT	PRE-PAUSAL SYLL	OPEN SYLL	FOLLOWING VCLESS C	PRE-PAUSAL SYLL * OPEN SYLL	PRE-PAUSAL SYLL* FOLLOWING VCLESS C
CA	***	***	***	0.02	***	0.07	**
ER	***	***	***	**	***	**	0.23
LF	***	0.06	0.96	0.46	0.03	0.41	0.78
LM	***	***	***	0.98	***	0.57	0.46
MM	***	***	0.02	0.57	0.20	0.06	0.57
MR	***	***	0.30	0.16	0.04	0.09	0.44
MRR	***	***	***	0.18	***	***	0.05
TF	**	0.20	0.01	0.05	***	0.13	0.33

Significance codes: ***: $p<0.001$, **: $p<0.01$

The effects of VOWEL LENGTH and FRONTNESS are robust across speakers. VʔV tokens are estimated to be between 36.57 ms (CA) and 114.00 ms (MR) longer than plain V tokens, and all speakers show a significant effect. All but two speakers distinguished front vowels from non-front vowels. CA showed the smallest significant difference, estimated at just 10.22 ms, compared to MRR, whose non-front vowels were estimated to be 17.42 ms longer than front vowel /e/.

The effects of PRE-PAUSAL SYLLABLE and FOLLOWING VOICELESS CONSONANT were less robust, appearing in only four and five speakers, respectively. The OPEN SYLLABLE effect is even less strong. This factor was signficant in the overall model, but only one speaker, ER, shows vowels in open syllables to be longer than vowels in closed syllables (by about 6.89 ms). Only two speakers (ER and MRR) showed significant interactions between PRE-PAUSAL SYLLABLE and OPEN SYLLABLE, and only CA showed a significant interaction between PRE-PAUSAL SYLLABLE and FOLLOWING VOICELESS CONSONANT.

Overall, the individual models indicate that the effects of VOWEL LENGTH and FRONTNESS are most robust, appearing for most speakers. The other effects are more variable, though vowel shortening before voiceless consonants also appears for half the speakers. The speakers also vary in their baseline speaking rates; CA is the fastest talker and MR is the slowest one.

4. DISCUSSION

The following sections discuss each effect on vowel duration considered in this study, beginning with the factors that contribute to the main questions about stress and phonemic length.

4.1. Stress

Stress, if present in Enenlhet, was predicted to appear as a significant effect of a vowel's position within a word; once other factors were controlled for, an independent effect of position was not found, providing no evidence for a fixed stress position in Enenlhet. However, this analysis does not account for several variables upon which stress may rely. Elliott's (2021) look at pitch and intensity as correlates of word-level prominence shows that in Enxet these cues depend on the morphological structure of the word, with different morpheme classes affecting pitch and intensity in different ways. If Enenlhet is similar, and different morphological categories have different stress-assignment patterns, the brute force analysis presented here will not have found them. Similarly, morphological classes that fall outside the stress domain would obscure a stress pattern, even if stress is assigned to a fixed position within the stress domain. Therefore, this study rules out fixed, lexical stress across the entire lexicon in Enenlhet, but it cannot conclusively argue that Enenlhet has no lexical stress at all. Rather, this description provides a starting point for a future investigation that accounts for factors not considered here.

4.2. Categorical Length Distinctions

All previous literature on Enlhet-Enenlhet languages agrees that Enenlhet does not make a phonemic length contrast (Unruh & Kalisch 2003, Elliott 2021), a conclusion that is supported by the results of this study. Phonemic length, if present, would surface in this analysis as a systematic effect of lexical item. Because the random intercept for lexical item significantly improved the maximal statistical model, the possibility of phonemic vowel length was further investigated through a comparison to Enxet, which all descriptions agree does have a length contrast. A list of lexical items with long vowels was extracted from Elliott (2021) using a Python script. This list was compared to the lexical items in the Enenlhet corpus, and probable cognates were extracted.[12] From this set, bisyllabic words whose first syllable contains a long vowel in Enxet were selected; see Table 3.6.

[12] Enenlhet cognates could be identified even without translations for every word in the Enenlhet database because of the lexical similarity between the two languages. In most cases, the only difference between the Enxet and Enenlhet was the presence of an orthographic long vowel in Enxet. Only lexical items which were certain cognates were included in this analysis.

Table 3.6: Data subset used for Enxet-Enenlhet cognate comparison

VOWEL QUALITY	N SPEAKERS	N LEXICAL ITEMS	N TOKENS
/a/	8	27	652
/e/	8	20	236
/o/	4	5	17
TOTAL	8	36	905

A model with the following structure was run on this subset to investigate whether vowels which are cognate with Enxet long vowels are longer than vowels in Enenlhet that are cognate with Enxet short vowels.[13]

```
Duration = (1|speaker) + (1|lexicalItem) + vowelQuality + wordPosition(right) +
prePausalWord + openSyllable + followingVoicelessConsonant +
wordPosition(right)*prePausalWord + wordPosition(right)*openSyllable+
wordPosition(right)*voicelessFollowingConsonant + prePausalWord*openSyllable
```

Since all the vowels that are cognate with Enxet long vowels are located in the first syllable of these disyllabic words, an effect in this model of a vowel's position within the word would be equivalent to an effect of phonemic length. If phonemic length is present in Enenlhet, first syllables will be longer than second syllables. There were no significant effects or interactions involving word position in this model, indicating that vowels in first syllables (cognate with Enxet long vowels) were about the same in duration as vowels in second syllables (cognate with Enxet short vowels).

I also ran the maximal model (from Table 3.4) on the subset of Enenlhet data presented in Table 3.6 and compared the model's predicted values to the actual observed values.[14] This comparison is shown in Figure 3.7.

[13] This model removes VOWEL LENGTH (V vs. VʔV) as all the tokens were orthographically plain Vs. It includes WORD POSITION (RIGHT) and PRE-PAUSAL WORD as separate factors. Interactions between WORD POSITION (RIGHT) and the other positional factors were included because WORD POSITION (RIGHT) was the main factor of interest. An interaction between PRE-PAUSAL WORD and OPEN SYLLABLE was also included because the interaction between PRE-PAUSAL SYLLABLE and OPEN SYLLABLE was significant in the overall model.

[14] As for the preceding model, VOWEL LENGTH was removed because all vowels in in the subset of words selected as cognates with Enxet words were plain Vs.

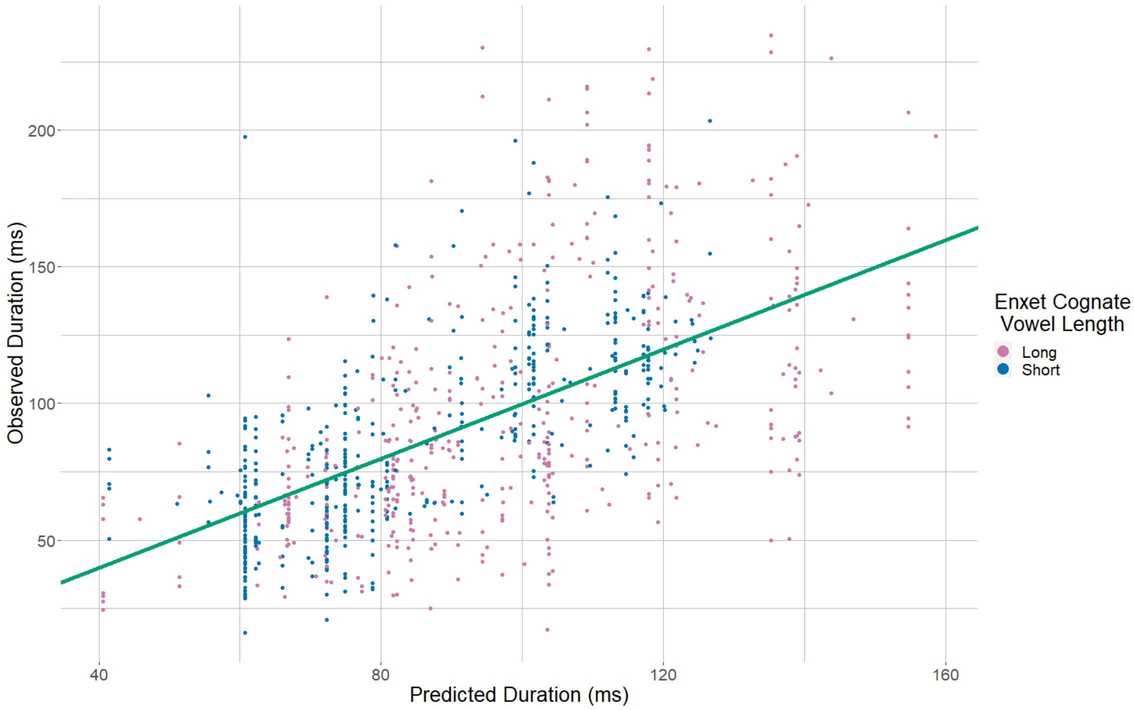

Figure 3.7. Comparison of predicted and actual durations of Enenlhet vowels from words with Enxet long vowel cognates

In Figure 3.7, if vowels that are cognate with Enxet long vowels had consistently greater durations than the model predicted, the pink points would cluster above the green trend line. Instead, we observe a relatively even scatter of both pink and blue dots, corresponding to Enxet long and short vowels, above and below the trend line.

As expected, these two examinations provide no evidence for a phonemic length contrast in Enenlhet. However, the random effect of lexical item contributed to the maximal model, suggesting that the word in which a vowel appears has some effect on the vowel's duration independently of the other factors. Whether this effect is due to a near-merger of historically contrastive vowel length, a nascent length contrast, or some other factor like syllable quantity not captured by the open-closed distinction is unclear.

There was one categorical length distinction in the corpus. As anticipated, the VʔV tokens were robustly longer than plain Vs, as we would expect them to be if they were sequences of three segments. However, the VʔVs were not, on average, twice as long as the plain Vs. As mentioned in Section 3.2, these VʔVs may also be one segment with a different inherent length for some reason other than vowel quality; for example, a phonemically specified voice quality change may result in a longer intrinsic length (see, e.g., Frazier 2009: 41). Comparison of glottalization in these tokens with the glottalization triggered by orthographic glottal stop more broadly shows that the VʔV tokens have more glottalization than the plain Vs, but they do not categorically differ from vowels with glottal stops on just one side or another. See Chapter 4 for a more thorough discussion of the VʔVs.

4.3. Pre-pausal Lengthening

As expected based on virtually all cross-linguistic literature on vowel duration, Enenlhet exhibits pre-pausal lengthening. This effect is also the largest effect in the statistical model. It is apparently limited to the immediately pre-pausal syllable, as the model did not show word-position effects indicating progressive lengthening. This pre-pausal lengthening is, however, still relatively small compared to other languages in which this effect has been studied. For example, Berkovits (1994) reports final vowel lengthening of 28 ms (58%) for unstressed final syllables, and 116ms (57%) lengthening for stressed final vowels. Cambier-Langeveld (1999) reports final lengthening for unaccented monosyllabic names (whole syllables) around 121.2 ms in Dutch and 141.5 ms in English. In Finnish, Nakai et al. (2009) report final lengthening ranging from 22 ms to 88 ms, depending on syllable type. The effects in Enenlhet are much smaller, though the pre-pausal lengthening was significant across four of the eight speakers. The magnitude of this effect may suggest that Enenlhet listeners do not use it as a perceptual cue to syntactic or morphological boundaries as speakers of other languages do.

If pre-pausal lengthening is not available as a prosodic cue to syntactic or morphological constitutents, it raises the question of what cues Enenlhet speakers use instead. One possibility is that morphological markers of boundaries are so robust that prosodic marking is not needed. Since the basic word order in Enlhet-Enenlhet languages is verb-first, it is possible that verbal person-marking prefixes are sufficient markers of the beginnings of major syntactic constituents. Alternatively, pre-pausal lengthening may be one of several acoustic cues to phrase boundaries, working in tandem with prosodic features like voice quality or pitch. Wheeler (2020) observes that single word utterances in Angaité often exhibited a fall-rise F0 pattern, which may function as a boundary tone. If a similar pattern is present in Enenlhet, speakers may use both cues together to identify constituent boundaries.

4.4. Following Consonant Voicing

As in other languages, vowels in Enenlhet were predicted to be shorter before voiceless consonants, and this effect did appear. Once again, the effect was quite small, estimated to be 10.37 ms. In comparison, Chen (1970) finds an average lengthening of 92.29 ms in English, 47.44 ms in French, 31.5 ms in Russian, and 27.3 ms in Korean; Letterman (1994) finds an average duration difference of 16 ms between vowels preceding voiceless versus voiced consonants in Sinhala.

Speakers are relatively more sensitive to changes in vowel duration compared to consonant duration (Huggins 1972), at least in English, and the just noticeable difference

(JND) for vowels depends on the total length of the segment as well as its position. However, Klatt and Cooper (1975) find JNDs in English ranging from 22–98 ms; the consonant voicing effect in Enenlhet is well below this threshold, which suggests that it is likely not perceptible to Enenlhet speakers. Since voicing conditioned vowel duration is unlikely to serve a perceptible (phonologized) purpose in Enenlhet, it is most likely attributable to a coarticulatory effect.

4.4. Vowel Quality

The effect of vowel quality on duration in Enenlhet was unpredicted. Most studies of the relationship between quality and duration have found that lower vowels have a longer inherent duration than higher vowels. Since Enenlhet has one low vowel /a/ and two mid vowels /e, o/, the mid vowels were predicted to be shorter than the low vowel. In this regard, /e/ patterned as predicted, being significantly shorter than /a/ for six speakers. However, post hoc analysis showed that /a/ and /o/ did not differ.

Chapter 2 shows that this duration pattern cannot be attributed to a height difference. Enenlhet makes only two distinctions in F1 (height); /e/ and /o/, the non-low vowels, are significantly different from /a/, but not from each other. The difference in duration found here may be a quirk of the lexical items included in the corpus; /o/ is the least frequent quality in the language, so it is possible that these apparent differences are due to the unbalanced sample. If /o/ appears primarily in open syllables or words that are usually pre-pausal, these two factors may confound examination of the effect of quality itself.

Though less cross-linguistically common, there is at least one other case of inherent duration differing based on FRONTNESS. Letterman (1994) reports that in Sinhala back /u:/

is shorter than non-back vowels /i:, a:/. In Enenlhet the difference is the opposite; front /e/ is shorter than non-front /a, o/. The condensed FRONTNESS did have a significant effect in the model, and it was also significant for six speakers, showing that, while cross-linguistically unusual, the front ~ non-front distinction is relatively robust.

4.5. Syllable Structure

As in many other languages, the effects on duration of differently sized constituents (syllables vs. phrases) are not strictly additive. As predicted, vowels in Enenlhet undergo open syllable lengthening, and this effect interacts with pre-pausal position. For vowels in pre-pausal position, open syllable lengthening is enhanced. The effect of open syllable lengthening was not very robust at the level of the speaker; the interaction between syllable structure and pre-pausal lengthening was only slightly more resilient, showing significant differences for two speakers. It's worth noting that two more speakers, TF and CA, would show a significant main effect of open syllables if alpha were set to 0.05 rather than 0.01.

5. CONCLUSION

This chapter presents an exploratory study of the phonetics of vowel duration in Enenlhet using a corpus of naturalistic monologues and interviews. It investigates the possibility of lexical stress in Enenlhet, which is a gap in previous descriptions of the language, and provides phonetic evidence supporting qualitative descriptions of Enenlhet as *not* having a phonemic length distinction (unlike its sister languages Enlhet and Enxet). The statistical analysis identified significant lengthening associated with immediately pre-pausal position, open syllables, non-front vowels, and VʔV sequences. Vowels preceding voiceless consonants are significantly shorter than vowels in other positions. Open syllable lengthening is enhanced in pre-pausal syllables. The effect of vowel quality is unexpected, and it does not map onto the low ~ non-low contrast seen in Chapter 2. Finally, the fact that VʔV sequences are longer than plain Vs, suggests that orthographic <V’V> sequences do behave differently from plain Vs – whether they should be treated as sequences of three segments or as longer vowels with an associated voice quality modification is discussed further in the next chapter.

Because this study is the first of its kind, not only for Enenlhet but for any Enlhet-Enenlhet language, the results necessarily require substantial further elaboration. The surprising effect of vowel quality on duration cannot be explained by vowel height. The unexpected effect of vowel quality on duration found here should be further examined through studies on samples with more balanced numbers of each vowel quality, as /o/ is less frequent in this corpus than /e/ and /a/. Though this study does not find any evidence for stress in Enenlhet, it does not rule out stress entirely. If stress is lexically assigned or the domain of stress assignment is not congruent with the orthographic word, it will not be apparent from this analysis. Further investigation of the possibility of lexical stress in Enenlhet will require levering a more detailed morphological analysis.

The effects found in this corpus are relatively small compared to those found in studies of other languages. This difference may be due to the nature of the data; many previous studies have relied on elicited or experimental data, which may be pronounced more carefully (and more slowly) than the naturalistic speech in the Enenlhet corpus. These differences in magnitude may also be genuine, language-specific differences, in which case future work should investigate whether speakers perceive these duration differences and whether they use them as perceptual cues to consonant voicing or constituent boundaries, as other languages have been shown to do. Additional work on unrelated languages in the area may also be informative as to whether the results of this study represent regional trends or family-specific patterns.

Chapter 4: Voice Quality

1. Introduction

Many of the recordings in the Enenlhet corpus contain non-modal voice qualities, including creaky voice, breathy voice, and de-voicing. There are no previous phonetic studies of glottal stop or voice quality in Enenlhet. However, as noted in Chapter 1, Elliott (2021) describes the participation of /ʔ/ in several complex morphophonological processes that mark it as unusual among the consonants of Enxet, and Wheeler (2020) notes voice quality modulations in Angaité that are not explained by the consonantal context. Though the distribution of glottal stop in Enenlhet is different than in Enxet (see Section 2), the unusual behavior of /ʔ/ in Enxet and of voice quality in Angaité motivate a close description of it in Enenlhet.

Since the other two studies included in this dissertation focus primarily on vowels, this study focuses narrowly on just one topic related to voice quality: orthographic /ʔ/ adjacent to vowels. There are three well-attested possible positions for a glottal stop adjacent to a vowel in Enenlhet: following (Vʔ), preceding (ʔV), or between two vowels of identical quality (VʔV).

Previous descriptions of Enenlhet treat the VʔV sequences as two vowels with a consonant between. This analysis is implicit in the Enenlhet orthography, which attempts to represent only phonemes of the language and writes these tokens as <a’a, e’e>, or <o’o> (Unruh, Kalisch, & Romero 2003: 300). This choice suggests that the /ʔ/ is analyzed as a phoneme in this position. However, Enenlhet /ʔ/ in general, and these VʔV cases in particular, have highly variable phonetic realizations, from complete closure, to extended periods of creaky voice, to no discernable change in phonation at all. This impression, based on the recordings from Heaton (2019–) is also consistent with Wheeler’s (2020) observation about glottal stops and Angaité voice quality. However, as noted in Chapter 1, these VʔV intervals are not always treated as sequences of three phonemes; Gomes (2013) argues that at least some of the glottal stops in Sanapaná are epenthetic, and Wheeler (2020) remains ambivalent about their status in Angaité. Furthermore, Chapter 3 provides some evidence that the VʔVs might not be a sequence of three phonemes; they are only about 1.5 times longer, on average, than the plain V intervals.

