Acoustic correlates and diagnostics of ATR in Boa-Leboale

Ahmed, Samuel

Citation

Ahmed, S. (2023). *Acoustic correlates and diagnostics of ATR in Boa-Leboale*.

Acoustic correlates and diagnostics of ATR in Boa-Leboale

Samuel Kumar Ahmed
s2817802
Leiden University

A Research Master's thesis supervised by:
Dr. E.D. Botma
Dr. G. de Wit

19 August 2023

Abstract

This thesis investigates a recurring issue found in fieldwork on Niger-Congo and Nilo-Saharan languages: how can a linguist tell if a given language has the vowels [ɪ ʊ], or [e o], or both? These two pairs of vowels – which differ in their height and ATR (‘advanced tongue root’) value – are notoriously difficult to tell apart acoustically, and their pronunciation varies considerably between languages and even speakers. Moreover, previous studies have only focused on how to distinguish these vowels in languages with both pairs; in languages with just one pair, how do we know if that pair is [ɪ ʊ] or [e o]?

To address this gap, the main acoustic study of this thesis looks at data from Boa-Leboale, a Bantu language spoken in the Democratic Republic of the Congo that has historically been described as only having [ɪ ʊ], not [e o]. Using the results and acoustic methodology of Starwalt’s (2008) landmark study, I construct a number of hypotheses to test two research questions; research question (I) investigates whether Boa-Leboale has one or two pairs of vowels and research question (II) attempts to identify whether a given set of vowels is more likely to be [ɪ ʊ] or [e o].

The results of this study indicate that Boa-Leboale does indeed have only one pair of vowels, and those vowels are most likely [ɪ ʊ], as suggested by previous descriptions. Given this promising result, I hope that further studies will be able to improve on the study presented here and solve the issue of identifying these vowels acoustically not just for Boa-Leboale, but for any language.

Acknowledgements

I would like to thank my supervisors, Bert Botma and Gerrit de Wit, for introducing me to the topic of this thesis, and in particular to Gerrit for sharing his data and insights with me throughout this process. This thesis was only possible thanks to the work of Gerrit de Wit, Dominique Banotanea Bapokanzo and the Boa team: Patrick Alipanazanga Masanga, Rev. Sébastien Banale Aleumba, Julienne Etindepua Embatia and, most importantly, Dieudonné Tazanaba Kadite, whose speech I was honoured to have the opportunity to analyse.

I also could not have finished this thesis without the support of my friends, my family and my boyfriend, who were all more than patient through this (unexpectedly) long process.

Contents

1 Introduction **5**

1.1 Topic background . . . 5

1.2 Research questions . . . 6

1.3 Thesis structure . . . 6

2 Literature background **8**

2.1 What is ATR? . . . 8

2.2 ATR vowel systems . . . 10

2.3 Distinguishing /1IU/ and /2IU/ 7-vowel systems . . . 12

2.3.1 Articulatory diagnostics . . . 12

2.3.2 Acoustic diagnostics . . . 13

2.3.3 Phonological diagnostics . . . 16

2.4 Starwalt (2008) . . . 17

2.4.1 Starwalt's methodology . . . 17

2.4.2 Starwalt's results . . . 21

2.4.3 Starwalt on distinguishing 7-vowel systems . . . 23

2.5 Conclusions . . . 24

3 Boa-Leboale **26**

3.1 Language introduction . . . 26

3.2 Boa-Leboale vowel system . . . 27

3.3 Vowel alternations . . . 28

3.3.1 Regressive assimilations . . . 28

3.3.2 Progressive assimilations . . . 30

3.3.3 Alternations overview . . . 32

3.4 Related vowel systems . . . 32

3.5 Conclusions . . . 34

4 Methodology **36**

4.1 Data collection and annotation . . . 36

4.2 Acoustic analysis . . . 37

4.2.1 Automatic measurement . . . 37

4.2.2 Points of measurement . . . 40

4.2.3 Formant frequency and bandwidth (F1, F2, F3, B1, ΔB1) 40
4.2.4 Formant amplitude (A1, A2, A1*–A2*) 43
4.2.5 Centre of gravity 45
4.2.6 Cepstral peak prominence 45
4.2.7 Data output 47
4.3 Statistical analysis 48
4.3.1 Descriptive statistics 48
4.3.2 Research questions and hypotheses 49
4.3.3 Significance tests 52

5 Results **56**
5.1 Descriptive statistics 56
5.1.1 Vowel chart 56
5.1.2 Breakdown by acoustic measure 56
5.1.3 Skewness, outliers and transformation 65
5.1.4 Comparing derived and underived vowels 69
5.2 Research question (I): One or two sets of Level 2 vowels? 71
5.2.1 Hypothesis 1: Unimodality 71
5.2.2 Hypothesis 2: Derived vs underived vowels 72
5.2.3 Discussion of RQ(I) results 74
5.3 Research question (II): /1IU/ or /2IU/? 76
5.3.1 Hypothesis 3: ΔB1 diagnostic 76
5.3.2 Hypothesis 4: A1*–A2* diagnostic 77
5.3.3 Hypothesis 5: Centre of gravity diagnostic 78
5.3.4 Hypothesis 6: CPP diagnostic 79
5.3.5 Hypothesis 7: Diagnostic using ΔB1 of /ɛ ʊ/ 79
5.3.6 Discussion of RQ(II) results 80
5.4 Summary of results and discussions 81

6 Conclusions **83**

A SpectralMeasures.Praat **89**

B BoaStats.R **96**

Chapter 1

Introduction

In this thesis, I present an acoustic study of Boa-Leboale, a Bantu language spoken by around 95,000 speakers in the Bas-Uélé Province in the Democratic Republic of the Congo. This study focuses on two vowels [ɪ ʊ] and the theoretical issues identified in previous literature surrounding the acoustic categorisation and identification of these vowels in many Niger-Congo and Nilo-Saharan languages, such as Boa-Leboale. I hope that this thesis contributes both to the description and documentation of the Boa languages and as well to the broader debate around the acoustic categorisation and identification of the vowels [ɪ ʊ] and [e o].

1.1 Topic background

Many Niger-Congo and Nilo-Saharan languages have vowel systems revolving around an opposition between two groups of vowels, usually labelled [+ATR] and [–ATR] by phonologists, where ATR stands for 'advanced tongue root'. The [+ATR] group will typically include /i u/ and sometimes /e o ə/, while the [–ATR] group will include /ɛ ɔ a/ and sometimes /ɪ ʊ/. The presence or absence of certain vowels gives rise to a number of commonly occurring vowel systems, like the 9-vowel system /i ɪ e ɛ a ɔ o ʊ u/, or the 7-vowel systems /i ɪ ɛ a ɔ ʊ u/ and /i e ɛ a ɔ o u/.

However, it is well-established in existing literature that it is very difficult to distinguish what I call the 'Level 2 vowels' – the [+ATR] vowels /e o/ and the [–ATR] vowels /ɪ ʊ/ – using traditional acoustic methods (see §2.3 and Casali (2008: 508–511)). There has been some success looking at indirect acoustic correlates of ATR, such as formant bandwidth and spectral flatness, but this has typically only been shown to distinguish the two sets of vowels in languages that have both, not identify if a language has one set of vowels or another. This is a challenge for descriptive linguists working on languages with 7 vowels: how can you tell if a 7-vowel language (like Boa-Leboale) has underlying [+ATR] /e o/ or [–ATR] /ɪ ʊ/?

In many languages, the easiest way to identify the vowels is through how they behave phonologically. Most languages with such vowel systems involve some kind of vowel harmony for ATR, so [+ATR] vowels tend to appear more with other [+ATR] vowels, and the same for [–ATR] vowels, both statically within roots and dynamically when affixes are added to a root. Boa-Leboale, however, only has a partial vowel harmony system. There do seem to be some restrictions on vowel co-occurence within roots, but the restrictions do not point towards any particular analysis in terms of ATR. Similarly, there are a number of vowel alternations induced by affixation, but they also defy a simple analysis in terms of ATR; rather, most of the alternations in Boa-Leboale involve a combination of [ATR] and [high] assimilation or dissimilation. As such, the phonology of Boa-Leboale means that, though previous work on the language has identified the underlying vowels as /ɪ ʊ/, it can just as easily be analysed as having underlying /e o/.

1.2 Research questions

The goal of this study is therefore to investigate whether acoustic methods can be used to determine the nature of the Boa-Leboale vowel system. This consists of two parts, each tied to a possible mistranscription or misanalysis of the vowel system.

Firstly, I want to see if I can use acoustic methods to determine whether the vowels described as /ɪ ʊ/ in Boa-Leboale are actually one set of vowels, or if there is any acoustic evidence that there might in fact be 9 vowels in the language (either underlyingly or on the surface). This is related to the fact that a number of languages with 9 vowels were mistakenly understood to have 7 vowels when descriptive linguists could not notice the difference between [e o] and [ɪ ʊ].

Secondly, I want to see if I can use acoustic methods to identify whether a given one of these ambiguous vowels is [+ATR] or [−ATR] based on its measurements relative to other vowels in the language. This is related to the potential mistranscription of [ɪ ʊ] as [e o] and vice versa.

Therefore the research questions for this thesis are presented in (1), where Level 2 vowels is a term used to refer to the ambiguous vowels which could be [+ATR] [e o] or [−ATR] [ɪ ʊ].

(1) I. Do the Level 2 vowels in Boa-Leboale include just one pair of vowels or two?
II. If Boa-Leboale has one pair of Level 2 vowels, are they [+ATR] /e o/ or [−ATR] /ɪ ʊ/? If it has both pairs of Level 2 vowels, which pair appears in which contexts?

I investigate these research questions using a number of hypotheses primarily based on the results of Starwalt (2008), a large scale study into 6 acoustic correlates of ATR in eleven 7- and 9-vowel languages. To test these hypotheses, I wrote a script for the speech analysis software Praat, which I ran on a number of annotated recordings provided by my supervisor Gerrit de Wit. This script took the necessary acoustic measurements for each of the annotated vowels and prepared them for statistical analysis, which was performed in R. Both the Praat and R scripts used in my analysis are given as appendices to this thesis.