This study takes up the question of the VʔV intervals by comparing voice quality in these tokens to the effects on voice quality generated by a /ʔ/ in other positions with respect to the vowel. The main questions that this study explores are: 1) What are the acoustic features associated with glottal stops adjacent to vowels? 2) Do VʔV differ from Vʔ and ʔVs in terms of the timing of these cues?

The remainder of this chapter uses the term "glottalization" to refer to the acoustic effects on adjacent vowels of a (presumed) underlying glottal segment. I use it in lieu of a more specific term to avoid implying any particular acoustic characterization of it in Enenlhet, since cross-linguistic research suggests a variety of acoustic features associated with glottalization (see Section 3).

2. Glottal Stop in Enenlhet

2.1. Distribution of Glottal Stop

Many languages impose distributional restrictions on glottal stop, such as limiting it to certain positions within a syllable or requiring flanking vowels to be identical (Borroff 2007). In Enxet, glottal stop is avoided as an onset, resulting in deletions when it would otherwise appear as one (Elliott 2021: 97). However, glottal stop does appear robustly in coda position in Enxet, both as an apparently underlying segment (e.g., in many verb stems) and as a part of an epenthesized rhyme (see Elliott (2021: 98–102) for more discussion of the latter). Glottal stop can appear between identical vowels in Enxet, though in these cases it is sometimes deleted, resulting in a surface long vowel. It also appears between non-identical vowels, at least in some varieties of the language (Elliott 2021: 97).

The distribution of /ʔ/ in Enenlhet is different than in its sister language, but, like in Enxet, /ʔ/ patterns differently than the other Enenlhet consonants. Enenlhet allows glottal stop in (apparent) word-medial onset position and in word-final position after any sonorant. Example (9) shows it word-medially, and (10) shows it word-finally.

(9) *asvesai'a*
/as.we.saj.ʔa/
'I am called/my name is'

(10) a. *apsese'*
/ap.se.seʔ/
'his cookie' (?)

b. *aknem'*
/ak.nemʔ/
'day, sun'

c. *nenekev'*
/ne.ne.kewʔ/
'Laguna Porã' (place name)

Word medially, it is robustly attested in C_V environments, where the consonant is one of the following sonorants: /j/, /w/, /m/, /n/. It can never precede another word-medial consonant (*VʔC). There are also no cases of a vowel with a glottal stop on either side in my corpus (*/ʔVʔ/) or of onsetless syllables followed by a /ʔ/ (*#Vʔ, *V.Vʔ), regardless of whether a following /ʔ/ would be syllabified as a coda or an onset. These apparent restrictions on where glottal stop can surface as a coda are surprising, given the apparent preference for it as a coda compared to an onset in Enxet. Refer to Appendix B for a complete list of the medial consonant clusters in which /ʔ/ is attested.

Only one token in my corpus has glottal stop after an obstruent; *svet'ak* /swetʔak/ 'you see'. However, Unruh, Kalisch, and Romero (2003) include other examples of /ʔ/ following a /t/, as well as following word-medial /k/, shown in (11).

(11) *mak'ak*
/makʔak/
'querer hacer, ir a hacer algo / want to do, go to do (stg.)'
Unruh, Kalisch, & Romero (2003: 46)

In some contexts, the final rhyme can be removed from the verb in (11), leaving a word-final /kʔ/ sequence, as in *mak'* [makʔ] 'viajar / travel' (Unruh, Kalisch, & Romero 2003: 48).

Orthographic glottal stop /ʔ/ also appears very frequently between identical vowels; some examples are shown in (12).

(12) a. *mo'ok*
/moʔok/
'other'

b. *pa'at*
/paʔat/
'grass'

c. *tape'e*
/tapeʔe/
'chicken'

Glottal stop rarely appears between vowels of non-identical qualities, though see (13).

(13) *ha'e*
/haʔe/
'allá / there' Unruh, Kalisch, & Romero (2003: 36)

Example (13) is the only token with /ʔ/ between non-identical vowels, in both my corpus and in the pedagogical grammar. This scarcity suggests that Enenlhet strongly prefers for the vowels on either side of an intervocalic glottal stop to be of an identical quality (if these are indeed sequences of three underlying segments). This restriction contrasts with Enxet, where, though glottal stop is more restricted in its distribution within syllables, it apparently has more freedom in which sounds can be adjacent to it.

An auditory impression of word-initial glottalization is frequent, but it is possible that it is not indicative of an underlying glottal stop phoneme in this position. The Enenlhet orthography (which does not write word-initial <’>) is based on the Enlhet orthography, about which Unruh and Kalisch (1999: 5) say that "*Enlhet* only has syllables with structure CV(C) – if initial glottal stop is recognized as a phoneme that, nevertheless, does not need to be written due to its complete predictability…"[15] The unique orthographic treatment of word-initial [ʔ] raises the possibility that it behaves differently than [ʔ] in other positions, which may suggest a different phonemic status.

The other notable location for orthographic glottal stop in Enenlhet is word-finally after sonorant consonants, as in (10b) and (10c). The status of glottal stop here is also somewhat unclear. Enlhet has word-final glottalized nasals and glides which act as a single phoneme (Unruh & Kalisch 1999: 7). This analysis is represented in the Enlhet dictionary and elsewhere (Unruh & Kalisch 1997, Ritchie Key & Comrie, n.d.). Were this analysis applied to Enenlhet, the examples in (10b)–(10c) would be treated as ending with CVC syllables with a glottalized coda consonant. Campbell (2012b: 270) also describes Enlhet

[15] Original Spanish: "El *Enlhet* tiene solamente sílabas del tipo CV(C) – si se reconoce la glotal inicial como fonema que, sin embargo, no necesita ser escrito por su predictabilidad total…" (translation mine)

in this way; I suspect that he also relies on Unruh and Kalisch's materials. Elliott (2021: 114), in contrast, does not suggest phonemic glottalized sonorants for Enxet.

The Enenlhet data are currently inconclusive. If glottalized sonorants only appear word finally, this would suggest that their appearance is conditioned by prosodic factors rather than being phonemic. However, there are a limited number of onsetless, word-medial syllables (n=14) in the corpus. Given this, the orthographic <'> in examples such as (9) could theoretically be part of a glottalized sonorant rather than treated as an onset to the following syllable. Furthermore, word-final sonorants are not always followed by an orthographic glottal stop, which suggests a contrast between the two. For example, Unruh, Kalisch, and Romero (2003) present: *iemmen* /jemmen/ 'water' (p. 24), *piapom* /p-japom/ 'your/his father' (p. 32), *peletav* /peletaw/ 'spoon' (p. 36). This study examines only vowels and therefore does not investigate the status of <'> after sonorants; it treats any orthographic <'> as an underlying /ʔ/ consonant.

To sum up, glottal stop in Enenlhet appears word-finally after any sonorant. Word-medially, it appears either between two vowels of identical quality or after /j, w, m, n/. Vowels cannot have a glottal stop on either side, glottal stop cannot follow an onsetless syllable, and it cannot precede another word-medial consonant. The VʔV (<a'a, e'e, o'o>) cases represent one of the two main environments for word-medial glottal stop, but their analysis based on descriptions of related languages is not fully clear.

2.2. Qualitative Examination of Glottal Stop

There are some cases of complete closures associated with glottal stop, as in Figure 4.1, which shows a period of silence followed by a small release burst in the final syllable of *nenekev'* /nenkewʔ/ 'Laguna Porã (place name)' (release burst circled on waveform).

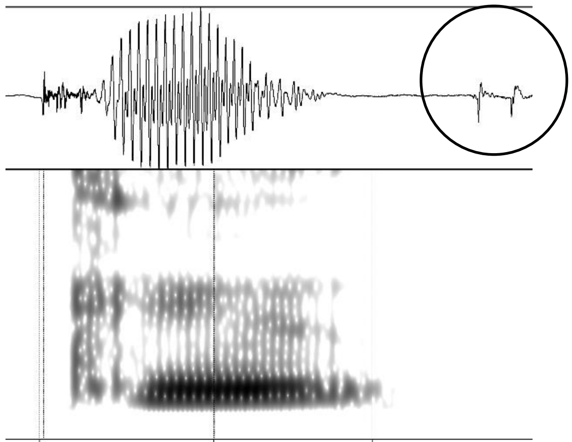

Figure 4.1. Full closure realization of glottal stop in *nenekev'* /nenekewʔ/ [ER; 70.020466s]

However, the most frequent realizations of /ʔ/ involve voice quality modulations of some kind. Sections 2.2.1 through 2.2.3 show some possible realizations of each of the four contexts for vowels with respect to glottal stop examined in this study: V, ʔV, Vʔ, and VʔV.

2.2.1. Plain V (Modal Voice) Realizations

Some of the speakers in the corpus are elderly, so some amount of creaky voice quality is expected even in vowels with no adjacent glottal segments. The degree of creak varies from speaker to speaker. Figures 4.2 through 4.4 show examples of /a/ with no adjacent glottal segments from three different speakers with varying baseline levels of creakiness.

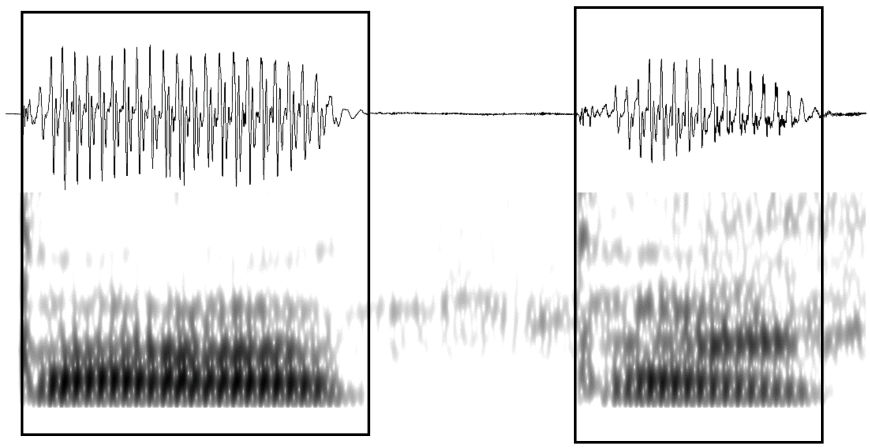

Figure 4.2. Two cases of /a/ with no adjacent glottal stop, from *tata* /tata/ ‘father’ [CA; 119.7635s]

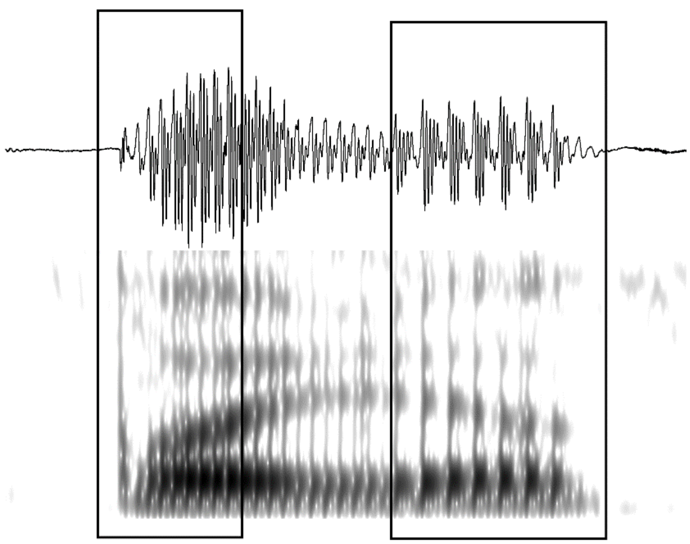

Figure 4.3. Final two /a/ from *askempaiak* /askempajak/, with no adjacent glottal stop [LF; 43.541s]

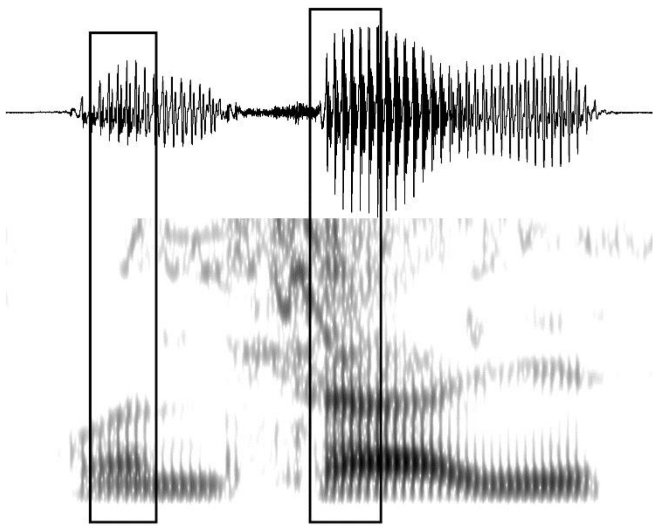

Figure 4.4. Vowels in *vanlha* /wanɬa/ [MM; 74.94381s]

The vowels in Figures 4.2–4.4 show varying intensities, some of which may be due to their phrasal position or adjacent consonants, and varying levels of aperiodicity. Notably, Figure 4.3 (speaker LF) shows a voice quality in the vowel after the glide that looks quite similar to what we might expect from an adjacent glottal stop (slow glottal pulses, sharply attenuated amplitude in each period). Figure 4.2 (speaker CA) shows a somewhat glottalized production of the second vowel, which has a much lower intensity and much less clear formants compared to the first vowel in the word.

2.2.2. ʔV and Vʔ Realizations

Preceding and following /ʔ/ typically result in voice quality changes located around the edge of the vowel nearest the glottal segment. Figure 4.5 shows an example of a medial glottal stop in a ʔV token realized as extreme aperiodicity and low intensity in the word *askelvetai'a* /askelwetajʔa/. The low intensity portion of the vowel is circled on the waveform. Note that it is preceded by two very slow, low intensity glottal pulses which could be categorized as the glottal stop itself if we were inclined to insist on a strictly linear realization of segments.

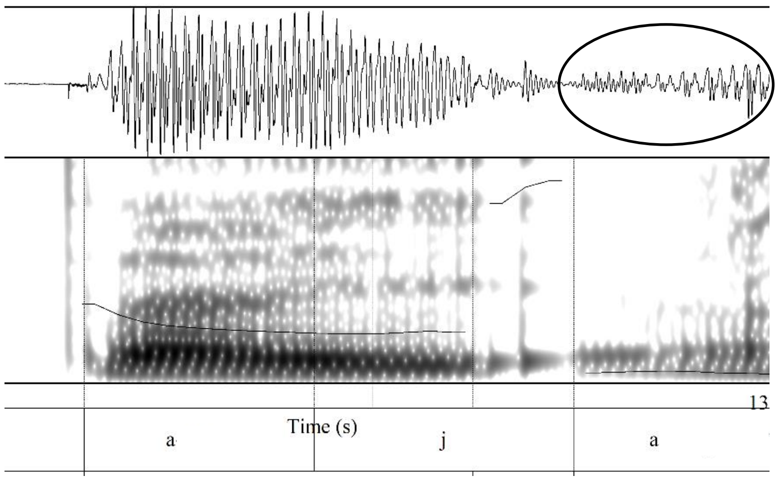

Figure 4.5. Glottalized realization of medial glottal stop (ʔV) from *askelvetai'a* /askelwetajʔa/ [ER; 137.487196s]

Figure 4.6 shows aperiodicity on a vowel preceding the glottal stop in *taiepe'* /tajepeʔ/. In this case, the Vʔ token shows sharply attenuated periods with very low intensity. The Praat pitch tracker (indicated by the black line on the spectrogram) clearly failed to accurately track the pitch during this vowel.

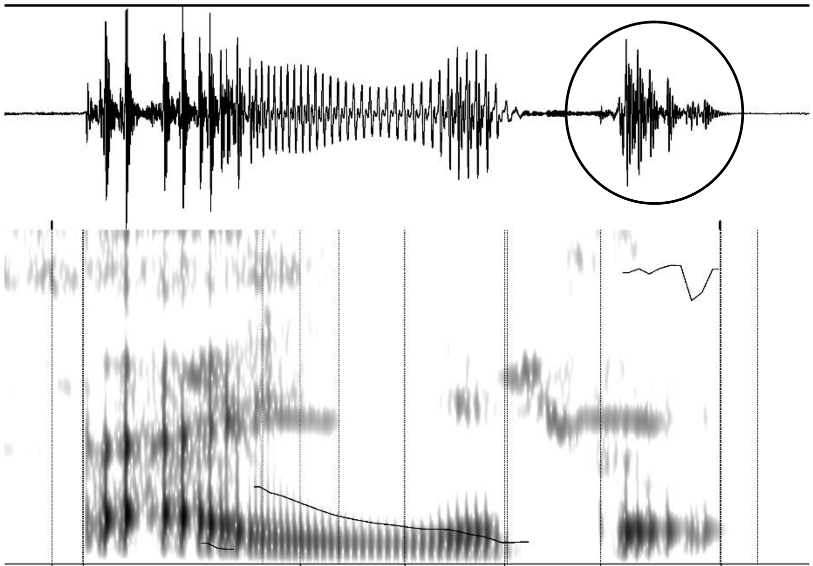

Figure 4.6. Glottalized realization of word-final glottal stop on preceding vowel (Vʔ) in *taiepe'* /tajepeʔ/ [MM; 409.419929s]

Figure 4.7 shows another ʔV token, on the left, followed by a Vʔ one, on the right. Both are exemplars of the low vowel /a/. The juxtaposition of these two cases shows the typical localization of the glottal effects to the edge of the vowel nearest the glottal stop.

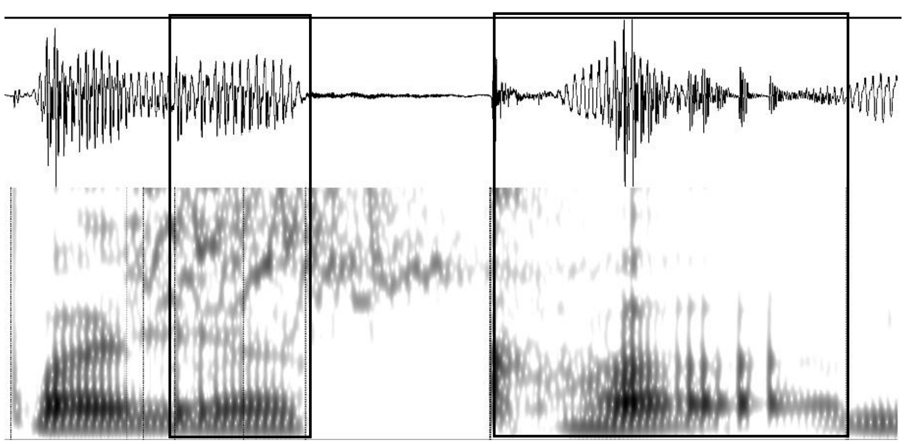

Figure 4.7. Vowels preceded (left) and followed (right) by a /ʔ/ [ER; 69.117448s]

In the ʔV case, there is very little, if any, aperiodicity visible in the spectrogram or the waveform. At most, there are one to two slightly slower glottal pulses at the extreme left edge. However, this vowel is quite low intensity in comparison to the preceding one (at the left edge of the spectrogram). Its intensity is more similar to the preceding /j/. The righthand Vʔ case shows much more aperiodicity, combined with low intensity and sharply attenuated pulses.