1.3 Thesis structure

This thesis is composed of 4 main content chapters, as well as this introduction and a final conclusion.

Chapter 2 introduces the background to the theoretical side of the thesis, particularly in terms of defining what ATR is, how it manifests in articulation, acoustics and phonology, and expanding on previous work addressing the problem of distinguishing between the Level 2 vowels.

Chapter 3 describes the phonology of Boa-Leboale, as described in previous work (e.g. Motingea 2005; de Wit 2022b). This includes the 7-vowel inventory, the co-occurrence restrictions in noun roots and a number of relevant vowel alternations. It also briefly discusses similar alternations in related languages to show how such alternations can be analysed in a number of ways.

Chapter 4 outlines the methodology of the main acoustic study; how the data was collected and annotated (§4.1), how the acoustic measurements were taken (including a detailed description of how the Praat script functions) (§4.2), and how the statistical analysis was performed in R (§4.3).

Chapter 5 presents the results of the acoustic study and how they answer, or fail to answer, the research questions put forward. The results chapter, along with the conclusions, also discusses a number of promising directions for future research.

Chapter 2

Literature background

Many languages of the Niger-Congo and Nilo-Saharan phyla have vowel systems characterised by the opposition and alternation of two sets of vowels: one including the vowels /i u e o/ and the other /ɪ ʊ ɛ ɔ/. These two sets are commonly known as [+ATR] and [–ATR] vowels respectively, where 'ATR' is an initialism for 'advanced tongue root', one of the primary articulatory ways that these vowels are distinguished from one another.

These languages typically have symmetrical vowel systems, where vowels can be paired up into [+ATR] ~ [–ATR] pairs like /i ~ ɪ/. Depending on which vowels are unpaired, we get different vowel systems. If there are no unpaired vowels, then we get a 10-vowel system such as /i ɪ u ʊ e ɛ o ɔ ə a/. If only /a/ is unpaired, we get a 9-vowel system: /i ɪ u ʊ e ɛ o ɔ a/. There are then two common 7-vowel systems: one where the mid vowels are unpaired: /i ɪ u ʊ ɛ ɔ a/, and one where the high vowels are unpaired: /i u e ɛ o ɔ a/. In this section, I will discuss why linguists can have difficulties distinguishing between these two types of 7-vowel system, and outline various ways they are predicted to be different from one another.

2.1 What is ATR?

Prior to Ladefoged's (1964) landmark phonetic study of various West African languages, [+ATR] and [–ATR] vowels were described as 'tense'/'lax' (borrowing terminology from European languages) or in terms of relative tongue body height (as countered by Stewart (1967a,b) and Pike (1967)). However, Ladefoged's study showed through X-ray radiography that the primary articulatory parameter that distinguished these two sets of vowels in Igbo was the retraction of the tongue body, leading to a narrower opening in the throat. Pike (1967) and Stewart (1967a,b) then labelled this contrast in terms of 'tongue root position'. Painter (1973) followed this up with another radiographic study, this time on Twi vowels, which also found that the primary articulatory difference was one of 'pharynx width'. Thus, in the late 1960s and early 1970s, the relevant feature came to be known as 'advanced tongue root' (hence the initialism 'ATR'), assuming that the movement of the tongue root was primary to distinguishing the two sets of vowels.

To this day, these vowel sets are still referred to as [+ATR] and [–ATR], though tongue root position is now rarely thought of as the only distinction between these vowels. The main other articulatory differences between [+ATR] and [–ATR] vowels are given in (2), per Casali (2008).

(2)
   a. [+ATR] vowels are often articulated with higher tongue bodies than their [–ATR] equivalents. Sometimes height is even the primary difference, e.g. in Ateso (Lindau & Ladefoged 1986).
   b. [+ATR] and [–ATR] vowels often have differences in laryngeal settings. [+ATR] vowels are 'unconstricted', sounding 'hollow', 'breathy' or 'resonant', while [–ATR] vowels are 'constricted', sounding 'tight', 'choked' or 'creaky'. Edmondson & Esling (2006) found that [–ATR] vowels in Kabiye showed aryepiglotto-epiglottal constriction, epiglotto-pharyngeal constriction and pharyngeal narrowing, while [+ATR] vowels were similar to a faucalised ('yawning') voice in terms of their lack of pharyngeal constriction.
   c. [+ATR] vowels can be articulated with a lower larynx than their [–ATR] equivalents, e.g. in Akan (Lindau 1979) and Kabiye (Edmondson & Esling 2006).

There is a reasonable amount of variation in how ATR is realised between languages and between speakers, but most cases involve some effect from tongue root position, tongue height and pharyngo-laryngeal settings.

However, articulary gestures are only half the picture when it comes to the phonetics of a particular feature, the other half being acoustic correlates and perceptual cues which, like articulation, can be varied. One of the primary acoustic differences between [+ATR] and [–ATR] vowels is the frequency of the first formant (F1) – this has been found time and again in instrumental studies on ATR vowel systems (e.g. Lindau 1978; Lindau & Ladefoged 1986; Hess 1992; Fulop et al. 1998; Casali 2002; Guion et al. 2004; Przezdziecki 2005; Starwalt 2008; Olejarczuk et al. 2019). This correlation is because lowering of F1 can be achieved through a number of articulatory mechanisms involved in ATR: expansion of the pharynx through advancement of the tongue root, raising of the tongue body and lowering of the larynx.

Other proposed acoustic correlates of ATR tend to vary by speaker and language, though they can also be linked to aspects of articulation already discussed. Raising the tongue body, along with reducing F1, also affects the frequency of the second formant (F2), while changes in laryngeal settings can be heard as 'voice quality', which correlates with a number of acoustic measurements, most notably, first formant bandwidth (B1), measures of so-called 'spectral tilt' or 'spectral flatness', and measures of periodicity. Spectral tilt refers to the extent to which amplitude decreases for higher frequencies – breathy sounds (like many [+ATR] vowels) typically have a steeper drop in acoustic intensity than modal or creaky voiced sounds (Gordon & Ladefoged 2001: 398). Such measures include, among many others, the normalised difference between the amplitudes of the first two formants (A1*–A2*), the normalised difference between the amplitudes of the first two harmonics (H1*–H2*) and centre of gravity. Measures of periodicity instead use the idea that different laryngeal settings will affect the regularity of the sound signal – breathy sounds are less periodic than modal voiced sounds in their higher frequencies since they have increased high frequency spectral noise (Hillenbrand et al. 1994); such measures include cepstral peak prominence (CPP) and harmonics-to-noise ratio (HNR) (both used by Olejarczuk et al. (2019)).

2.2 ATR vowel systems

[+ATR] and [−ATR] vowels typically fall into so-called ‘harmonic pairs’ – that is, pairs of vowels which are identical for other features (backness, height, rounding). This harmonic pairing is less common with low vowels: many languages have the low [−ATR] vowel /a/, but not its [+ATR] counterpart, like Liko, whose vowel system is shown in (3).

(3) The 9-vowel system of Liko (de Wit 2015)

	[+ATR]		[−ATR]	
High	i	u	ɪ	ʊ
Mid	e	o	ɛ	ɔ
Low			a	

In Liko, all eight high and mid vowels fall into harmonic pairs: /i ~ ɪ/, /u ~ ʊ/, /e ~ ɛ/ and /o ~ ɔ/. These vowels alternate within their harmonic pairs in different phonological environments, as can be seen with the noun class prefixes li ~ lɪ and mu ~ mʊ in (4).

(4)	a.	li-kúbú	‘5-umbilical cord, navel’
		mu-pumí	‘3-door’
	b.	lɪ-sɪ́sɪ́	‘5-oil palm tree’
		mʊ-tɪ́wɪ	‘3-advice’

The examples in (4) demonstrate a widespread phenomenon in languages with ATR vowel systems called vowel harmony, where the vowels in a particular word tend to all have the same value for ATR: they are all [+ATR] (as in 4a) or all [−ATR] (as in 4b). Here we say that the vowels in each word ‘agree’ or ‘harmonise’ for the feature ATR.

In many languages, unlike Liko, vowels other than /a/ are also left unpaired. This gives two different possible 7-vowel systems: one where the high vowels are unpaired (as in 5a), e.g. in Yoruba (Allen et al. 2013), and one where the mid vowels are unpaired (as in 5b), e.g. in Kinande (Gick et al. 2006) and Ethiopian Komo (Olejarczuk et al. 2019).

(5) a. Type A

	[+ATR]		[−ATR]	
High	i	u		
Mid	e	o	ɛ	ɔ
Low			a	

b. Type B

	[+ATR]		[−ATR]	
High	i	u	ɪ	ʊ
Mid			ɛ	ɔ
Low			a	

Alternatively, ATR and height can be considered together, with the vowels occupying different ‘levels’, where higher level vowels have greater height and are [+ATR]. In this case, the two types of vowel system would look very similar to one another, as can be seen in (6).

(6) a. Type A

Level 1	i	u
Level 2	e	o
Level 3	ɛ	ɔ
Level 4	a	

b. Type B

Level 1	i	u
Level 2	ɪ	ʊ
Level 3	ɛ	ɔ
Level 4	a	

The difference between the two types of 7-vowel system can then be clearly seen as the difference between the Level 2 vowels, whether they are [+high, –ATR] or [–high, +ATR]. Following this approach, Starwalt (2008) refers to different vowel systems by how many levels there are (for 7 vowels, this is 4Ht) and whether the Level 2 vowels are high or mid vowels (H or M). As such, in her system, Type A would be a 4Ht(M) system and Type B a 4Ht(H) system. Casali (2008), however, calls them /1IU/ and /2IU/ 7-vowel systems respectively, referring to the number of front-back pairs of high vowels present in each system. I will adopt Casali's conventions in this thesis, but use the phrase 'Level 2 vowels' to refer to vowels between /i u/ and /ɛ ɔ/, which could be /ɪ ʊ/ or /e o/.