Figure 4.8 shows another ʔV case, this time in a word-final syllable. The spectrogram shows slow glottal pulses, and the waveform shows the triangular periods which are seen in many of these glottalized vowels. Both these acoustic features are primarily located in the first half of the vowel. The second half of the vowel looks to be basically modal (compare to examples in Section 2.2.1), albeit with a low intensity.

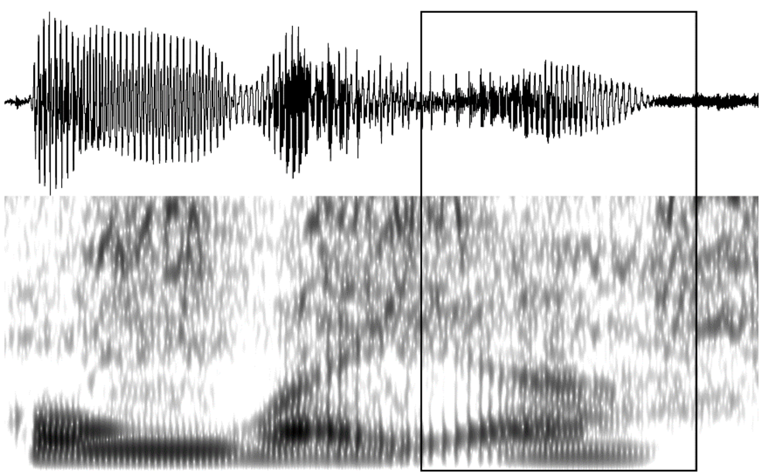

Figure 4.8. Final vowel (ʔV) in *nemaheiangkongvai'a* /nemahejaŋkoŋwajʔa/ [TF; 168.3366333s]

Finally, Figure 4.9 shows another Vʔ token from a word-final vowel. The vowel is very short, with only about four glottal pulses visible in the spectrogram. The last period of the waveform clearly shows a sharply attenuated period similar to those seen in the first half of the vowel in Figure 4.8.

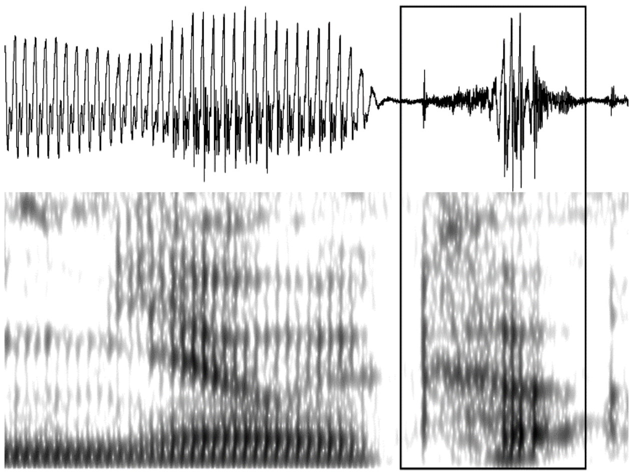

Figure 4.9. Final /a/ in *nengiavekha'* /neŋjawekhaʔ/ [MR; 169.309003s]

The vowels in Figures 4.5–4.9 are in a variety of consonantal contexts, but all were taken from non-pre-pausal words. The extent of aperiodicity and lowered intensity varies from speaker to speaker and vowel to vowel. For example, in Figure 4.9, the vowel is so short that the interval only contains four glottal pulses, and determining where aperiodicity begins is difficult, though the last period is further from the other three and is more sharply attenuated. In contrast, the vowel followed by /ʔ/ in Figure 4.7 shows a long period of aperiodic voicing in the second half of the vowel. The vowels preceded by glottal stop are similar. There is a relatively long period of low intensity glottal pulses in Figure 4.8, and a shorter, more aperiodic interval in Figure 4.7. In each of these cases, the glottalization is most notable toward the edge of the vowel adjacent to the glottal stop.

2.2.3. VʔV Realizations

The VʔV tokens are also realized in a variety of ways, though glottalization tends to be located toward the middle of the interval. Figure 4.10 shows a case with a decrease in intensity around the middle of the vowel (circled), accompanied by irregular and sharply attenuated glottal periods.

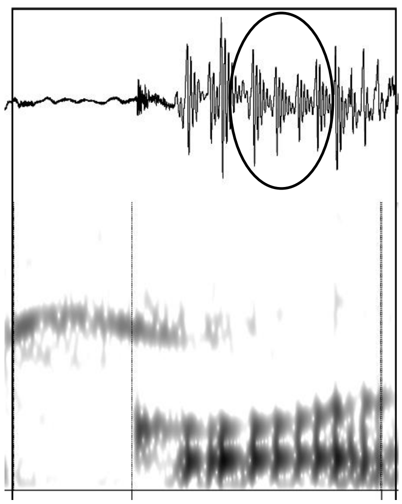

Figure 4.10. Glottalized realization of *o'o* in *ko'o* /koʔo/ [MM; 276.680631s]

The circled portion of the waveform in Figure 4.10 corresponds to the three lowest amplitude periods in this interval. The spectrogram, on the other hand, shows approximately evenly spaced, but very sharply attenuated glottal periods throughout the entire duration.

Figure 4.11 shows the same word with an intensity decrease in the middle of the *o'o* interval but no visible aperiodicity or spectral irregularity.

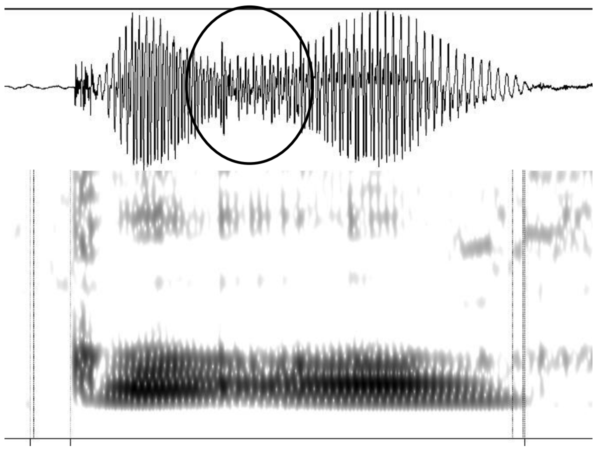

Figure 4.11. Glottalized realization of *o'o* in *ko'o* /koʔo/ [ER; 42.614s]

In Figure 4.10, the entire interval is relatively short, with aperiodicity and lower intensity throughout. In Figure 4.11, the portions on either side of the circled part of the waveform appear to be totally modal, with high intensity, clear formants, and regular periods. The circled intensity dip, which is also visible in the lighter portion of the spectrogram, is the only indicator of the /ʔ/.

In contrast, Figure 4.12 shows a complete closure realization of *a'a* in *pa'ang* /paʔaŋ/. Here, the second half of the interval also has a much lower intensity. The closure is about 19 ms long, which is shorter than many of the other stop closures but not uniquely short.

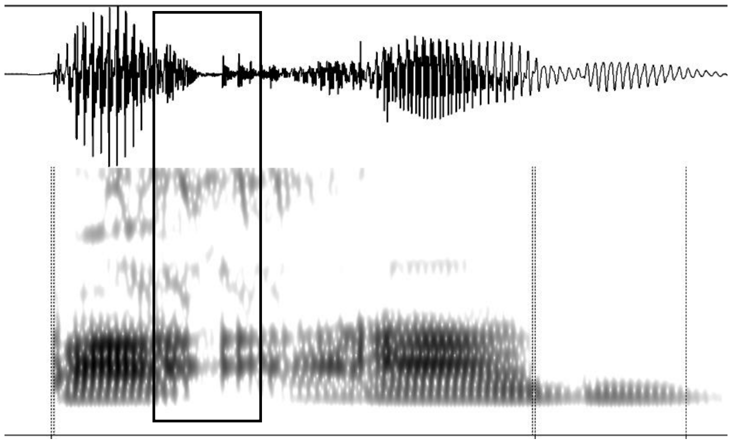

Figure 4.12. Full closure realization of *a'a* in *pa'ang* /paʔaŋ/ [LM; 136.889s]

Figure 4.13 shows an example with minimal (if any) glottalization. In Figure 4.13, the entire /aʔa/ interval is boxed. The glottal stop is perhaps indicated by a relatively sharp decrease in intensity, marked by the orange line, around the midpoint of the interval.

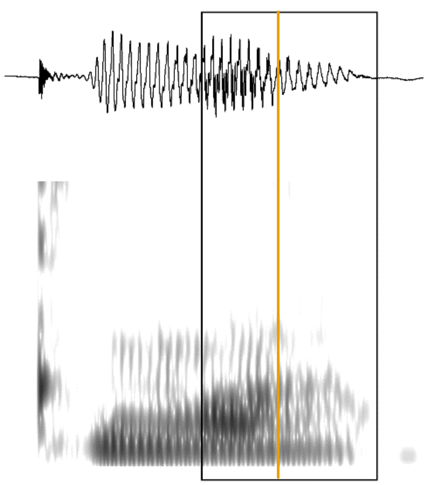

Figure 4.13. Token of *ma'a* /maʔa/ [CA; 210.148949s]

This realization of VʔV is fairly typical for this speaker. Often these sequences are produced with little visible aperiodicity and only small decreases in intensity. Compare

Figure 4.13 to Figure 4.2, which is a plain vowel with no adjacent glottal stop produced by the same speaker.

Figures 4.14–4.16 show some additional realizations of /aʔa/ sequences with aperiodicity and decreased intensity toward the middle of the interval.

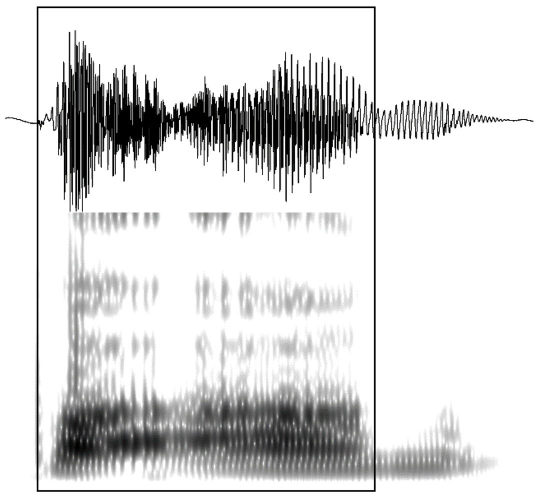

Figure 4.14. Example of *pa'ang* /paʔaŋ/ with glottalization toward vowel midpoint [ER; 80.245s]

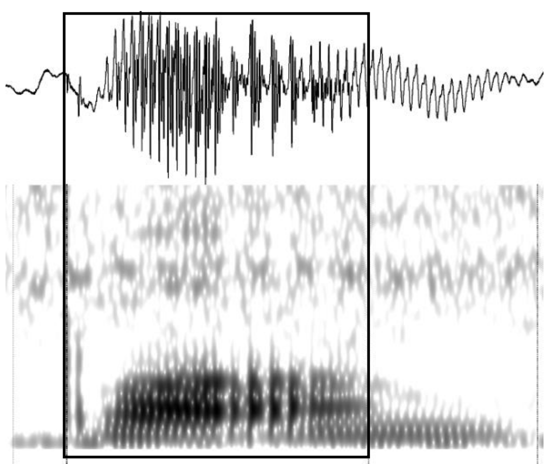

Figure 4.15. Exmaple of *pa'ang* /paʔaŋ/ with glottalization toward vowel midpoint [LF; 369.16s]

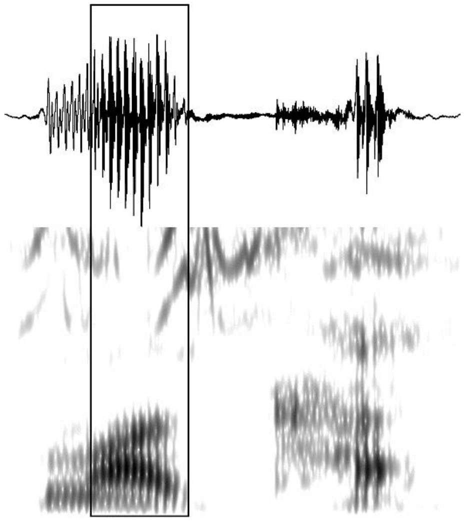

Figure 4.16. Example of *ma'akha'* /maʔakhaʔ/ with minimal aperiodicity but dip in intensity toward vowel midpoint [MR; 365.572311s]

In Figures 4.14–4.16, the VʔV intervals are boxed. In Figures 4.14 and 4.15, there is a clear decrease in intensity and periodicity in roughly the middle third of the vowel. In Figure 4.14, this decrease is accompanied by a near total loss of energy in the upper half of the spectrum, while Figure 4.15 instead shows the familiar sharply attenuated periods and very slow glottal pulses. Figure 4.16, in contrast, has a small dip in intensity but not a clear change in periodicity. This may be because Figure 4.16 shows a very short instance of /aʔa/ from a multisyllabic word, while Figures 4.14 and 4.15 are taken from monosyllabic words and have longer durations.

3. Methods

Refer to Chapter 1 for information about the participants, corpus, and vowel segmentation procedures used in this study. Section 3.1 discusses annotation and classification of the tokens used in this study, and Section 3.2 discusses measurement.

3.1. Annotation and Classification

The same procedure that was used in Chapters 2 and 3 to identify adjacent consonants for each vowel was used in this study. Based on their adjacent consonants, vowels were sorted into one of four VOWEL CONTEXTS: no adjacent glottal stop (PLAIN V), preceding glottal stop (ʔV), following glottal stop (Vʔ), and VʔV. The VʔV category had already been marked separately during the boundary correction process described in Chapter 1, so the exact same set of tokens that were identified as VOWEL LENGTH: VʔV in Chapter 3 were classified as VʔV here. Since there were no cases of a glottal stop on either side of a vowel or of a glottal stop between two non-identical vowels, these categories accounted for the entire dataset.

Since the primary interest in this study is the acoustic effects of glottal stop on adjacent vowels, other factors that are known to affect the selected voice quality measures were included in the analysis in order to rule them out.

In word- or utterance-final position, Redi and Shattuck-Hufnagel (2001) report phrase-final glottalization in American English, and Seyfarth and Garellek (2015) report both word-final /t/ glottalization and phrase-final creak in American English. English and Spanish speakers can use this final glottalization to disambiguate structurally ambiguous sentences (Crowhurst 2018). French uses word- and utterance-final /ʔ/ with a range of purposes (Malécot 1975), and Bellem and Watson (2014) report predictable pre-pausal glottalization in Ṣan’āni, Mehreyyet, and Mahrīyōt Arabic.

For this study, utterance breaks and pre-pausal syllables were identified as in Chapters 2 and 3. Each vowel was classified as either immediately pre-pausal (word-final syllable in a pre-pausal word), or not pre-pausal (all other cases) – this variable is the same as the PRE-PAUSAL SYLLABLE factor used in Chapter 3 to account for pre-pausal lengthening.

Glottal stop insertion or glottalization is also common in initial positions. For example, Mitterer, Kim, and Cho (2019) report epenthetic /ʔ/ before underlyingly vowel-initial words in Maltese; this process is sensitive to prosodic structure. Pompino-Marschall and Żygis (2010) report initial glottalization in vowel-initial words and accented syllables in German. Finally, Dilley and Shattuck-Hufnagel (1995) report both word-and phrase-initial glottalization in American English, and Garellek (2012a) finds that the strength of word-initial glottalization in American English depends on prosodic structure.

Unruh and Kalisch's (1999) description of the Enlhet orthography, combined with auditory impressions of the Enenlhet corpus, suggest that orthographically vowel-initial words may be glottalized. Syllable structures and a vowel's position within a word were also computed in the same way as described in Chapters 2 and 3. Each vowel was marked for its SYLLABLE STRUCTURE/POSITION. This factor categorized each vowel as either belonging to an onsetless, word-initial syllable, or not.

Surrounding /h/ also affects voice quality, so each vowel was also noted as having an ADJACENT /h/ (either immediately preceding or following) or not, as /h/ was expected to result in breathier phonation on the adjacent vowel (Stevens 1998: 426). The annotated factors and baseline values for each one are shown in Table 4.1.

Table 4.1: Factors and baseline values for voice quality analysis

FACTOR	POSSIBLE LEVELS	BASELINE
VOWEL CONTEXT	V, VʔV, ʔV, Vʔ	V
ADJACENT /h/	no adjacent /h/, adjacent /h/	no adjacent /h/
PRE-PAUSAL POSITION	pre-pausal syllable, not pre-pausal syllable	not pre-pausal
SYLLABLE STRUCTURE/ POSITION	onsetless, word-initial syllable; not onsetless, word-initial syllable	not onsetless, word-initial

3.2. Measurement

There are many possible acoustic measures of voice quality, and, because there is no previous study of glottalization in Enenlhet, I could make no hypotheses about which are most active in this language. In general, glottalization is associated with lower pitch, aperiodicity, changes in intensity, and low spectral tilt values. The most easily identifiable realization of /ʔ/ is that of a glottal closure followed by release. However, at least in English, glottal stop is rarely realized as a full closure (Pierrehumbert & Talkin 1992). Borroff (2007) summarizes possible realizations of /ʔ/ in a wide variety of languages (Arbore, Chamicuro, Chemehuevi, Kashaya, Kekchi, Makassarese, Nez Perce, St'at'imcets, Sundanese, Tukang Besi, Wichita, Yapese, Yatzachi Zapotec, Yucatec Maya, Yurok); in addition to a full closure, these realizations include longer periods for vocal fold pulses, decreases in F0 or amplitude, and increased jitter.

Because of its variable realization and effects on adjacent sonorants, I selected a range of frequently-used measures: midpoint harmonics-to-noise ratio (HNR), the difference between the amplitude of the first and second harmonics (H1-H2), F0, and overall intensity. Table 4.2 summarizes these measures, which are described in more detail in Sections 3.2.1 through 3.2.4.[16]

Table 4.2: Voice quality measures used in Enenlhet study

CUE	DEFINITION	VALUES ASSOCIATED WITH GLOTTALIZATION
HNR	intensity of harmonics vs. aperiodic noise	lower (more noise/less prominent harmonics)
Intensity	overall intensity of the signal	lower
H1-H2	amplitude of first harmonic – amplitude of second harmonic	lower (lower open quotient/more damping)
F0	frequency of the first harmonic	lower

[16] I also measured shimmer, jitter, HNR averaged across each third of the vowel, and CPP. Preliminary analysis suggested that the four selected measures were the most robust, and they also capture a range of variability in periodicity, intensity, and spectral tilt.

Vowels were divided into thirds and voice quality measurements were taken from each third to assess changes across time, which was expected to be critical in distinguishing the Vʔ, ʔV, and VʔV cases. The average value of each acoustic factor across the entire vowel interval was also recorded.