Like with the 9-vowel systems, 7-vowel languages often have vowel harmony for the feature ATR. However, since fewer vowels are paired underlyingly, these languages often have allophonic, non-phonemic vowels which surface in harmonic contexts. For example, Komo only has 7 phonemic vowels /i ɪ u ʊ ɛ ɔ a/ (a /2IU/ system), but the vowels /ɛ ɔ a/ surface as [e o ə[1]] in [+ATR] harmonic contexts (Olejarczuk et al. 2019). This can be seen in (7), with the alternating vowels in bold.

(7)	a.	[k**ɛ́**ʃ-ʊ́]	'thresh-DD1'
		[d**ɔ̀**t'-ʊ́]	'ask-DD1'
		[h**á**m-ʊ́]	'yawn-DD1'
	b.	[k**é**ʃ-úk]	'thresh-DD2'
		[d**ò**t'-úk]	'ask-DD2'
		[h**ə́**m-úk]	'yawn-DD2'

The non-high [+ATR] vowels [e o ə] here are only found as harmonic allophones of /ɛ ɔ a/ in Komo, which is why we still call this a 7-vowel system. We also find an equivalent allophony in some /1IU/ 7-vowel systems like Ijẹṣa and Ekiti Yoruba, with an underlying 7-vowel system /i u e ɛ o ɔ a/ but surface [ɪ ʊ] as [–ATR] allophones of /i u/ (Orie 2003). This compares to Standard Yoruba, where only the mid vowels obey ATR harmony (Elesin 2018: 251–252). For roots with mid vowels, as in (8a), the nominalising prefix must harmonise for ATR, but as (8b) shows, the same is not true for the high vowel /i/, which can appear before a [+ATR] or [–ATR] root. In Standard Yoruba, then, there are seven vowels underlyingly *and* on the surface.

One recurring problem in fieldwork on these 7-vowel languages has been identifying which type of system they have. This is primarily because Level 2 vowels are not consistently different in terms of F1, the primary acoustic correlate of ATR discussed so far, and there are also many other respects in which these vowels are similar to one another. How, then, do linguists determine if a given 7-vowel

---

[1] Here the symbol [ə] is used for the [+ATR] equivalent of the low vowel /a/. This sound is transcribed in a variety of ways by different linguists, including /ʌ/, /ɐ/, /æ/ or just using the IPA advanced tongue root diacritic /a̘/.

(8) a. i. **ɛ̀-** tɔ́
PREF- right
‘right’

ii. **ɔ̀-** rɛ́
PREF- to.befriend
‘friendship’

iii. **è-** to
PREF- to.plan
‘plan’

iv. **o-** gbó
PREF- old
‘old age’

b. i. **ì-** fɛ́
PREF- to.love
‘love’

ii. **ì-** pè
PREF- to.call
‘call’

language is /1IU/ or /2IU/?

2.3 Distinguishing /1IU/ and /2IU/ 7-vowel systems

To answer this question, I will discuss the ways in which the Level 2 vowels are similar to one another, and how they might be distinguished, in their articulation, acoustics (including impressionistic judgements) and also in their phonological behaviour in vowel harmony. Since /ɪ ʊ/ and /e o/ do not form a harmonic pair (they differ in terms of height as well as ATR), the same correlates discussed in Section 2.1 cannot always be applied. As such, I will survey some of the evidence we have for how these vowels are distinct from one another, for now excluding the evidence presented by Starwalt’s (2008) landmark study on 11 different languages (see §2.4).

2.3.1 Articulatory diagnostics

Most of Section 2.1 discusses the articulatory correlates of ATR, and for good reason: ATR, like many other phonological features, is primarily defined in articulatory terms (e.g. tongue root advancement, pharyngeal expansion). Therefore, any discussion of how to distinguish these vowels should start with discussing how they might differ in how they are articulated.

Starting with the X-ray radiographic studies from the 1960s-70s, Ladefoged (1964: 38) gives tracings from his study on Igbo which, when overlaid, show that the tongue root of [o] lies between the advancement of [u] and [ɔ], but is clearly more advanced than [ʊ], which is comparable to [ɔ].[2] Painter (1973) found that the low and back vowels were much more retracted than the front vowels, regardless of ATR value. As such, the [+ATR] back vowels were sometimes equivalent to [–ATR] front vowels, and it was not possible to separate [+ATR] and [–ATR] vowels purely in terms of raw tongue root position. Taking the front and back vowels separately, however, he concluded that [e] and [o] had wider pharynxes and more advanced and raised tongue positions than [ɪ] and [ʊ], as might be expected. Similarly, though she does not explicitly discuss it, Lindau (1979) includes tracings of the vowels of each of the speakers in her study of Akan. From these tracings, we can see that there is a good deal of variation between speakers in the relative tongue root positions and body heights

---

[2] Igbo does not have the phoneme */ɛ/, so it is harder to compare the front vowels. From what I could tell, the tongue root advancements of [ɪ] and [e] were comparable to one another – somewhere between [i] and [a].

of /ɪ ʊ/ and /e o/. For the front vowels, [e] consistently has a more advanced tongue root than [ɪ], while there is little consistency in terms of tongue body height (two speakers have very similar vowel heights, one [ɪ] is higher than [e], one the inverse). For the back vowels, [ʊ] consistently has a higher tongue body height than [o] and there is little consistency in tongue root position, much like with [e] and [ɪ]. As such, we can see that even tongue root position is not always sufficient to distinguish these vowels for every speaker.

In a more recent ultrasound study, Gick et al. (2006: 14–16) investigated the vowels of Kinande, a 7-vowel /2IU/ language with surface [e o]. They found that [ɪ] and [ʊ] were both categorically distinct from the [+ATR] high vowels [i u] in terms of tongue root retraction, but that this was not true for the surface [+ATR] mid vowels [e o], which both overlapped with [i u] in their tongue root retraction. They did find that [e o] each had closer average tongue root advancement to [i u] than [ɛ ɔ], and that the inverse was true for [ɪ ʊ], which were closer to [ɛ ɔ]. For the back vowels, however, there was no significant difference in tongue root position between [o] and [ʊ]. As such, while the front vowels could easily be distinguished in terms of tongue root position, the same was not true of the back vowels. This disparity between front and back vowels was also found by Hudu (2010) in his study of Dagbani, another /2IU/ language. All five participants pronounced both [e] and [i] with a more advanced tongue root than both [ɛ] and [ɪ], but only one had any significant difference between [o ʊ], and for them it was in the unexpected direction ([ʊ] had a more advanced tongue root than [o]).

These studies show a general pattern in the front vowels where [e] is articulated with a more advanced tongue root than [ɪ]. As such, we have a possible diagnostic criterion for an unknown front vowel between [i] and [ɛ] – if its tongue root is closer in advancement to [i] than [ɛ], then this it is more likely to be [e]. However, this criterion does not always work for every single study given, and it fails entirely for the back vowels, which often show very little difference in tongue root position between [o] and [ʊ].

2.3.2 Acoustic diagnostics

One other problem with using a diagnostic based on articulation is that articulatory data is much harder for fieldworkers to have access to. The studies described above are either X-ray radiographic studies – which are no longer practised to avoid adverse health risks – and ultrasound studies, which are becoming more achievable in a fieldwork setting, especially with the experimental set-up used by Gick et al. (2006), and are relatively unobtrusive. However, even ultrasound studies require specialist equipment that must be carried to the fieldwork location and require the speaker to stay very still. As such, there is still a barrier to getting good articulatory data from every language, as can be seen by the small number of articulatory studies conducted so far.

While not every fieldworker has access to ultrasound imaging equipment, they will all typically use an audio recorder that can capture acoustic data. If an acoustic analysis can distinguish between these vowels, then this is a much more achievable and cost-effective alternative to requiring an articulation study for each language.

2.3.2.1 Formants (F1, F2, B1)

As already mentioned, the main acoustic distinction between [+ATR] and [–ATR] harmonic pairs is F1, the frequency of the first resonating frequency of the vowel (formant), since expansion of the pharyngeal cavity has the effect of lowering F1. However, F1 is also the primary acoustic correlate of vowel height. As such, when we have a cross-height pair such as [e] and [ɪ], we have a conflict of expectations: based on just vowel height (ignoring ATR), [ɪ] would have lower F1 than [e], but based on just ATR (ignoring height), [e] would have lower F1 than [ɪ]. Various studies on the acoustics of ATR vowel systems have found differing results: Guion et al. (2004) found that F1 was higher for [e o] for some speakers of Maa than for [ɪ ʊ], while Lindau (1978) found the inverse case for Akan back vowels (F1 was higher for [ʊ] than [o]). The most common finding across these studies, however, is no significant difference between F1 values for the Level 2 vowels, as found by Casali (2002) in his study on Nawuri. This finding is backed up by Lindau (1978) and Hess (1992) for Akan front vowels, Jacobson (1978) for Dholuo front vowels, Fulop et al. (1998) for Degema, Gick et al. (2006) for Kinande back vowels, among many others. Hess (1992) also notes that even when the formant values for the Level 2 vowels were the same in Kwawu Akan, they were still perceptually distinct, meaning there must be some information in the acoustic signal other than the value of F1 which the listener can use to identify the vowels.

Similar results are found for F2, with some studies finding a difference between these vowels in terms of their peripherality. Ladefoged (1964) found that Igbo [ɪ ʊ] were more peripheral than [e o]; [ɪ] had higher F2 than [e] and [ʊ] had lower F2 than [o]. Remijsen et al. (2011) and Anderson (2007) found similar results for Shilluk and Foodo respectively. (Jacobson (1980) found the same for Shilluk front, but not back, vowels.) Guion et al. (2004) found only one Maa speaker in their study distinguished [e] and [ɪ] by F2, but no speakers distinguished [o] and [ʊ] with F2. Many other acoustic studies found no meaningful difference between either pair for F2 (e.g. Lindau 1978; Hess 1992; Casali 2002; Przezdziecki 2005).

While the F1 results from different studies contradict one another, at least the studies for F2 found either nothing or agreed that [ɪ ʊ] were more peripheral than [e o]. I could find no study where they found the inverse (that [ɪ ʊ] were significantly less peripheral than [e o]).[3] There is, however, a third formant measure which has been found to distinguish the Level 2 vowels: the bandwidth of the first formant. Hess (1992) found that [–ATR] vowels in Akan had significantly wider first formants than [+ATR] vowels, and that this held equally for same-height harmonic pairs as well as the cross-height pairs discussed here. Przezdziecki (2005), however, did not consistently find this difference between speakers of different Yoruba dialects.