3.2.1. Harmonics-to-Noise Ratio (HNR)

HNR measures the intensity of the harmonics of the speech signal relative to the intensity of the surrounding noise, with lower values correlated to lower prominence of speech harmonics. HNR has been correlated with voice quality in a variety of languages, including Arapaho (Whalen, DiCanio, Geissler, & King 2016), American English (Peña, Davidson, & Orosco 2021; Shue, Chen, & Alwin 2010), Javanese (Wayland, Gargash, & Jongman 1994), Ju|’hoansi (Miller 2007), and White Hmong (Fulop & Golston 2008; Garellek 2012b). Increased spectral noise is associated with both breathy phonation (e.g., Shue, Chen & Alwan 2010; English) and creaky voice (e.g., Whalen et al. 2016; Arapaho).

HNR was extracted using a Praat script that calculated HNR at the midpoint of each third of the vowel. The harmonicity settings used to calculate HNR were as follows: glottal time step = 0.01, minimum pitch = 75, silence threshold = 0.01, periods per window = 1.0.

3.2.2. Overall Intensity

In some languages, small dips in overall intensity correlate with a perception of glottalization, even when the signal remains periodic. Examples include: Standard Danish (Fisher-Jørgensen 1989), Hawai’ian (Davidson 2021), Yalálaz Zapotec (Avelino 2010), and Yucatec Maya (Frazier 2009). Gordon and Ladefoged (2001) also mention decreased intensity as an acoustic and perceptual cue to glottalization in their cross-linguistic survey of phonation type. Average intensity was extracted from each third of the vowel via Praat script. Praat's default settings were used to generate the intensity object: pitch floor = 100, time step = 0, mean not subtracted.

3.2.3. H1-H2

The difference in amplitude between the first and second harmonics, or H1-H2, is one of many measures of spectral tilt/balance that have been used to measure glottalization in recent years. During glottalized phonation, the amplitude of the first harmonic is sharply attenuated compared to in modal phonation, resulting in lower H1-H2 values in glottalized tokens. Conversely, higher H1-H2 values (with lower amplitude in the higher harmonics) are associated with breathiness (Stevens 1998; Ahn 2000). H1-H2 has been shown to correlate to glottalization in a wide range of languages. For example, Blankenship (2002) finds lower H1-H2 associated with laryngealization in Mazatec, and Avelino, Shin, and Tilsen (2011) see that it correlates with rearticulated (laryngealized) vowels in Yucatec Maya. Keating, Garellek, and Kreiman (2015) list low H1-H2 as characteristic of prototypical creaky voice, vocal fry, diplophonic voice, aperiodic voicing, and tense voicing, citing work on Chong, English, Hmong, Ju|'hoansi, Mazatec, Mpi, Taiwanese, Trique, Yi languages, and Zapotec. H1-H2 was expected to be a robust indicator of glottalization regardless of the exact type(s) present in Enenlhet. It was calculated in Praat as an average across each third of the vowel. The script generated an Ltas (1-to-1) object and subtracted the mean intensity of H2 (dB) across the interval from the mean intensity of H1(dB).

3.2.4. Fundamental Frequency

Low pitch (F0) is frequently cited as a prototypical correlate of glottalization and /ʔ/. Thurgood (2004) finds that low F0 is a key factor differentiating "slack" and “emphatic” voice in Javanese; Kirk, Ladefoged, and Ladefoged (1993) note lower F0 as a feature of creaky phonation in Jalapa de Díaz Mazatec; and DiCanio (2012) reports F0 lowering and perturbation as correlates of creaky phonation in Itunyoso Trique. In German, both Pompino-Marschall and Żygis (2010) and Brunner and Żygis (2011) use F0 lowering as a proxy for glottalization. Dilley, Shattuck-Hufnagel, and Ostendorf (1996); Redi and Shattuck-Hufnagel (2001); and Peña, Davidson, and Orosco (2021) all note F0 lowering as a feature of creak in American English. Davidson (2020) provides a cross-linguistic review of literature associating F0 dips with glottalization. Lower F0 is also a possible realization of a /ʔ/. Hillenbrand and Houde (1996) note that a small pitch dip, with or without additional amplitude changes or aperiodicity, is often sufficient to cue a perception of a glottal stop in American English, and Davidson (2021) finds low and irregular F0 associated with glottal stops in Hawai’ian.

In order to determine the optimal pitch settings in Praat, which is necessary for accurate F0 and voice quality measures, at least 20 tokens from each speaker were manually examined and settings were adjusted until good pitch tracking and glottal pulse recognition were achieved in most cases. Male speakers used a pitch floor of 70 Hz and a pitch ceiling of 250 Hz. Female speakers used a pitch floor of 100 Hz and a ceiling of 300Hz. The other settings, as follows, were found to work equally well for all speakers: time step= 0.0, max number of candidates = 15, attenuation at ceiling = 0.03, silence threshold = 0.03, voicing threshold = 0.45, octave cost = 0.01, octave jump cost = 0.5, voiced/unvoiced cost = 0.14.

3.3. Hypotheses

My hypotheses about how the four vowel contexts will differ are primarily based on the qualitative examination of glottal stop in Section 2.2, which showed variable realizations with glottalization mostly localized near the portion of the vowel adjacent to the orthographic glottal stop. For the VʔV tokens, I predict an increase in glottalization (lower HNR, F0, H1-H2, and intensity) at the midpoint, compared to the tokens with a preceding or following /ʔ/, which I expect to manifest glottalization more in the first or last third of the vowel, respectively. If the VʔV tokens are indeed two vowels with a glottal stop between, this midpoint glottalization should look like glottalization found at the edges of the ʔV and Vʔ tokens. If a different pattern is found, further language-specific morphosyntactic description will be required to characterize the VʔV cases.

I also expect to find increased glottalization on vowels before pauses and vowels in onsetless, word-initial syllables. I also expect increased breathiness, indicated by higher H1-H2 values, for vowels adjacent to /h/. Vowels adjacent to /h/ may also have a lower intensity and lower HNR due to breathiness.

3.4. (Non)Exclusion Criteria

Unlike for the studies in Chapters 2 and 3, no outliers were removed from this analysis. The vast majority of tokens had at least one measurement which was undefined, due either to issues with automatically extracting the relevant measurements from a messy signal, or simply because some portions of the vowel were devoiced or too aperiodic to reliably measure. Therefore, tokens with some *N/A* values were retained. Because of these *N/A* values, statistical methods (robust distances or Mahalanobis distances) could not be used to identify outliers. The final corpus contained 15,683 vowels. Table 4.3 shows the breakdown of the dataset based on vowel category.

Table 4.3: Number of tokens in each vowel category

VOWEL CATEGORY	# TOKENS	# LEXICAL ITEMS
V	14,124	2,009
VʔV	749	81
Vʔ	378	162
ʔV	432	191

During the initial annotation phase, I marked 1,172 (7.5% of the corpus) tokens as highly glottalized/aperiodic; I also marked four of the VʔV tokens as having full glottal closures.[17] The vast majority of the VʔV tokens were realized with some kind of voice quality change toward the midpoint of the vowel but no full closure (see Section 4).

4. RESULTS

Individual statistical models were constructed for each acoustic measure, which were used as the dependent variables. Random intercepts were included for SPEAKER and LEXICAL ITEM. The four factors in Table 4.1 were fixed effects. As in the other studies, alpha was set to 0.01 to minimize the chance of false positive results due to the unbalanced nature of the dataset. Speaker-by-speaker analyses were also conducted using the same models. Finally, the measures averaged across the entire vowel were used as the dependent variables for another set of four statistical models. The models for each time point were used to determine if the three vowel categories with adjacent /ʔ/ differed in the timing of glottalization, and the models with the averaged measures were used to determine if there were global differences between the four categories.

The following sections examine statistical models for each of the selected acoustic cues. The models for each third of the vowel are presented in detail, with individual speaker

[17] I didn't develop a firm metric for what amount of silence constituted a closure, and I did not exhaustively examine the corpus to mark these cases; the actual number of VʔVs realized with full closures might be slightly higher than this, though I'm confident that it is a genuinely rare realization.

results and averaged models mentioned only briefly. Table 4.4 shows the number of tokens for each speaker, at each time point.

Table 4.4: Number of tokens for each speaker at each time point

	HNR			INTENSITY			F0			H1-H2		
SPEAKER	1ST	2ND	3RD	1ST	2ND	3RD	1ST	2ND	3RD	1ST	2ND	3RD
CA	7513	7530	7503	7583	7583	7574	5734	6899	6664	7583	7583	7583
ER	820	820	819	820	820	820	632	786	770	820	820	820
LF	360	360	358	360	360	360	268	338	336	360	360	360
LM	607	607	606	608	608	608	488	595	584	608	608	608
MM	614	614	613	615	615	615	338	506	526	615	615	615
MR	455	455	455	455	455	455	340	424	437	455	455	455
MRR	2958	2948	2967	2967	2967	2967	2133	2825	2758	2967	2967	2967
TF	704	707	703	707	707	707	506	662	663	707	707	707

4.1. Harmonics-to-Noise Ratio

Figure 4.17 graphs the mean HNR values for each category at the midpoint of each third of the vowel.

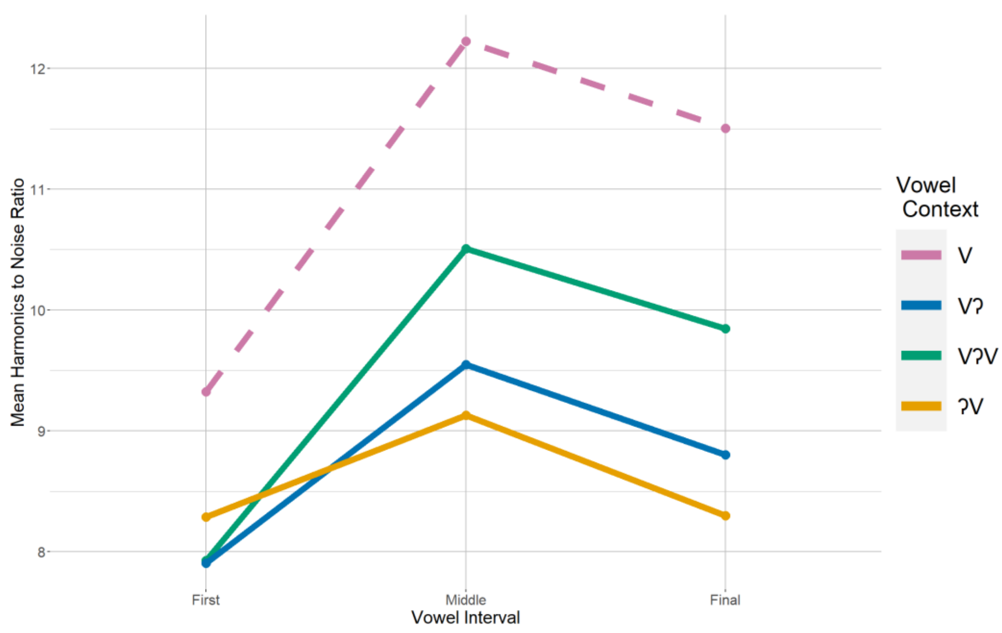

Figure 4.17. Mean HNR at each timepoint, for each vowel category

HNR is expected to be highest at vowel midpoint for the V category, and we do see this pattern (pink dashed line). For the ʔV cases, HNR is expected to rise more steeply between the first third of the vowel and the midpoint than it does for the plain Vs. However, the figure shows a very shallow rise. The Vʔs are expected to be opposite, with a steep fall in HNR between the midpoint and the last third of the vowel. Instead, Figure 4.17 shows a relatively steep rise between the first third and the midpoint and a shallower decline between the midpoint and the last third of the vowel. The VʔVs are expected to have a lower HNR at midpoint which, since HNR in general tends to be lower at the edges of the vowel (see the plain Vs) might manifest as relatively flat HNR values across the entire interval. Instead, the VʔVs show a steep increase in HNR between the first third and the midpoint and a shallower decline between the midpoint and the last third of the vowel. In fact, the trajectories for the plain Vs and the VʔVs appear to be nearly parallel, though the VʔVs have a lower HNR overall.

The variability in the HNR measures is quite large for all four vowel contexts, ranging from 5.86 to 6.62 in the first third of the vowel, 5.65 to 6.21 in the second, and 8.30 to 11.51 in the third. The plain Vs have the largest standard deviations in each third of the vowel.

The statistical models suggest that difference between the plain Vs and each of the other contexts is significant at all three time points. Table 4.5 shows the results from the three models of HNR.

Table 4.5: Results of by-timepoint models of HNR

```
HNR = (1|speaker) + (1|lexicalItem) + vowelContext + adjacentH +
wordInitialOnsetlessSyllable + pre-PausalSyllable
```

FACTOR	ESTIMATE	s	d.f	t	p
		FIRST THIRD			
(INTERCEPT)	9.72	0.48	7.33	20.40	<0.001
VʔV	-2.96	0.43	8795.80	-6.91	<0.001
Vʔ	-1.65	0.37	14408.41	-4.46	<0.001
ʔV	-0.94	0.33	15328.54	-2.86	<0.01
ADJACENT /h/	-0.85	0.20	14291.45	-4.25	<0.001
WORD INITIAL, ONSETLESS SYLLABLE	-2.35	0.23	15535.75	-10.09	<0.001
PRE-PAUSAL SYLLABLE	-0.18	0.13	15460.95	-1.41	0.16
		MIDDLE THIRD			
(INTERCEPT)	13.73	0.88	7.09	15.58	<0.001
VʔV	-2.87	0.38	8240.69	-7.51	<0.001
Vʔ	-2.40	0.33	14281.72	-7.28	<0.001
ʔV	-1.97	0.29	15317.42	-6.73	<0.001
ADJACENT /h/	-0.62	0.18	14118.92	-3.48	<0.001
WORD INITIAL, ONSETLESS SYLLABLE	-3.00	0.21	15549.59	-14.39	<0.001
PRE-PAUSAL SYLLABLE	-1.00	0.11	15427.89	-8.92	<0.001
		LAST THIRD			
(INTERCEPT)	12.82	0.72	7.13	17.75	<0.001
VʔV	-2.00	0.37	9075.65	-5.37	<0.001
Vʔ	-1.87	0.32	14495.99	-5.87	<0.001
ʔV	-2.05	0.29	15344.17	-7.17	<0.001
ADJACENT /h/	0.53	0.17	14352.98	3.06	<0.01
WORD INITIAL, ONSETLESS SYLLABLE	-1.86	0.20	15506	-9.20	<0.001
PRE-PAUSAL SYLLABLE	-1.57	0.11	15311.5	-14.46	<0.001

Since the primary interest is the difference between the four vowel contexts, post hoc pairwise comparisons were conducted to determine whether these differences are statistically significant. Table 4.6 shows these results, with the significant results shaded grey.

Table 4.6: Pairwise comparisons of vowel quality in the HNR model

COMPARISON	FIRST THIRD	MIDDLE THIRD	LAST THIRD
V – VʔV	β=2.96, p<0.0001	β=2.87, p<0.0001	β=2.00, p<0.0001
V – Vʔ	β=1.65, p<0.0001	β=2.40, p<0.0001	β=1.87, p<0.0001
V – ʔV	β=0.94, p=0.02	β=1.97, p<0.0001	β=2.05, p<0.0001
VʔV – Vʔ	β=-1.31, p=0.08	β=-0.47, p=0.78	β=-0.13, p=0.99
VʔV – ʔV	β=-2.02, p<0.001	β=-0.90, p=0.23	β=0.05, p=1.00
Vʔ – ʔV	β=-0.71, p=0.45	β=-0.43, p=0.75	β=0.18, p=0.97

The pairwise comparisons show that Vs have a greater HNR than VʔVs and Vʔs across the vowel and a greater HNR than ʔVs in the last 2/3 of the vowel. Since HNR is always expected to be highest at the midpoint, at least for modal vowels, the fact that the plain Vs have a higher HNR than the VʔV cases at midpoint is not surprising. However, glottalization was expected to primarily be located in the middle of these cases, so the fact that plain Vs have a higher HNR the VʔVs in the first and last thirds of the vowel is unexpected. Likewise, glottalization in the Vʔ cases was expected in the last third of the vowel, so the fact that these tokens have a lower HNR than vowels not adjacent to /ʔ/ across the entire vowel is also surprising.

The ʔV cases are even more unexpected. Glottalization, indicated by lower HNR, was most expected in these tokens in the first third of the vowel, as the /ʔ/ is adjacent to the left edge. Instead, there is no difference in HNR between the plain Vs and the ʔVs at the left edge, but there *is* a difference in the remaining 2/3 of the vowel.

The other surprising result revealed by these pairwise comparisons is that in the first third of the vowel, VʔVs have a lower HNR than ʔVs. The VʔVs were expected to be relatively more modal than the ʔVs in the first third of the vowel and instead, the VʔVs have a lower HNR than the ʔVs there. This result is a logical consequence of the previous two; VʔVs have a lower HNR than plain Vs in the first third of the vowel, and ʔVs do not differ from plain Vs during the same interval.

Briefly, the other fixed effects in Table 4.5 are as predicted. Word-initial, onsetless syllables have a lower HNR across the entire vowel. Vowels adjacent to /h/ also have a lower HNR across the entire vowel. Pre-pausal vowels have a lower HNR than non-pre-pausal vowels in the last 2/3 of the vowel. This lower HNR is assumed to be due to glottalization for the pre-pausal and word-initial vowels and due to breathiness in the vowels adjacent to /h/, but HNR does not distinguish the two sources.

When each speaker was analyzed individually, none of them showed the expected patterns for the VʔV, ʔV, and Vʔ categories. Some speakers (LM, TF, MM, MRR) showed basically the same patterns as seen in the overall model, and others (most notably CA) showed constant HNR across the vowel and no real differences between the four vowel contexts.

The model of HNR averaged across the entire vowel interval shows what we expect based on the models in each third of the vowel. HNR is significantly lower for each of the /ʔ/-adjacent vowel categories, but there are no significant differences between the three categories.

4.2. Overall Intensity

Figure 4.18 shows the mean intensity values averaged across each third of the vowel for each category. The variation in intensity is similar to the variability seen in the HNR measures; it ranges from 5.29 dB to 7.57 dB in the first third, 5.76 dB to 6.89 dB in the second third, and 6.80 dB to 8.87 dB in the last third. There are no patterns in terms of which vowel context showed greater or less variability.

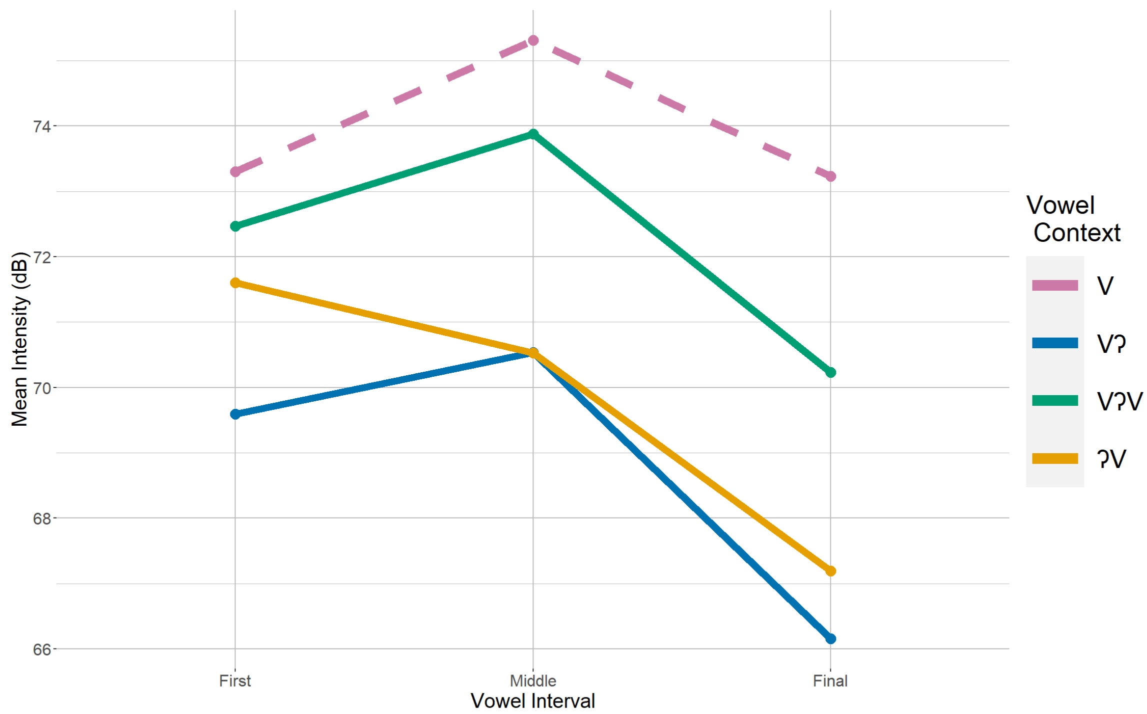

Figure 4.18. Mean overall intensity at each timepoint, for each vowel category

Like for HNR, overall intensity is expected to be highest at the midpoint of the vowel. The plain Vs show the expected pattern. The ʔVs are expected to show a steeper increase between the first third and middle thirds than the plain Vs; instead, Figure 4.20 shows that intensity in the ʔVs decreases across the entire vowel.

The Vʔs do show about the trajectory that was predicted – intensity increases between the first and middle third of the vowel and decreases more steeply between the middle and last thirds. However, the Vʔs have a lower intensity in each third of the vowel compared to the plain Vs, which was not necessarily predicted, since the effect of the following glottal was primarily expected in the last third of the vowel.

Once again, in Figure 4.20 the VʔVs are similar to the plain Vs, though intensity is lower in each third of the vowel than in the plain Vs. However, the expected intensity trajectory for the VʔVs was, as for HNR, expected to be relatively flat, or perhaps even lower at midpoint than in the first and last thirds. Therefore, the fact that the VʔVs show the same trajectory as the plain Vs is unexpected.

Table 4.7 shows the results of the statistical models of intensity in each third of the vowel.

Table 4.7: Results of by-timepoint models of overall intensity

`Intensity = (1|speaker) + (1|lexicalItem) + vowelContext + adjacentH + wordInitialOnsetlessSyllable + pre-PausalSyllable`

FACTOR	ESTIMATE	s	d.f.	t	p
FIRST THIRD					
(INTERCEPT)	73.30	1.49	7.02	49.33	<0.001
VʔV	-1.87	0.40	10490.09	-4.63	<0.001
Vʔ	-2.01	0.34	14972.83	-5.84	<0.001
ʔV	-0.98	0.30	15596.04	-3.23	<0.01
ADJACENT /h/	-3.11	0.19	14898.02	-16.70	<0.001
WORD-INITIAL, ONSETLESS SYLLABLE	-2.33	0.22	15669.69	-10.77	<0.001
PRE-PAUSAL SYLLABLE	-2.53	0.12	15400.96	-21.93	<0.001
MIDDLE THIRD					
(INTERCEPT)	75.87	1.54	7.01	49.39	<0.001
VʔV	-1.30	0.34	9659	-3.79	<0.001
Vʔ	-2.17	0.29	14760	-7.40	<0.001
ʔV	-2.92	0.26	1530	-11.25	<0.001
ADJACENT /h/	-1.32	0.16	14660	-8.36	<0.001
WORD-INITIAL, ONSETLESS SYLLABLE	-0.60	0.19	15670	-3.24	<0.01
PRE-PAUSAL SYLLABLE	-4.24	0.10	15450	-42.94	<0.001
LAST THIRD					
(IINTERCEPT)	73.99	1.47	7.01	50.41	<0.001
VʔV	-1.84	0.44	6501	-4.19	<0.001
Vʔ	-3.14	0.38	13640	-8.19	<0.001
ʔV	-3.46	0.34	15060	-10.15	<0.001
ADJACENT /h/	0.30	0.21	13450	1.45	0.15
WORD-INITIAL, ONSETLESS SYLLABLE	0.01	0.25	15520	0.04	0.97
PRE-PAUSAL SYLLABLE	-5.90	0.13	15610	-45.11	<0.001

The models again suggest that overall intensity is lower in each vowel category compared to the plain Vs. Pairwise comparisons of these four contexts appear in Table 4.8, with significant results shaded grey.

Table 4.8: Pairwise comparisons of vowel category in the intensity models

COMPARISON	FIRST THIRD	MIDDLE THIRD	LAST THIRD
V – VʔV	β=1.87, p<0.0001	β=1.30, p<0.001	β=1.84, p<0.001
V – Vʔ	β=2.01, p<0.0001	β=2.17, p<0.0001	β=3.14, p<0.0001
V – ʔV	β=0.98, p=0.007	β=2.92, p<0.0001	β=3.46, p<0.0001
VʔV – Vʔ	β=0.14, p=0.99	β=0.87, p=0.20	β=1.30, p=0.10
VʔV – ʔV	β=-0.89, p=0.28	β=1.62, p<0.001	β=1.62, p=0.12
Vʔ – ʔV	β=-1.03, p=0.10	β=0.75, p=0.20	β=0.32, p=0.92

The pairwise results for intensity look extremely similar to the HNR results. Plain Vs have a higher intensity than VʔVs and Vʔs across the entire vowel and a higher intensity than ʔVs in the last 2/3 of the vowel.

As in Section 4.1, these results are unexpected. The VʔVs were expected to differ from the plain Vs mostly in the middle third rather than across the entire interval. The Vʔs were expected to differ mostly in the last third, rather than across the whole interval. The ʔVs were expected to differ primarily in the first third of the vowel, the only interval which doesn’t show a difference between the ʔVs and the plain Vs.

The other significant pairwise comparison shows that VʔVs have a higher intensity than the ʔVs at midpoint, which is also unexpected, as the VʔVs were expected to have the lowest intensity of all four groups at the midpoint.

Once again, the individual models for each speaker show that no speakers produced the expected differences in intensity across time for the four vowel contexts. Most speakers showed patterns similar to those seen in the overall models, though TF and LF showed lower intensities than the other speakers across the entire vowel (this could be due to a difference in microphone placement or overall speaking volume). CA once again showed very little difference between the four vowel contexts. Likewise, the model of intensity averaged across the entire interval shows significantly lower intensity for each of the /ʔ/-adjacent categories compared to plain Vs but no significant differences between the three categories.

The other fixed effects in Table 4.7 behave as they were predicted to. Vowels adjacent to /h/ have a lower intensity than vowels not adjacent to /h/ in the first 2/3 of the vowel. Since the ADJACENT /h/ factor did not distinguish between a preceding and following /h/, it is somewhat surprising that this decreased intensity is not present in the last third of the vowel. It's possible that this is due to an asymmetry in the distribution of /h/ in the corpus. Word-initial, onsetless syllables have a lower intensity than vowels in other positions for the first 2/3 of the vowel; the lower intensity appears to spread from the left edge, as expected. Pre-pausal vowels have a lower intensity than non-pre-pausal vowels in all three thirds.

4.3. Fundamental Frequency

Figure 4.19 shows the mean F0 at each timepoint for each vowel category. The plain Vs show the least variability in F0, and the Vʔs show the most. Standard deviations range from 77.49 Hz to 101.44 Hz in the first third, 46.50 Hz to 86.36 Hz at the midpoint, and 45.72 Hz to 130.98 Hz in the last third of the vowel.

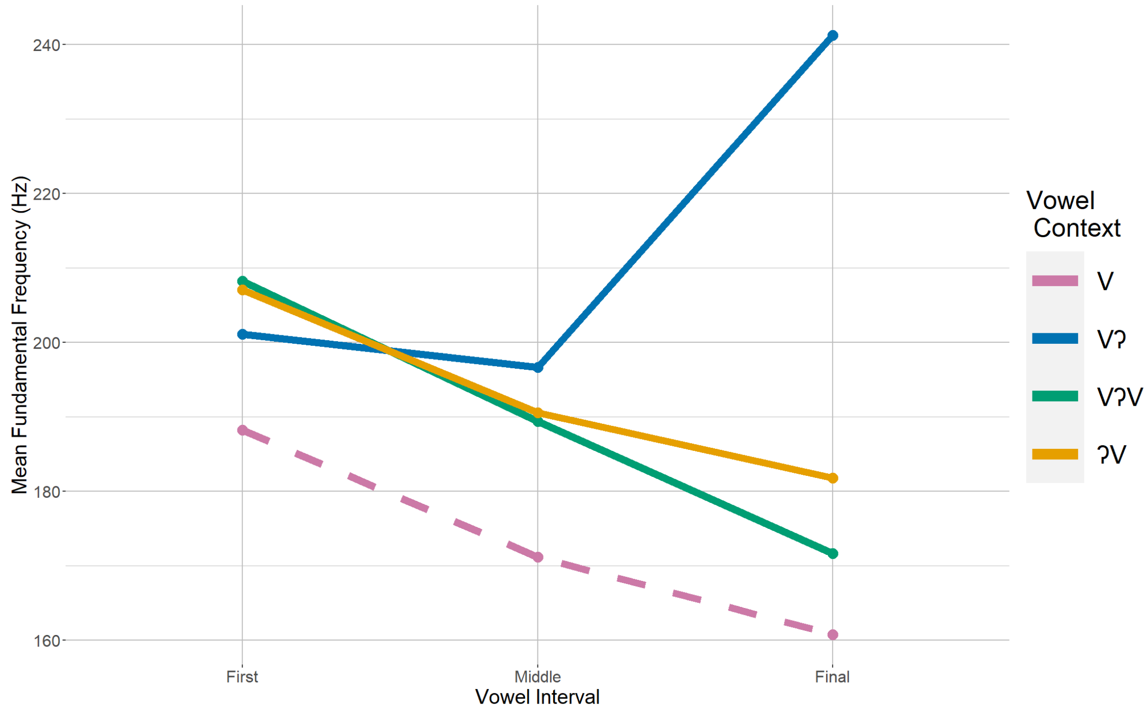

Figure 4.19. Mean F0 at each timepoint, for each vowel category

The F0 measures show a different trend than intensity and HNR, and they were not at all what I expected to find. F0 was expected to be the highest for the plain Vs, as low F0 is a typical correlate of glottalization. However, the plain Vs in Figure 4.21 show the lowest F0 values across each third of the vowel.

The ʔVs were expected to show an increase in F0 between the first third and the midpoint, and instead they show a steady decrease across time. A decrease across the vowel is what was expected for the Vʔs, and instead Figure 4.21 shows a very sharp increase in F0 between the midpoint and last third of the vowel. Since the standard deviations in these tokens were so large, especially in the last third of the vowel, that I suspect that Praat failed to accurately track the pitch when glottal pulses were very slow or irregular, resulting in pitch doubling or octave jumps which don't really represent the speech signal.

The VʔVs were anticipated to have a U-shaped trajectory, since glottalization was expected in the middle third of the vowel. However, these tokens also show a relatively linear decrease in F0 across the vowel.

Table 4.9 shows the results of the statistical models from each third of the vowel, and Table 4.10 presents the post hoc pairwise comparisons between the vowel categories (shaded grey), which is the primary factor of interest.

Table 4.9: Results of by-timepoint models of F0

```
F0 = (1|speaker) + (1|lexicalItem) + vowelContext + adjacentH +
wordInitialOnsetlessSyllable + pre-PausalSyllable
```

FACTOR	ESTIMATE	s	d.f.	t	p
FIRST THIRD					
(INTERCEPT)	172.15	12.25	7.06	14.05	<0.001
VʔV	30.07	5.41	2398.63	5.56	<0.001
Vʔ	26.76	5.06	8664.46	5.29	<0.001
ʔV	13.23	4.05	10665.51	3.27	<0.01
ADJACENT /h/	17.38	2.66	7886.6	6.54	<0.001
WORD-INITIAL, ONSETLESS SYLLABLE	49.26	2.99	11094.52	16.46	<0.001
PRE-PAUSAL SYLLABLE	-2.64	1.64	11558.24	-1.61	0.11
MIDDLE THIRD					
(INTERCEPT)	160.93	11.67	7.01	13.8	<0.001
VʔV	24.56	2.57	1247.39	9.57	<0.001
Vʔ	36.31	2.53	9481.29	14.37	<0.001
ʔV	13.89	2.28	12575.52	6.09	<0.001
ADJACENT /h/	-0.38	1.32	7762.07	-0.29	0.77
WORD-INITIAL, ONSETLESS SYLLABLE	28.69	1.55	13072.44	18.49	<0.001
PRE-PAUSAL SYLLABLE	1.13	0.87	14030.99	1.29	0.2
LAST THIRD					
(INTERCEPT)	152.1	11.71	7.01	12.98	<0.001
VʔV	15.7	2.81	1793	5.58	<0.001
Vʔ	79.2	2.86	10870	27.71	<0.001
ʔV	11.13	2.55	12910	4.37	<0.001
ADJACENT /h/	1.56	1.35	8610	1.15	0.25
WORD-INITIAL, ONSETLESS SYLLABLE	11.28	1.6	12860	7.05	<0.001
PRE-PAUSAL SYLLABLE	11.14	0.94	13750	11.81	<0.001

Table 4.10: Pairwise comparisons between the vowel categories in the F0 models

COMPARISON	FIRST THIRD	MIDDLE THIRD	LAST THIRD
V – VʔV	β=-30.1, p<0.0001	β=-24.6, p<0.0001	β=-15.7, p<0.0001
V – Vʔ	β=-26.8, p<0.0001	β=-36.3, p<0.0001	β=-79.2, p<0.0001
V – ʔV	β=-13.2, p=0.006	β=-13.9, p<0.0001	β=-11.13, p<0.0001
VʔV – Vʔ	β=3.3, p=0.97	β=-11.7, p=0.005	β=-63.5 p<0.0001
VʔV – ʔV	β=16.8, p=0.05	β=10.7, p=0.008	β=4.57, p=0.61
Vʔ – ʔV	β=13.5, p=0.14	β=22.4, p<0.0001	β=68.06, p<0.0001

As Figure 4.19 suggests, plain Vs have a significantly higher F0 than VʔVs and Vʔs across the entire vowel. These results are unexpected but consistent between the figure and the model. The plain Vs also show a lower F0 than the ʔVs in the last 2/3 of the vowel (note that this difference would also be significant in the first third if we set the alpha to 0.05 rather than 0.01). Again, this is opposite what I expected to find.

There are three other significant comparisons in the model which show differences between the three contexts in which vowels are adjacent to /ʔ/. I expected to find differences between these categories, but the differences in the model were not the ones I anticipated. These comparisons show that the VʔVs have a lower F0 than the Vʔs in the last third of the vowel; the last third of the vowel is where I expected the Vʔs to have the lowest F0 of all categories (this difference would also be significant at midpoint if we used alpha = 0.05). The Vʔs also show a higher F0 than the ʔVs in the last 2/3 of the vowel. Again, I expected the ʔVs to be approaching modal voice during this portion of the vowel and for Vʔs to become steadily more glottalized during the same time frame, so these comparisons are opposite the expected finding.

The other fixed effects in the model also did not show the expected behavior, though these results are consistent with what was seen for the vowels adjacent to /ʔ/. Vowels in word-initial, onsetless syllables have a higher F0 for the first 2/3 of the vowel than vowels in other syllables. This shows an effect stretching from the left edge, as was seen for intensity and HNR, but the increase in F0 is opposite my prediction. The pre-pausal syllable effect seems weaker, appearing only in the last third of the vowel; like the others, it manifests as a higher F0 in the last third of the vowel.

Vowels adjacent to /h/ also have a higher F0 in the first third of the vowel than vowels not adjacent to /h/. This result is particularly strange because, unlike for intensity and HNR, an adjacent /h/ was not expected to affect F0, and it was not expected to pattern like pre-pausal vowels or word-initial vowels, which were expected to be glottalized. Increased F0 is well-documented after voiceless consonants (e.g., House & Fairbanks 1953 Stevens 1998: 457), but this effect does not explain why /h/ specifically might raise F0, given that all Enenlhet obstruents are voiceless.

Once again, none of the speakers clearly show the anticipated F0 lowering due to glottalization, or the expected timing of pitch changes based on the location of the orthographic /ʔ/. In these comparisons, the last third of the vowel almost always shows that vowels adjacent to glottals have higher F0 than plain Vs. Only one speaker, CA, clearly shows the lower F0 in plain V tokens that was seen in the overall model.

The model of F0 averaged across each vowel shows significantly higher F0 for VʔV and Vʔs compared to plain Vs. Pairwise comparisons additionally indicate that ʔVs have a lower F0 than VʔVs and Vʔs but do not differentiate ʔVs and plain Vs.

4.4. H1-H2

Figure 4.20 graphs mean H1-H2 across each third of the vowel. The plain Vs have the least variability of the four vowel categories. Standard deviations range from 31.94 dB to 52.11 dB in the first third of the vowel, 24.35 dB to 40.79 dB in the middle third of the vowel, and 32.12 dB to 63.38 dB in the last third of the vowel. In the last 2/3 of the vowel, the ʔVs are the category with the greatest variability.