2.3.2.2 Spectral tilt and periodicity (A1*–A2*, CoG, CPP)

Measures of spectral tilt – generally speaking, the idea that [+ATR] sounds will have more energy in the higher frequencies than [–ATR] sounds – are numerous and various in their implementations and rates of success. Periodicity is another feature of the acoustic signal that is thought to correlate with differences in voice quality, and has similarly been measured in a number of ways.

---

[3] Note that this is not the case for same-height harmonic pairs; for pairs like /i ɪ/, Fulop et al. (1998) did find the inverse result in Degema (less peripheral F2 values for [–ATR] vowels).

There are many spectral tilt measures which I will not discuss in this thesis.[4] Instead, I will focus on the two measures investigated by Starwalt (2008), which my study is modeled on. The first measure is A1*–A2* – the normalised difference in amplitude between the first and second formants – which was found to correlate at least somewhat with ATR for Degema by Fulop et al. (1998) and for Maa by Guion et al. (2004). Furthermore, Guion et al. (2004) found that two Maa speakers had a significant difference for normalised A1–A2 between [e] and [ɪ], though this difference was not found for the other speakers, nor for the back vowels. The second spectral tilt measure here is centre of gravity (CoG), which measures the average frequency of acoustic energy. Anderson (2007) found that most Foodo speakers distinguished significantly between [e o] and [ɪ ʊ], even when F1 values overlapped. Beyond the work by Colleen Starwalt (Anderson 2007; Starwalt 2008), I have not found much research into CoG as a measure of spectral tilt. I will discuss her further findings on both A1*–A2* and CoG in a later section (§2.4.2.2).

Like with spectral tilt, there are many measures of periodicity that I will not discuss here (including RPK, peak-to-average ratio (as in Hillenbrand et al. 1994) and harmonics-to-noise ratio (as in Olejarczuk et al. 2019)). The one measure I will discuss is Cepstral Peak Prominence (CPP), which Hillenbrand et al. (1994) found to be correlated with breathy voice and Olejarczuk et al. (2019) found to be a significant correlate of ATR within harmony pairs in Komo, with [+ATR] vowels having lower periodicity than [–ATR] vowels, in line with the alignment of [+ATR] vowels with breathiness. Whether or not this can be used as a reliable way to distinguish the Level 2 vowels from one another remains to be investigated.

(9) Summary of acoustic diagnostics of Level 2 vowels:

a. **F1**: There is no consistent difference in F1 between [e o] and [ɪ ʊ]; some studies find significant differences but in contradictory directions.

b. **F2**: In some studies, F2 for [ɪ ʊ] is more peripheral than [e o], but in many cases there is no significant effect.

c. **B1**: Hess (1992) found that B1 was significantly higher for [ɪ ʊ] than for [e o] in Akan, but Przezdziecki (2005) did not find the same effect in Yoruba.

d. **A1*–A2***: Two speakers of Maa were found to have significantly higher A1*–A2* for [ɪ ʊ] than for [e o] (Guion et al. 2004), but there are many cases where this difference is not significant (as in the other Maa speakers) or was not investigated (as with the Degema study by Fulop et al. (1998)).

e. **CoG**: One study (Anderson 2007) found that the Level 2 vowels were distinguished significantly by centre of gravity for three out of four speakers.

f. **CPP**: One study (Olejarczuk et al. 2019) found that CPP significantly correlated with ATR for harmonic pairs, but no study has been conducted on using CPP to distinguish the Level 2 vowels.

---

[4] For example, Gordon & Ladefoged (2001) measure H2–f0 and A1–f0; Hillenbrand et al. (1994) have BRI, H/L and H1/H2; Kingston et al. (1997) measure H1–H4; Olejarczuk et al. (2019) measure H1*–H2*, H2*–H4*, H1*–A1*, H1*–A2*, H1*–A3*, A1*–A3* and A2*–A3*, in addition to A1*–A2*, which I do discuss here.

2.3.3 Phonological diagnostics

Not much space here will be dedicated to the phonological behaviour of /e o/ compared to /ɪ ʊ/, since this is primarily a phonetic study. However, taking even just Standard Yoruba example (8) from earlier, it is clear that phonological behaviour, particularly in vowel harmony, can be a strong indication of whether a language has a /1IU/ or /2IU/ system. In the Yoruba example, mid vowel prefixes ɛ-, ɔ-, e- and o- need to agree in ATR with the vowel of the noun root, while the prefix i- could appear before both [+ATR] and [–ATR] roots. This strongly indicates that /i/ is harmonically unpaired in Yoruba, otherwise we might expect to find the presence of [–ATR] [ɪ] in a [–ATR] environment. Since we do not, this gives a /1IU/ vowel system with the vowels /i u e o ɛ ɔ a/.