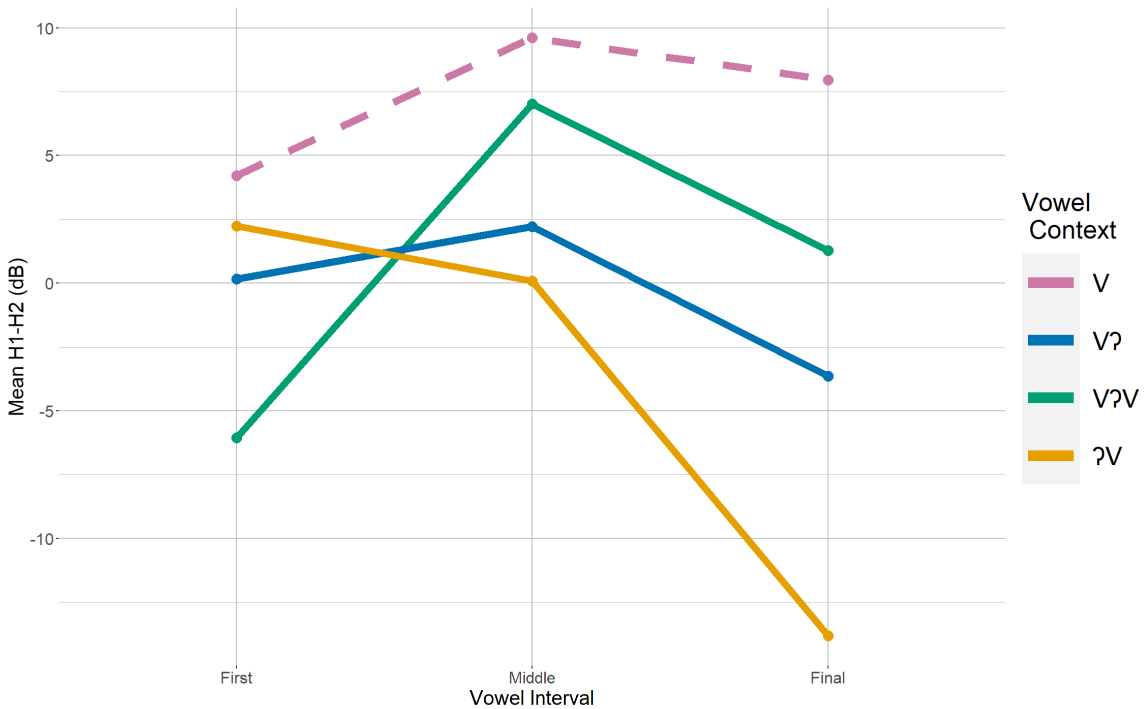

Figure 4.20. Mean H1-H2 at each timepoint, for each vowel category

H1-H2 was also expected to be highest at vowel midpoint for the Vʔ, ʔV, and plain Vs. This pattern does appear for the plain Vs and the Vʔs. The plain Vs have a higher H1-H2 than all other categories across the entire vowel. The Vʔs, though the changes in H1-H2 fit with my predictions, show a lower H1-H2 than the plain Vs across the entire vowel, whereas I expected them to be quite similar in at least the first third of the vowel.

I expected H1-H2 to be lowest during the first third of the vowel for the ʔV cases, and Figure 4.22 shows that the opposite occurs. H1-H2 decreases across these tokens, with a very sharp decrease between the midpoint and last third of the vowel; this is the portion that I expected to be most like the plain Vs.

The VʔVs were also expected to have a U-shaped H1-H2 trajectory, or at least to be relatively flat (given that H1-H2 is expected to be a bit lower at the edges of vowels). However, H1-H2 sharply increases between the first third of the vowel and the middle third for the VʔVs. This increase is steeper than for the plain Vs, contra my expectation.

Table 4.11 shows the results of the statistical models of H1-H2 at each timepoint.

Table 4.11: Results of by-timepoint models of H1-H2

```
H1-H2 = (1|speaker) + (1|lexicalItem) + vowelContext + adjacentH +
wordInitialOnsetelessSyllable + pre-PausalSyllable
```

FACTOR	ESTIMATE	s	d.f.	t	p
FIRST THIRD					
(INTERCEPT)	5.38	1.62	7.33	3.31	<0.01
VʔV	-8.4	2.01	1787.66	-4.19	<0.001
Vʔ	-2.66	1.9	10309.49	-1.4	0.16
ʔV	-0.47	1.71	12759.18	-0.28	0.78
ADJACENT /h/	-0.6	1.02	9591.43	-0.59	0.56
WORD-INITIAL, ONSETLESS SYLLABLE	-1.44	1.23	14414.1	-1.17	0.24
PRE-PAUSAL SYLLABLE	-3.46	0.67	15481.58	-5.17	<0.001
MIDDLE THIRD					
(INTERCEPT)	13.47	1.13	7.98	11.96	<0.001
VʔV	-2.93	1.5	1103.27	-1.96	0.05
Vʔ	-3.37	1.44	8766.66	-2.34	0.02
ʔV	-5.67	1.3	11485.62	-4.37	<0.001
ADJACENT /h/	-1.23	0.77	7905.86	-1.59	0.11
WORD-INITIAL, ONSETLESS SYLLABLE	-2.65	0.94	13853.39	-2.82	<0.01
PRE-PAUSAL SYLLABLE	-7.26	0.51	15283.04	-14.27	<0.001
LAST THIRD					
(INTERCEPT)	13.4	1.13	8.41	11.84	<0.001
VʔV	-2.21	1.98	1354.85	-1.12	0.26
Vʔ	-2.39	1.92	9875.34	-1.25	0.21
ʔV	-14.36	1.73	12234.13	-8.28	<0.001
ADJACENT /h/	3.74	1.03	9000.53	3.62	<0.001
WORD-INITIAL, ONSETLESS SYLLABLE	-0.9	1.26	14242.97	-0.71	0.48
PRE-PAUSAL SYLLABLE	-15.16	0.68	15324.07	-22.27	<0.001

Table 4.12 shows the post-hoc pairwise comparisons that were used to investigate the differences between the different vowel contexts, with significant results shaded grey.

Table 4.12: Pairwise comparisons between vowel categories in H1-H2 models

COMPARISON	FIRST THIRD	MIDDLE THIRD	LAST THIRD
V – VʔV	β=8.40, p<0.001	β=2.95, p=0.20	β=2.21, p=0.68
V – Vʔ	β=2.66, p=0.50	β=3.37, p=0.09	β=2.39, p=0.60
V – ʔV	β=0.47, p=0.99	β=5.67, p<0.0001	β=14.36, p<0.0001
VʔV – Vʔ	β=-5.74, p=0.14	β=0.44, p=1.00	β=0.19, p=1.00
VʔV – ʔV	β=-7.92, p=0.01	β=2.75, p=0.49	β=12.15 p<0.0001
Vʔ – ʔV	β=-2.18, p=0.82	β=2.30, p=0.61	β=11.97, p<0.0001

Table 4.12 shows few significant differences between these categories, although spectral tilt is the measure most clearly associated with glottalization. The plain Vs have a higher H1-H2 than the VʔVs in the first third, where the two groups were expected to be similar. The Vs also show higher H1-H2 than the ʔVs for the last 2/3 of the vowel. The two groups were expected to be relatively similar for these two intervals. Notably, the Vs do not differ from the ʔVs in the first third, where glottalization and therefore lower H1-H2 was most expected for the ʔVs.

There are two significant differences between the groups of vowels adjacent to /ʔ/, though neither was expected. The VʔVs have a higher H1-H2 than the ʔVs in the last third of the vowel. They were expected to be about the same at this point. The Vʔs also show a higher H1-H2 than the ʔVs in the last third; this is the point where the Vʔs were predicted to have the lowest H1-H2 values of all the vowel contexts.

Examination of the data separately for each speaker shows that H1-H2 is level across the vowel for most speakers, and in no cases does it dip in the center for the VʔV tokens, as would be expected if glottalization were located in the center of those intervals. The statistical model of H1-H2 averaged across the interval was similarly unenlightening.

The other factors included in the model do behave mostly as I predicted. Pre-pausal vowels have a lower H1-H2 across the vowel, and vowels in word-initial, onsetless syllables have a lower H1-H2 at midpoint. These effects are consistent with greater glottalization in these tokens, which is what I expected to find. Vowels with an adjacent /h/ have a higher H1-H2 in the last third of the vowel, which is consistent with breathiness, again, expected adjacent to /h/. It is surprising that ADJACENT /h/ did not result in a similar increase in H1-H2 in the first third of the vowel, since the factor coded for /h/ on either side. Distributional irregularities in the dataset might explain this asymmetry.

5. DISCUSSION

I expected adjacent glottal sounds to result in changes in acoustic features associated with glottalization, and they did. Vowels adjacent to glottal stops have a lower HNR, overall intensity, and H1-H2 compared to vowels not adjacent to glottal stops, and they have a higher F0. However, the statistical models show increased noise and lower intensity spread across the entire vowel interval rather than localized in the part of the vowel closest to the adjacent /ʔ/.

My main prediction about differences between the four categories was based on qualitative observation of the data suggesting that the timing of glottalization differed depending on the location of the glottal stop. This observation is also consistent with Wheeler's (2020) report for Angaité; there, I show a number of examples with increasing acoustic cues to glottalization (jitter, shimmer, and intensity) toward the edge of the vowel adjacent to /ʔ/. I also observe decreases in intensity and increases in aperiodicity toward the midpoint of some long vowels.[18]

However, the statistical models of data from each time point were unable to identify timing differences. It may be that this investigation did not find differences between Vʔ, ʔV, and VʔVs because I used a relatively course-grained time dimension. Since the corpus contains naturalistic speech, and since the duration study in Chapter 3 indicated a broad range of vowel durations, some quite short, I only divided the vowels into thirds to avoid measurement issues due to very short intervals. However, the statistical models suggest that the effects of adjacent glottal stops spread across large portions of the vowel.

[18] Having now examined the Enenlhet data, I am convinced at least some of the cases I identified in Angaité are also VʔVs. The time course of the acoustic changes is similar to Enenlhet, and at least some of the tokens appear to be cognate with Enenlhet lexical items that have a VʔV.

Therefore, if there are differences between the categories like in Figure 4.21, this analysis is unable to identify them.

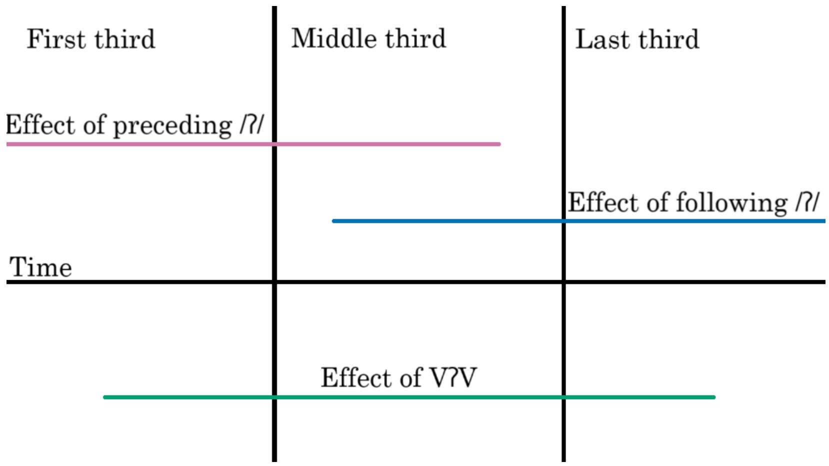

Figure 4.21. Possible distribution of glottalization in each glottal-adjacent vowel category with respect to time

A more fine-grained time domain, which divides each vowel into fifths (or smaller) would be able to identify timing differences between Vʔ, ʔV, and VʔV categories with more precision.

In addition to a finer-grained time dimension, a variety of other factors might account for the somewhat unexpected results of this study. Voice quality measures are sensitive to background noise, as measuring spectral tilt and F0 requires accurate identification of individual glottal periods, so some of the unexpected results may be due to measurement errors.

Though measurement errors are possible, and future controlled studies of glottalization in Enenlhet are certainly needed, there are several reasons to take the unexpected direction of the F0 results as reliable. First, Wheeler (2020) shows frequent word-final pitch increases in Angaité, which are often associated with increased glottalization. Most of the words I used for that study were produced in isolation in elicitation, conflating word-final and phrase-final phenomena. These results suggest that, at least in pre-pausal syllables, we might really expect a higher F0.

Secondly, though glottalization is commonly associated with lower pitch, in some cases glottalization is associated with high pitch, especially on preceding vowels. In Yucatec Maya, for example, a phonologized high tone precedes glottalized segments (Frazier 2013). Danish *stød* is also produced with an initial high pitch before creaky voice (Fisher-Jørgensen 1989), vowels before glottalized sonorants in Coatlán-Loxicha Zapotec are produced with high pitch (Plauché, de Azcona, Roengpitya & Weigel 1998), and following /ʔ/ has resulted in the development of high or rising pitch contours in many languages (e.g., Burmese, Lalu, Middle Chinese; experimental results from Arabic also show this effect; Hombert 1978). Glottal squeak, described by Redi and Shattuck-Hufnagel (2001: 423) for American English, is also characterized by an abrupt, high F0. In their corpus, glottal squeak usually accompanied other acoustic correlates of glottalization, such as extremely low intensity. Therefore, there is some cross-linguistic evidence for a relationship between /ʔ/ and a high F0, though this is not the typical association.

Finally, PRE-PAUSAL SYLLABLE and WORD-INITIAL, ONSETLESS SYLLABLE showed a higher F0 for vowels in these positions. As expected, these positions were also associated with significantly lower intensity, H1-H2, and HNR, similar to what was observed for the vowels adjacent to /ʔ/. The fact that these two factors in addition to the three glottal-adjacent levels of VOWEL CONTEXT behaved as predicted suggests that Praat was relatively successful at identifying harmonics and formants in most cases. Since both positional factors and VOWEL CONTEXT behaved as expected on the other acoustic measures, and because there is cross-linguistic evidence of a relationship between some types of glottalization and high pitch, it is plausible, albeit unexpected, that /ʔ/ in Enenlhet is linked to a high F0.

Many of the speakers who contributed to the corpus are also relatively elderly, and the effect of age on voice quality is well documented (see e.g., Iseli, Shue, & Alwan 2006). Of the speakers whose ages are known, two are in their 30s (MM: 35, MR: 39), and together their recordings total 14:20. The two other speakers whose ages are known are in their 70s (CA: 72, MRR: 77), and their recordings account for well over half the corpus. Recall also that some speakers showed aperiodicity and low intensity even in contexts where glottalization was not expected. Therefore, if there are differences between the Vʔ, VʔV, and ʔVs, but they are quite small, speakers' baseline creakiness may obscure them.

As noted above, the pre-pausal syllables; word-initial, onsetless syllables; and vowels adjacent to /h/ pattern roughly as expected. Pre-pausal syllables correlate with greater glottalization, particularly in the later portions of the vowel. Word-initial, onsetless syllables are also associated with the same changes in the acoustic measures as the vowels with adjacent /ʔ/, suggesting that, as in Enlhet, there is a preceding [ʔ] before orthographically vowel-initial words. Determining if this glottal stop is phonemic requires further collaboration with native speakers.

An adjacent /h/ also results in acoustic effects similar to adjacent /ʔ/, with the exception of raising H1-H2 in the final third of the vowel. This higher H1-H2 correlates to greater breathiness, which is the anticipated effect of an adjacent /h/. The effect of an adjacent /h/ on F0 was unexpected, as there is no cross-linguistic evidence that /h/ affects F0. The fact that /h/ was associated with an increase in F0 might be explained by increased breathiness (as indicated by the H1-H2 model) causing pitch tracking errors; future work addressing F0 measures for the vowels in different glottal stop contexts would also be useful in examining the effect of /h/.

6. CONCLUSION

This chapter has investigated the acoustic correlates of glottalization that appear in vowels preceded and followed by /ʔ/ and in sequences of identical vowels separated by a glottal stop (<a’a, e’e, o’o>). Statistical analysis showed that ʔV, Vʔ, and VʔVs had lower HNR, intensity, and H1-H2 than vowels not adjacent to /ʔ/. These results are consistent with greater glottalization in these tokens, though these cues are unpredictably located and highly variable from speaker to speaker. An analysis with a more fine-grained time dimension is needed to determine if there are more subtle differences in timing between the three glottal-adjacent categories.

In contrast to these three measures, F0 is higher in vowels adjacent to /ʔ/ compared to other segments; this result is unexpected. However, higher F0 is associated with glottalization in some contexts in a variety of languages, though this is less frequent than glottalization linked to low F0. High F0 adjacent to /ʔ/ is also potentially consistent with Wheeler’s (2020) observations about final glottalization in Angaité. A study with clearer recordings and a more balanced sample of ages is needed to clarify this effect.

This description of voice quality is the first acoustic study of voice quality in this language family, and as such it represents an important step in the description of Enenlhet. It also points toward the role that holistic language description plays in phonetic description. This study investigates, but crucially leaves open, the question of the status of VʔVs and differences in the timing of glottalization dependent on the location of the glottal stop. Ultimately, further description of syllable structure, morphology, and acoustic features of other segments is necessary to provide a more robust analysis.

Chapter 5: Conclusion and Future Directions

The three studies in this dissertation showcase a range of variability in the Enenlhet vowel system. This chapter briefly describes directions for future research on Enenlhet based on the results of the studies presented here (Section 1) and discusses the role of corpus studies and naturalistic data in phonetic description more broadly (Section 2).

1. New Directions for Enenlhet Research

Each study in this dissertation suggests many directions for future research. To keep this section relatively brief, I discuss one or two open questions from each chapter.

1.1. Vowel Quality

Perception studies are needed to flesh out how speakers discriminate between Enenlhet vowel qualities, given the small F1 range and the overlap in the center of the F1 x F2 space. Since the /a/ category stretches into F1 ranges also occupied by /e/ and /o/, perception studies can evaluate the relative weight of F1 and F2 in discriminating between the non-low vowels and /a/. Perception studies are also needed to investigate the weight of other cues, such as duration and intensity, in making category distinctions. They can also provide insight into how consonantal context affects vowel categorization, perhaps especially in the case of nasals and glottals, both of which have substantial effects on the acoustic salience vowel formants (due to antiformants and aperiodicity, respectively).

In addition to future perception studies of Enenlhet vowel discrimination, more detailed work is needed to determine the phonemic categories of the language. As discussed previously, Chapter 2 relies on Enenlhet orthography and assumes that the grapheme used to represent each vowel corresponds to an underlying phoneme; that is, I have assumed that everything written with <a> corresponds to some phoneme /a/. Since there are some minimal pairs or minimal environments in the corpus, and as Chapter 2 does turn up the expected categories, this assumption seems reasonable as a starting point. However, the same native speaker assisted on all the transcriptions in Heaton (2019–), and this speaker's phonemic categories/orthographic choices may not correspond to the categories in other speakers' grammar. It may be the case that not all speakers have the same phonemes in each word, or even the same phoneme inventory. Given the methodological limitations in Chapter 2, along with the wide variation between speakers, future work is needed to thoroughly probe speaker-specific variability.

1.2. Vowel Duration

One of the main topics of investigation in the vowel duration study was lexical stress, and Chapter 3 leaves this question open. To determine whether lexical stress is present in Enenlhet, and where, a phonetic study that accounts for Enenlhet morphological structure is needed. Having not found a fixed lexical stress position suggests that if there is stress, either it is not fixed, or it depends on some other factors. Elliott (2021) provides some preliminary evidence that Enxet stress is sensitive to syllable structure and morphological structure. Chapter 3 did not account for morphological structure because a morphological analysis was not readily available. Furthermore, the WORD POSITION (RIGHT) and WORD POSITION (LEFT) factors in Chapter 3 used orthographic word edges as word boundaries. However, assuming that an optimal phonological or morphological word can be identified in Enenlhet, the orthographic word may not correspond to either, and, as noted in Chapter 1, there is some reason to suspect that this is the case in Enenlhet (e.g., clitics written as separate words). Even within the phonological word, the stress domain (if present) may be limited to certain syllables. Therefore, future work should attend to morphological structure and use a more nuanced definition of the "word". This could be achieved either by limiting the study to a particular morpheme class (nouns, or perhaps verb stems), or by adjusting the word position variable(s) to attend to morphological structure.

1.3. Voice Quality

As indicated in Chapter 4, a study of the Vʔ, ʔV, and VʔV contexts with a more fine-grained time dimension is necessary to uncover the timing differences between the three. However, Chapter 4 also presents a larger question. Chapter 2 shows that the Enenlhet vowels do not maximally use the F1 x F2 space, and Chapter 4 shows a high degree of variability in phonation. Qualitatively, I have observed that pre-pausal syllables in most utterances are devoiced, with devoicing sometimes spreading across the final two or even three syllables in an utterance. Many speakers also demonstrate a high degree of aperiodicity and relatively low intensity, even when vowels are not adjacent to a /ʔ/ or major prosodic boundary. As in Enxet, vowels are often deleted, especially when they do not fall into the prefix or stem syllables (see Elliott 2021 for a more thorough description of Enxet vowel deletion). All these features suggest a phonology which does not particularly prioritize vowels. Future work on Enenlhet should investigate the ways in which the variable phonation observed in Chapter 4 is related to these other features of Enenlhet vowels and, subsequently, how this unusual vowel system operates in the language's phonology more generally.

2. The Role of Corpus Studies in Phonetic Research

These three studies of Enenlhet reveal an enormous range of inter-speaker (and intra-speaker) variation. The data from the Enenlhet corpus show a much higher degree of variability than is usually observed in laboratory studies. This variability surfaces in the results in a variety of ways. Contextual effects on vowel duration are quite small in comparison to what has been attested cross-linguistically. While the effects of these factors in Enenlhet may indeed be smaller than in many other languages, speech rate differences between speakers and within monologues are likely at least partially responsible. Vowel quality is also highly variable; in this way, Enenlhet is similar to its sister languages. However, Chapter 2 shows that speakers' vowel systems are apparently quite flexible, with some speakers showing a broader F2 spread than others. Notably, the speakers whose vowel systems appear most compact in the F2 dimension are those with the greatest number of tokens. The data are unbalanced because they are drawn from a corpus, and there is no way to determine whether additional tokens from the less-represented speakers might reveal plots similar to CA and MRR, who made up the majority of the corpus. Finally, Chapter 4 showed massive variability in voice quality between speakers and contexts which will require many further studies to fully untangle.

While the nature of corpus data somewhat prevents me from making firm conclusions about the aspects of the Enenlhet vowel system that I set out to investigate, this uncertainty is a critical part of the investigative process. Since Enenlhet is under-studied, there are basically two possible routes to investigate any aspect of its grammar (in this case, phonetics). One option is to select a specific phenomenon and conduct targeted studies that carefully document that phenomenon while controlling for all others. The other option is the one that this dissertation presents: examine highly variable speech, perhaps providing much less generalizable or clear results but simultaneously sketching the shape of the language's quirks as it is most frequently experienced and produced by speakers.

Most well-studied languages already have vast amounts of data available for analysis. Due to these large corpora and robust previous research, dialect variability, inter-speaker variation, and speaker-specific idiosyncrasies, are already easily observable. From this starting point, phoneticians isolate phenomena for further research, and they contextualize their results within this well-attested linguistic variability. Anyone reading an experimental study of “English”, “German”, or “Japanese” understands that the results from in the carefully designed dataset represent specific effects produced by *some* speakers under *specific* conditions, and not necessarily by *every* speaker in *all* conditions. In other words, we tacitly situate experimental results within our knowledge of the language based on many other studies of that language and through many available works build up our understanding of the “language” as a cohesive, variable, system.

The situation of Enenlhet, and other under-studied languages, is very different. For these languages, robust corpora are not available, or at least not familiar to many linguists. For this reason, it is not trivial to select a specific phenomenon for research and for readers to situate experimental results in the context of the language’s overall variability. The range of variability is unknown. The studies here provide a window into the variability present in spontaneous Enenlhet speech, and, in doing so, they provide a starting point for the types of comparisons and questions that studies of more well-studied languages take as a given. Detailing the variability present in an Enenlhet corpus allows future work to pick out specific areas of interest to describe and contextualize against the observed variability in the corpus.

This set of studies indicates that detailed phonetic description based on a relatively un-controlled corpus, even of under-described languages, even when a large corpus and background literature is not available, is achievable. Not only is such a goal achievable, but it represents what I have come to see as a critical step in language description: phonetic corpus studies provide valuable information about the range of variability available to the phenomenon under study. Experimental work, in contrast, is narrower in focus, providing more close-grained information. Both are critical to the project of fully describing a linguistic system.

Appendix A

Abbreviations used in this dissertation are as follows. Abbreviations in morphological glosses are adapted from the source material where necessary to maintain consistency.

1	FIRST PERSON
2	SECOND PERSON
3	THIRD PERSON
CAUS	CAUSATIVE
DIRECT	DIRECT
IMP	IMPERATIVE
INDIR	INDIRECT
FEM	FEMININE
MASC	MASCULINE
NONFIRST	NON-FIRST PERSON
PL	PLURAL
PRIM	PRIMATIVE
PRO	PRONOUN
REAL	REALIS
REP	REPETITIVE
SG	SINGULAR

Appendix B

The following tables present examples of the attested consonant clusters. All examples are taken from Unruh, Kalisch, and Romero (2003). The top row indicates the first consonant in the cluster, and the leftmost column indicates the second consonant in each cluster. Relevant clusters have been bolded for ease of reading. All examples appear using the Enenlhet orthography. Column and row headers are in the IPA.

C2	**C1**					
	p	**t**	**k**	**q**	**m**	**n**
p	a**pp**ok		ngelie**kp**elkak	ngko**qp**alhqatek	**mp**aihak	
t	ne**pt**ame	ko**tt**eie	**kt**oskama	nenna**qt**engkeskama		ngkosa**nt**a'
k	ngka**pk**e'	a**tk**a	nempo**kk**anma	ngkoia**qk**anmok	kopaso**mk**om'	seppeiomapa**nk**ok
ʔ	mmo**p'**a'	ngkelio**t'**a'	mmama**k'**ak		hanetolhe**m'**a	la**n'**o
q	a**pq**anet					
m	a**pm**angkake'	nempakhe**tm**a	**km**eiakhelha	ngkopa**qm**etek	ne**mm**ahai'a	pke**nm**o
n	**pn**aqtoskama	ngko**tn**ehek	a**kn**ekhak	asia**qn**ek	a**mn**ek	**nn**aivehe'
s	**ps**akho	a**ts**oma'	lhvase**ks**ek	pa**qs**olval		**ns**etko'ok
h	ie**ph**opai'a	ngkona**th**oho'	se**kh**e'	nna**qh**apa	ngkoiangvo**mh**o'	apie**nh**ak
ɬ	a**plh**engkek		**klh**ekmo	mpa**qlh**ek		ene**nlh**et
j	**pi**apom	a**ti**anvakha	**ki**eto			a**ni**a'
w	**pv**esai'a	a**tv**ok	**kv**eno	akia**qv**atem		atia**nv**omhok
l	**pl**oka		a**kl**oma	iota**ql**a		

C2	**C1**						
	ŋ	**s**	**h**	**ɬ**	**j**	**w**	**l**
p		a**sp**atmek					ngke**lp**epma'a
t		**st**ahak		a**lht**a	ngka**it**e		ske**lt**elhnama
k	**ngk**olhong	aspene**sk**ek		**lhk**e	niava**ik**e	pta**vk**e	iame**lk**et
ʔ	elialhe**ng'**a				apveta**i'**ak	a**v'**alhok	
q	e**ngq**anet			heie**lhq**ak			a**lq**ama
m		mena**sm**a		a**lhm**a'	na**im**ong		aktaqma**lm**a
n		anva**sn**ek		a**lhn**ankok	a**in**ek		
s	ngkelketa**ngs**engke				kelpa**is**akmek		etaqme**ls**ap
h	e**ngh**aikok	a**sh**ankek			a**ih**angvomok	asa**vh**o	
ɬ					skela**ilh**ek		
j	e**ngi**aha'	a**si**enmemaha		ense**lhi**e'			e**li**ota'
w	e**ngv**a'	nie**sv**ehe'		ko**lhv**a'	apa**iv**oma		ke**lv**ana
l	ne**ngl**eklema	a**sl**oka	ngketlhengahlkehe'	a**lhl**eng'ak			ia**ll**ehe'

References

Abramson, Arthur S. & Nianqi Ren. 1990. Distinctive vowel length: Duration vs. spectrum in Thai. *Journal of Phonetics* 18. 79–82.

Adelaar, Willem F. H. 1984. Grammatical vowel length and the classification of Quechua dialects. *International Journal of American Linguistics* 50(1). 25–47. URL: https://www.jstor.org/stable/1265196.

Adelaar, Willem F. H. & Pieter Muysken. 2004. *The languages of the Andes.* Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511486852.

Ahn, Hyunkee. 2000. The H1*-H2* measure. *Korean Society of Speech Sciences* 7(2). 85–95.

Audacity Team. 2022. Audacity®: Free audio editor and recorder [Computer program]. Version 3.2. URL: https://www.audacityteam.org.

Avelino, Heriberto, Eurie Shin & Sam Tilsen. 2011. The phonetics of laryngealization in Yucatec Maya. In Heriberto Avelino (ed.), *New perspectives in Mayan linguistics,* 1–20. Newcastle upon Tyne: Cambridge Scholars Publishing.

Avelino, Heriberto. 2010. Acoustic and electroglottographic analyses of nonpathological, nonmodal phonation. *Journal of Voice* 24(3). 270–280. DOI: https://doi.org/10.1016/j.jvoice.2008.10.002.

Bani-Yasin, Raslan & Jonathan Owens. 1987. The phonology of a Northern Jordanian Arabic dialect. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 137(2). 291–331.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67. 1–48. DOI: https://doi.org/10.18637/jss.v067.i01.

Beaufort, Richard & Alain Ruelle. 2006. eLite: système de synthèse de la parole à orientation linguistique. *Journées d'études sur la parole* 26. Dinard, France.

Becker-Kristal, Roy. 2010. *Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus*. UCLA Ph.D. dissertation.

Beguš, Gašpar. 2017. Effects of ejective stops on preceding vowel duration. *Journal of the Acoustical Society of America* 142(4). 2168–2184. DOI: https://doi.org/10.1121/1.4007728.

Bellem, Alex & Janet C. Watson. 2014. Backing and glottalization in three SWAP language varieties. In Manueal Giolfo (ed.), *Arabic linguistics: Traditional and new theoretical approaches* (*Journal of Semitic Studies* supplement 34), 169–207. Oxford: Oxford University Press.

Benguerel, André-Pierre & Helen A. Cowen. 1974. Coarticulation of upper lip protrusion in French. *Phonetica* 30. 41–55.

Benguerel, André-Pierre. 1971. Duration of French vowels in unemphatic stress. *Language and Speech* 14(4). 383–391. DOI: https://doi.org/10.1177/002383097101400406.

Bennett, Ryan. 2018. La tensión vocálica en el kaqchikel de Sololá, Guatemala: un estudio preliminar. In Esther Herrera Zendejas & Verónica Reyes Taboada (eds.), *Fonología segmental: procesos e interacciones*, 13–46. Mexico City: Colegio de México. DOI: https://doi.org/10.2307/j.ctvkwnq26.4.

Berkovits, Rochele. 1994. Durational effects in final lengthening, gapping, and contrastive stress. *Language and Speech* 37(3). 237–250.

Blankenship, Barbara. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30. 163–191. DOI: https://doi.org/10.006/jpho.2001.0155.

Boersma, Paul & David Weenink. 2023. Praat: Doing phonetics by computer [Computer program]. Version 6.3.18. URL: http://www.praat.org/.

Borroff, Marianne. 2007. *A landmark underspecification account of the patterning of glottal stop*. Stony Brook University Ph.D. dissertation.

Bradlow, Ann R. 1993. *Language-specific and universal aspects of vowel production and perception: A cross-linguistic study of vowel inventories.* Cornell University Ph.D. dissertation.

Bradlow, Ann R. 1995. A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America* 97. 1916–1924. DOI: https://doi.org/10.1121/1.412064.

Brunner, Jana & Marzena Żygis. 2011. Why do glottal stops and low vowels like each other? *International Congress of Phonetic Sciences (ICPhS)* 17. 376–379. URL: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/index.htm.

Buckley, Eugene. 1991. Kashaya closed-syllable shortening and prosodic syllabification. In Dawn Bates (ed.), *West Coast Conference on Formal Linguistics* 10. 65–74. Stanford, CA: Center for the Study of Language and Information.

Bye, Patrik, Elin Sagulin & Ida Toivonene. 2009. Phonetic duration, phonological quantity and prosodic structure in Inari Saami. *Phonetica* 66. 199–221.

Byrd, Dani & Elliot Saltzman. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31. 149–180.

Cambier-Langeveld, Tina. 1997. The domain of final lengthening in the production of Dutch. *Linguistics in the Netherlands.* DOI: https://doi.org/10.1075/avt.14.04cam.

Cambier-Langeveld, Tina. 1999. The interaction between final lengthening and accentual lengthening: Dutch versus English. *Linguistics in the Netherlands* 16(1). 13–25.

Campbell, Lyle. 2012a. Classification of the indigenous languages of South America. In Lyle Campbell & Verónica Grondona (eds.), *The indigenous languages of South America: A comprehensive guide* (World of Linguistics [WOL] 2), 59–166. Berlin: de Gruyter Mouton. DOI: https://doi.org/10.1515/9783110258935.59.

Campbell, Lyle. 2012b. Typological characteristics of South American indigenous languages. In Lyle Campbell & Verónica Grondona (eds.), *The indigenous languages of South America: A comprehensive guide* (World of Linguistics [WOL] 2), 259–330. Berlin: de Gruyter Mouton. DOI: https://doi.org/10.1515/9783110258035.

Campbell, Lyle. 2013. Language contact and linguistic change in the Chaco. *Revista Brasileira de Linguística Antropológica* 5(2). 259–291. DOI: https://doi.org/10.26512/rbla.v5i2.16268.

Caplow, Nancy. 2016. Stress patterns and acoustic correlates of stress in Bati Tibetan. *Himalayan Linguistics* 15(2). 1–49.

Cartes Jara, Horacio Manuel, José Molinas Vega, Elizabeth Barrios Kück & Yolanda Barrios Sosa (eds.). 2015. *Atlas de comunidades de pueblos indígenas en Paraguay 2012: Familia lingüística lengua mascoy*. Fernando de la Mora, Paraguay: Dirección General de Estadística, Encuestas y Censos. URL: www.dgeec.gov.py.

Carvalho, Silvia M. Schmuziger. 1992. Chaco: Encruzilhada de povos e “melting pot” cultural. In Manuela Carneiro da Cunha (ed.), *História dos índios no Brasil* (2nd. ed.), 457–474. São Paulo, Brazil: Editora Schwarcz Ltda.

Chelliah, Shobhana L. 2001. The role of text collection and elicitation in linguistic field-work. In Paul Newman & Martha Ratliff (eds.), *Linguistic fieldwork*, 152–165. Cambridge: Cambridge University Press.

Chelliah, Shobhana L. & Willem J. de Reuse. 2010. Semantics, pragmatics, and text collection. In Shobhana L. Chelliah & Willem J. de Reuse (eds.), *Handbook of descriptive linguistic fieldwork*, 314–448. Dordrecht: Springer.

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

Coto-Solano, Rolando & Sofía Flores Solórzano. 2017. Comparison of two forced alignment systems for aligning Bribri speech. *CLEI Electronic Journal* 20(1). Paper 2.

Coto-Solano, Rolando, Sally Akevai Nicholas & Samantha Wray. 2018. Development of natural language processing tools for Cook Islands Maōri. *Australian Language Technology Association Workshop*. 26–33.

Crowhurst. Megan. 2018. The joint influence of vowel duration and creak on the perception of internal phrase boundaries. *Journal of the Acoustical Society of America* 143(3). EL 147– EL 153. DOI: https://doi.org/10.1121/1.5025325.

Davidson, Lisa. 2020. The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages. *WIREs Cognitive Science*. DOI: https://doi.org/10.1002/wcs.1547.

Davidson, Lisa. 2021. Effects of word position and flanking vowel on the implementation of glottal stop: Evidence from Hawaiian. *Journal of Phonetics* 88. https://doi.org/10.1016/j.wocn.2021.101075

Delattre, Pierre. 1966. A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics (IRAL)* 4(3). 183–198.

DiCanio, Christian. 2012. Coarticulation between tone and glottal consonants in Itunyoso Trique. *Journal of Phonetics* 40. 162–176.

DiCanio, Christian, Hosung Nam, Jonathan D. Amith, & Rey Castillo García. 2015. Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec. *Journal of Phonetics* 48. 45–49.

Dilley, Laura & Stefanie Shattuck-Hufnagel. 1995. Individual differences in the glottalization of vowel-initial syllables. *Journal of the Acoustical Society of America* 97. 3418. DOI: https://doi.org/10.1121/1.412997.

Dilley, Laura, Stefanie Shattuck-Hufnagel & Mari Ostendorf. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24. 423–444. DOI: https://doi.org/10.1006/jpho.1996.0023.

Dinard, France S., Richard Beaufort & Alain Ruelle. 2006. eLite: système de synthèse de la parole à orientation linguistique. In *Proceedings XXVIe Journées d'Etude sur la Parole*, 509–512.

Disner, Sandra Ferrari. 1983. Vowel quality: The relation between universal and language-specific factors. UCLA Ph.D. dissertation (*UCLA Working Papers in Phonetics* 58).

ELAN. 2022. [Computer program]. Nijmegan: Max Planck Institute for Psycholinguistics, The Language Archive. Version 6.4. URL: https://archive.mpi.nl/tla/elan

Elliott, John. 2016. For bilinguals, Enxet vowel spaces smaller than Spanish. *The Journal of the Acoustical Society of America* 140(4). 1307.

Elliott, John. 2021. *A grammar of Enxet Sur.* Manoa, HI: University of Hawaii at Manoa Ph.D. dissertation.

Enenlhet. 2022. *Catalogue of Endangered Languages*. Manoa, HI: University of Hawaii at Manoa. Accessed 7 July 2022. URL: https://endangeredlanguages.com/lang/2141.

Enlhet. 2022. *Catalogue of Endangered Languages.* Manoa, HI: University of Hawaii at Manoa. Accessed 4 December 2023. URL: https://endangeredlanguages.com/lang/4490.

Epps, Patience & Lev Michael (eds.). 2023. *Amazonian languages: An international handbook* (Handbooks of Linguistics and Communication Science [HSK] 44.1–44.2). Berlin: de Gruyter Mouton.

Esposito, Anna. 2002. On vowel height and consonantal voicing effects: Data from Italian. *Phonetica* 59. 197–231.

Fabre, Alain. 2013. Los pueblos del Gran Chaco y sus lenguas, primera parte: Los enlhet-enenlhet del Chaco Paraguayo. *Suplemento Antropológico* 40(1). Last modified: 24 December 2013. URL: https://alain-fabre.fi/ling.fi/afkirjoitukset.htm.

Fant, Gunnar, Lennart Nord & Anita Kruckenberg. 1987. Segmental and prosodic variabilities in connected speech: An applied data-bank study. *International Congress of Phonetic Sciences (ICPhS)* 6. 102–105.

Félix-Brasdefer, Cesar J. 2007. Natural speech vs. elicited data: A comparison of natural and role play requests in Mexican Spanish. *Spanish in Context* 4(2). 159–185.

Fintoft, Knut. 1961. The duration of some Norwegian speech sounds. *Phonetica* 7. 19–39.

Fisher-Jørgensen, Eli. 1989. Phonetic analysis of the stød in Standard Danish. *Phonetica* 46. 1–59.

Flege, James E. 1989. Differences in inventory size affect the location but not the precision of tongue positioning in vowel production. *Language and Speech* 32(2). 123–147.

Fletcher, Janet. 1991. Rhythm and final lengthening in French. *Journal of Phonetics* 19. 193–212.

Frazier, Melissa. 2009. *The production and perception of pitch and glottalization in Yucatec Maya.* University of North Carolina at Chapel Hill Ph.D. dissertation.

Frazier, Melissa. 2013. The phonetics of Yucatec Maya and the typology of laryngeal complexity. *STUF*, *Akademie Verlag* 66 (1). 7–21.

Fulop, Sean & Chris Golston. 2008. Breathy and whispery voicing in White Hmong. *Journal of the Acoustical Society of America* 123(5). 3883. DOI: https://doi.org/10.1121/1.2935803.

Garellek, Marc. 2012a. Word-initial glottalization and voice quality strengthening. *UCLA Working Papers in Phonetics* 111. 92–122.

Garellek, Marc. 2012b. The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics* 40. 152–161.

Gendrot, Cédric & Martine Adda-Decker. 2007. Impact of duration and vowel inventory size on formant values of oral vowels: Automated formant analysis of eight languages. *International Congress of Phonetic Sciences (ICPhS)* 16. 1417–1420.

Goldman, Jean-Philippe. 2011. EasyAlign: An automatic phonetic alignment tool under Praat [Praat plugin]. Version 04.2012. URL: http://latlcui.unige.ch/phonetique/easyalign.php.

Gomes, Antonio Silva. 2013. *Sanapaná uma lingua maskoy: Aspectos gramaticais.* São Paulo: Universidade estadual de Campinas, Instituto de estudio da linguagem Ph.D. dissertation.

Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29. 383–406. DOI: https://doi.org/10.006/jpho.2001.0147.

Guillaume, Antoine. Forthcoming. Tacanan languages. In Patience Epps & Lev Michael (eds.), *Amazonian languages: An international handbook* (Handbooks of Linguistics and Communication Science [HSK] 44.3). Berlin: de Gruyter Mouton.

Guion, Susan G. 2003. The vowel systems of Quichua-Spanish bilinguals. *Phonetica* 60. 98–128. DOI: https://doi.org/10.1159/000071449.

Gysel, Jens van. 2017. Temporal predicative particles in Sanapaná and the Enlhet-Enenlhet language family (Paraguay): A descriptive and comparative study. Leiden University M.A. thesis. URL: https://hdl.handle.net/1887/52531.

Gysel, Jens van. 2022. The influence of language shift on Sanapaná vowels: An exemplar-based perspective. *Language Variation and Change* 34(1). 107–135. DOI: https://doi.org/10.1017/S0954394522000023.

Hardin, Joanna & David Rocke. 2005. The distribution of robust distances. *Journal of Computational and graphical Statistics* 14(1). 1–19. DOI: https://doi.org/10.1198/106005X77685.

Heaton, Raina. 2019–. Enenlhet documentation. https://aila.utexas.org. *The Archive of the Indigenous Languages of Latin America*. Access: public. PID ailla:266555.

Hernandez Laura A., Scott James Perry & Benjamin V. Tucker. 2023. The effect of stress and speech rate on vowel quality in spontaneous Central Mexican Spanish. In Radek Skarnitzl & Jan Volín (eds.), *International Congress of Phonetic Sciences (ICPhS)* 20. 694–698.

Hidden Markov Model Toolkit (HTK) [portable toolkit]. Cambridge, UK: University of Cambridge. Version 3.4. URL: https://htk.eng.cam.ac.uk.

Hillenbrand, James M. & Robert A. Houde. 1996. Role of F_0 and amplitude in the perception of intervocalic glottal stops. *Journal of Speech and Hearing Research* 39. 1182–1190.

Hillenbrand, James M., Michael J. Clark & Terrance M. Nearey. 2001. Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America* 109. 748–763.

Hockey, Beth Ann & Zsuzsanna Faygal. 1999. Phonemic length and pre-boundary lengthening: An experimental investigation on the use of durational cues in Hungarian. In John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey (eds.), *International Congress of Phonetic Sciences (ICPhS)* 14. URL: http://www.internationalphoneticassociation.org/icphs/icphs1999.

Holliday, Nicole & Sean Martin. 2017. Vowel categories and allophonic lowering among Bolivian Quechua-Spanish bilinguals. *Journal of the International Phonetic Association* 48(2). 199–222. DOI: https://doi.org/10.1017/S0025100317000512.

Hombert, Jean-Marie. 1978. Consonant types, voice quality, and tone. In Victoria Fromkin (ed.), *Tone: A linguistic survey*, 77–111. New York: Academic Press.

Hothorn, Torsten, Achim Zeileis, Richard W. Farebrother, Clint Cummins, Giovanni Millo & David Mitchell. 2022. *Lmtest: Testing Linear Regression Models.* [R Package]. Version 3.1-3. URL: https://cran.r-project.org/web/packages/lmtest/index.html

House, Arthur S. & Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25(1). 105–113.

Hubert, Mia, Michiel Debruyne & Peter Rousseeuw. 2017. Minimum covariance determinant and extensions. *WIREs Computational Statistics* 10(3). DOI: https://doi.org/10.1002/wics.1421.

Huggins, A. W. F. 1972. Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America* 51(4.2). 1270–1278. DOI: https://doi.org/10.1121/1.1912971.

Iseli, Markus, Yen-Liang Shue & Abeer Alwan. 2006. Age- and gender dependent analysis of voice source characteristics. *International Conference on Acoustics, Speech, and Signal Processing.* Tolouse, France: Institute of Electrical and Electronics Engineers Signal Processing Society.

Jong, Kenneth de & Bushra Adnan Zawaydeh. 1999. Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics* 27. 3–22. DOI: https://doi.org/0095-4470/99/010003.

Kalisch, Hannes. 2019. Orden de los argumentos en enlhet. In Valeria A. Belloro (ed.), *La interfaz sintaxis-pragmática: Estudios teóricos, descriptivos y experimentales*, 143–177. Berlin: de Gruyter Mouton.

Keating, Patricia, Marc Garellek & Jody Kreiman. 2015. Acoustic properties of different kinds of creaky voice. In The Scottish Consortium for ICPhS 2015 (ed.), *International Congress of Phonetic Sciences (ICPhS)* 18. URL: https://www.internationalphoneticassociation.org/icphs/icphs2015

Kirk, Paul L., Jenny Ladefoged & Peter Ladefoged. 1993. Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec. In Anthony Mattina and Timothy Montler (eds.), *American Indian linguistics and ethnography in honor of Laurence C. Thompson*, 435–450. Missoula, MT: University of Montana.

Klatt, Dennis H. & William E. Cooper. 1975. Perception of segment duration in sentence contexts. In Antonie Cohen, Seib Nooteboom (eds.) *Structure and Process in Speech Perception* (Communication and Cybernetics 11), 69–89. Berlin/Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-642-81000-8_5.

Klatt, Dennis. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59(5). 1208–1221. DOI: https://doi.org/10.112/1.380986.

Korkmaz, Yunus & Aytuğ Boyacı. 2018. Examining vowels' formant frequency shifts caused by preceding consonants for Turkish Language. *Journal of Engineering and Technology* 2(2). 38–47.

Lass, Roger. 1984. Vowel system universals and typology: Prologue to theory. *Phonology Yearbook* 1. 75–111.

Lehiste, Ilse. 1970. *Suprasegmentals.* Cambridge, MA: MIT Press.

Lenth, Russell V., Ben Bolker, Paul Buerkner, Iago Giné-Vázquez, Maxime Herve, Maarten Jung, Jonathon Love, Fernando Miguez, Hannes Riebl & Henrik Singmann. 2022. *Estimated marginal means aka least squares means.* [R Package]. Version 1.8.2. URL: https://CRAN.R-project.org/package=emmeans

Letterman, Rebecca S. 1994. A phonetic study of Sinhala syllable rhymes. *Working Papers of the Cornell Phonetics Laboratory* 9. 155–181.

Liljencrants, Johan & Björn Lindblom. 1972. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48(4). 839–862.

Lindblom, Bjorn & Karin Rapp. 1973. Some temporal regularities of spoken Swedish. In Gunnar Fant & Marcel A.A. Tatham (eds.), *Auditory analysis and perception of speech*, 1–59. London: Academic Press.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John Ohala & Jeri Jaeger (eds.), *Experimental phonology*, 13–44. Orlando: Academic Press.

Lippus, Pärtel, Eva Liina Asu, Pire Teras & Tuuli Tuisk. 2013. Quantity-related variation of duration, pitch and vowel quality in spontaneous Estonian. *Journal of Phonetics* 4. 17–26.

Loporcaro, Michele. 2015. *Vowel length from Latin to Romance.* Oxford: Oxford University Press.

Maddieson, Ian & Karen Emmorey. 1985. Relationship between semi-vowels and vowels: Cross-linguistic investigations of acoustic difference and coarticulation. *Phonetica* 42(4). 163–174. DOI: https://doi.org/10.1159/000261748.

Maddieson, Ian. 1985. Phonetic cues to syllabification. In Victoria A. Fromkin (ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged*, 203–221. Orlando/London: Academic Press.

Maddieson, Ian. 2013. Vowel quality inventories. In Matthew Dryer & Martin Haspelmath (eds.), *WALS Online* (v. 2020.3) [Data set]. Zenodo. DOI: https://doi.org/10.5281/zenodo.7385533.

Malécot, André. 1975. The glottal stop in French. *Phonetica* 29. 51–63.

Manelis Klein, Harriet E. & Louisa R. Stark. 1977. Indian languages of the Paraguayan Chaco. *Anthropological Linguistics* 19(8). 378–401.

Martin, James G. & H. Timothy Bunnell. 1982. Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance* 8(3). 473–488.

Meliá, Bartomeu, Luís Farré Maluquer & Alfonso Pérez Peñasco. 1995. *El guaraní a su alcance.* Asunción, Paraguay: Centro de Estudios Paraguayos "Antonio Guasch".

Miller, Amanda L. 2007. Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics* 35(1). 56–84. DOI: https://doi.org/10.1016/j.wcon.2005.11.001.

Mitleb, Fares M. 1984. Voicing effect on vowel duration is not an absolute universal. *Journal of Phonetics* 12. 23–27.

Mitterer, Holger, Sahyang Kim & Taehong Cho. 2019. The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language* 108. DOI: https://doi.org/10.1016/j.jml.2019.104034.

Moran, Steven, Daniel McCloy & Richard Wright (eds.). 2019. *PHOIBLE 2.0.* Jena: Max Planck Institute for the Science of Human History. Accessed on 15 February 2024. URL: http://phoible.org.

Nakai, Satsuki, Sari Kunnari, Alice Turk, Kari Suomi & Rikka Ylitalo. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 37. 29–45.

Nooteboom, Sieb. 1972. *Production and perception of vowel duration: A study of durational properties of vowels in Dutch.* Rijksuniversiteit Utrecht PhD dissertation.

Öhman, S. E. G. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39. 151–168. DOI: https://doi.org/10.1121/1.1909864.

Oller, D. Kimbrough. 1973. The effect of position in utterance on speech segment duration in English. *The Journal of the Acoustical Society of America* 54. 1235–1247. DOI: https://doi.org/10.1121/1.1914393.

Ortega-Llebaria, Marta & Pilar Prieto. 2011. Acoustic correlates of stress in Central Catalan and Castilian Spanish. *Language and Speech* 54(1). 72–97.

Peña, Jailyn, Lisa Davidson & Shmico Orosco. 2021. The independnence of phrasal creak and segmental glottalization in American English. *JASA Express Letters* 1. https://doi.org/10.1121/10.0005601.

Pierrehumbert, Janet & David Talkin. 1992. Lenition of /h/ and glottal stop. In Gerard J. Docherty & D. Robert Ladd (eds.), *Papers in Laboratory Phonology,* vol. 2: *Gesture, Segment, Prosody,* 90–127. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511519918.005.

Plauché, Madelaine C., Rosemary Beam de Azcona, Rungpat Roengpitya & William F. Weigel. 1998. Glottalized sonorants: A phonetic universal? In Benjamin Bergen, Madelaine C. Plauché & Ashlee C. Bailey (eds.), *Berkeley Linguistcs Society (BLS)* 24, 381–390.

Pompino-Marschall, Bernd & Marzena Żygis. 2010. Glottal marking of vowel-initial words in German. *ZAS Papers in Linguistics* 52. 1–17.

Powys, E. L. G. 1929. *A practical lengua grammar.* Oxford University M.A. thesis.

R Core Team. 2023. *R: A language and environment for statistical computing* [Computer program]. Vienna, Austria: R Foundation for Statistical Computing. Version 4.2.1. DOI: https://www.R-project.org.

Redi, Laura & Stefanie Shattuck-Hufnagel. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29. 407–429. DOI: 10.006/jpho.2001.0145.

Riggs, Daniel. 2016. *Praat scripting tutorial* [website]. Accessed 12 July 2022, URL: www.praatscriptingtutorial.com.

Ritchie Key, Mary & Bernard Comrie (eds.). n.d. Sanapana. *Intercontinental Dictionary Series.* Leipzig: Max Planck Institute for Evolutionary Anthropology. URL: https://ids.clld.org/contributions/299. Accessed 27 November 2023.

Rosen, Kristin M. 2005. Analysis of speech segment duration with the lognormal distribution: A basis for unification and comparison. *Journal of Phonetics* 33. 411–426.

Rossum, Guido & Fred Drake. 1995. *Python reference manual.* Amsterdam: Centrum voor Wiskunde en Informatica Amsterdam.

Rousseeuw, Peter. 1985. Multivariate estimation with high breakdown point. In W. Grossman, G. Plfug, I. Vincze & W. Werz (eds.), *Mathematical statistics and applications, Volume B*, 283–297. The Netherlands: Dordrecht-Reidel.

Sanker, Chelsea. 2020. A perceptual pathway for voicing-conditioned vowel duration. *Journal of the Association for Laboratory Phonology* 11(1). DOI: http://dx.doi.org/10.5334/labphon.268.

Saunders, Gladys E. 1978. Evolution of vowel length in Gallo-Italian. *General Linguistics* 18. 14–27.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée & Christian Abry. 1997. Major trends in vowel system inventories. *Journal of Phonetics* 25. 233–253.

Seyfarth, Scott & Marc Garellek. 2015. Coda glottalization in American English. In The Scottish Consortium for ICPhS 2015 (ed.), *International Congress of Phonetic Sciences (ICPhS)* 18. URL: https://www.internationalphoneticassociation.org/icphs/icphs2015.

Shue, Yen-Liang, Gang Chen & Abeer Alwan. 2010. On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. *Interspeech.* 26–30 September, Makuhari, Chiba, Japan. 33-37.

Speeter Beddor, Patrice, James D. Harnsberger & Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30. 591–627.

Speeter Beddor, Patrice, Rena Arens Krakow & Louis M. Goldstein. 1986. Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook* 3. 197–217.

Stevens, Kenneth N. & Arthur S. House. 1963. Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research* 6. 111–128.

Stevens, Kenneth N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In Edward E. David & Peter B. Denes, *Human communication: A unified view*, 51–66. New York: McGraw Hill.

Stevens, Kenneth N. 1989. On the quantal nature of speech. *Journal of Phonetics* 17. 3–46.

Stevens, Kenneth N. 1998. *Acoustic phonetics.* Cambridge, MA/London: MIT Press.

Tallman, Adam. 2020. Beyond grammatical and phonological words. *Language and Linguistic Compass* 14e. DOI: https://doi.org/10.1111/lnc3.12364.

Thurgood, Ela. 2004. Phonation types in Javanese. *Oceanic Linguistics* 43(2). 277–295. URL: https://www.jstor.org/stable/3623359.

Turk, Alice E. & Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35. 445–472.

Ueyama, Motoko. 1997. An experimental study of vowel duration in phrase-final contexts in Japanese. *UCLA Working Papers in Phonetics* 97. 174–182.

Unruh, Ernesto & Hannes Kalisch. 1997. *Moya'ansaeclha' Nengelpayvaam Nengeltomha Enlhet* (Biblioteca Paraguaya de Antropología, 27). Ya'alve-Saanga/Asunción: Comunidad Enlhet/ CEADUC.

Unruh, Ernesto & Hannes Kalisch. 1999. Escribir en idioma Enlhet para hacer crecer lo propio. *Suplemento Antropológico* 34(1). 101–176.

Unruh, Ernesto & Hannes Kalisch. 2003. Enlhet-Enenlhet: Una familia lingüística chaqueña. *Thule, Rivista italiana di studi americanistici* 14/15. 207–231.

Unruh, Ernesto, Hannes Kalisch & Manolo Romero. 2003. *Enlhet Apaivoma. Nentengiai'a nengiangveiakmoho neliatekamaha enenlhet apaivoma: Guía para el aprendizaje del idioma materno toba.* Ya'alve Saanga/Chaco paraguayo: Nengvaanemqueskama Nempayvaam Enenlhet. URL: www.enlhet.org.

Vaux, Bert & Bridget Samuels. 2015. Explaining vowel systems: Dispersion theory versus natural selection. *The Linguistic Review* 32(3). 573–599. DOI: https://doi.org/10.1515/tlr-2014-0028.

Wayland, Ratree, Scott Gargash & Allard Longman. 1995. Acoustic and perceptual investigation of breathy voice. *Journal of the Acoustical Society of America* 97(5). 3364. DOI: https://doi.org/10.1121/1.413011.

Whalen, Douglas, Christian DiCanio, Christopher Geissler & Hannah King. 2016. Acoustic realization of a distinctive, frequent glottal stop: The Arapaho example. *Journal of the Acoustical Society of America* 139. 2212–2213. DOI: https://doi.org/10.1121/1.4950615.

Wheeler, Paige-Erin. 2020. *Consonants and syllable structure in Angaité.* Austin, TX: University of Texas at Austin M.A. thesis.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis.* [R package] New York: Springer-Verlag New York. Version 3.4.2. URL: https://ggplot2.tidyverse.org.

Wood, Elizabeth. 2020. *Vowels and stress in Chichicastenango K'iche'.* University of Texas at Austin M.A. thesis.

Younes, Munther. 1995. On vowel shortening in Palestinian Arabic. In Mushira Eid (ed.), *Perspectives on Arabic linguistics VII*, 157–172. Amsterdam: John Benjamins.

Vita

Paige Erin Wheeler was raised in Indiana. She earned her B.A. in Spanish, with a concentration in Linguistics, from Grinnell College in 2016 and her M.A. in Linguistics from the University of Texas at Austin in 2020. In addition to her interest in phonetics and Indigenous South American languages, she has collaborated on research on diversity and access to higher education in the U.S., and she is an enthusiastic classroom instructor. She is an avid reader, a lover of guinea pigs and cats, a somewhat-chaotic player of tabletop RPGs, and a viciously competitive opponent in the board game Clue.

Permanent address (or email): paigeew95@gmail.com

This dissertation was typed by Paige Erin Wheeler.