Similarly, in the Komo example (7), the presence of a [+ATR] vowel in the suffix leads to allophonic changes in the root verb, cf. [d**ɔ̀**t'-ʊ́] 'ask-DD1' and [d**ò**t'-úk] 'ask-DD2'. In this case, it is quite easy to see from this phonological alternation that the vowel of the suffix -ʊ́ 'DD1' is a [–ATR] high vowel, while the [o] in [dòt'-úk] is a [+ATR] mid vowel, simply because they alternate with [u] and [ɔ] respectively. The differing behaviour of roots before suffixes with the vowels /ʊ/ and /u/ strongly indicates that Komo does distinguish [+ATR] and [–ATR] high vowels, and is thus a /2IU/ system (albeit with derived surface [e o ə] in [+ATR] contexts).

However, in languages where the vowel harmony system is more restricted, we might not find these kinds of diagnostic alternations. For example, in Boa-Leboale (a 7-vowel system), the number of vowel alternations within verbs is rather limited. One alternation is that the negative future suffix -ɪ is realised as -i after a verb root with /i u/ in it, as in example (10).[5]

(10) a. i. tí- dɔp **-ɪ** =ndɛ́
1S.NEG- frapper -N.F =FUT
'je ne frapperai pas'

ii. tí- mʊnd **-ɪ** =ndɛ́
1S.NEG- watch -N.F =FUT
'je ne regarderai pas'

b. i. tí- sumb **-i** =ndɛ́
1S.NEG- brûler -N.F =FUT
'je ne brûlerai pas'

ii. tí- kil **-i** =ndɛ́
1S.NEG- arriver -N.F =FUT
'je n'arriverai pas'

This example can be easily analysed in terms of spreading of [+ATR], in which case Boa-Leboale has a /2IU/ system. However, we could easily replace all instances of /ɪ ʊ/ with /e o/ and still produce a functional analysis of the system in terms of spreading of [+high] instead. In this analysis, the examples in (10a) would have only [–high] vowels, while the examples in (10b) would only have [+high] vowels. As such, there is an ambiguity between ATR-spreading and height-spreading systems that cannot always entirely be distinguished phonologically.

---

[5] Boa-Leboale glosses throughout this thesis will be given with the French glosses from their original texts to avoid a mistranslation into English.

Beyond looking directly at active harmonic patterns, there have been several suggestions about general principles of harmony, such as whether there is ATR dominance or not and the behaviour of /a/ in harmonic contexts, which could further distinguish /1IU/ and /2IU/ systems through phonology (see Casali 2003, 2008, 2016; Rose 2018).

I will return to this question of how to distinguish the two types of 7-vowel systems at the end of this chapter (§2.4.3), where I will discuss the various acoustic approaches to this question from Starwalt (2008) who, beyond being the model for my own study, specifically intended to use her study to propose possible ways to answer this question.

2.4 Starwalt (2008)

The previous discussion on acoustic differences between [e o] and [ɪ ʊ] (see §2.3.2) missed out one major work, Starwalt (2008) which looked at the acoustic differences between these vowels (and all other vowels) in 11 different Niger-Congo languages of varying vowel systems. Given the scope of her study, both in terms of languages and acoustic correlates considered, Starwalt (2008) naturally forms the basis for much of this thesis's main study's methodology.

2.4.1 Starwalt's methodology

Starwalt's study is a large-scale investigation into a variety of acoustic correlates of ATR in six Bantu languages (Dibole, Kinande, Londo, LuBwisi, Mbonge, Mbosi), three Kwa languages (Foodo, Ikposo, Tuwuli) and two Edekiri languages (Ekiti-Yoruba and Ifè). For each language, Starwalt created a dataset of recordings of at most five speakers (three men and two women) pronouncing words from a word-list. Each word-list included two words for each vowel and each word was repeated 10 times, for a total of 20 tokens per vowel per speaker per language. These vowel tokens were then analysed in Praat 4.3, a software package designed for speech signal processing and analysis (the most recent version of which is Boersma & Weenink (2021)), to take measurements of F1, F2, F3, B1, A1, A2 and CoG.[6] These measurements were then processed to give results for a number of acoustic correlates described above, namely F1, F2, B1, ΔB1, A1*–A2* and CoG.

Below I give an outline of how each of these acoustic measurements were taken and processed in Starwalt's original study. For further details on her methodology, see Starwalt (2008), while Chapter 4 details my own methodology and the ways in which it differs from Starwalt's.

2.4.1.1 F1, F2, F3

To measure the frequencies of the first three formants, Starwalt used the in-built formant measurement system in Praat to average formant values over a short interval in the centre of each vowel (5–10 glottal pulses). This can be seen in Figure 2.1. This short interval is part of the steady state of the vowel and was chosen to mitigate the effects of any formant transitions from the preceding or following consonant, and seems to have been selected by hand for each vowel under consideration.

---

[6] Starwalt also measured duration and fundamental frequency, but does not report these results since there were no clear significant correlations between either measurement and ATR. Since the results are not reported, I will not discuss them either, here or for my study.

For the most part, Starwalt used the default Praat settings for formant measurement: 5 formant tracks with a maximum frequency of 5000Hz for male speakers and 5500Hz for female speakers. Occasionally, she adjusted these settings on a case-by-case basis, when Praat's automatic formant tracking misidentified F1.

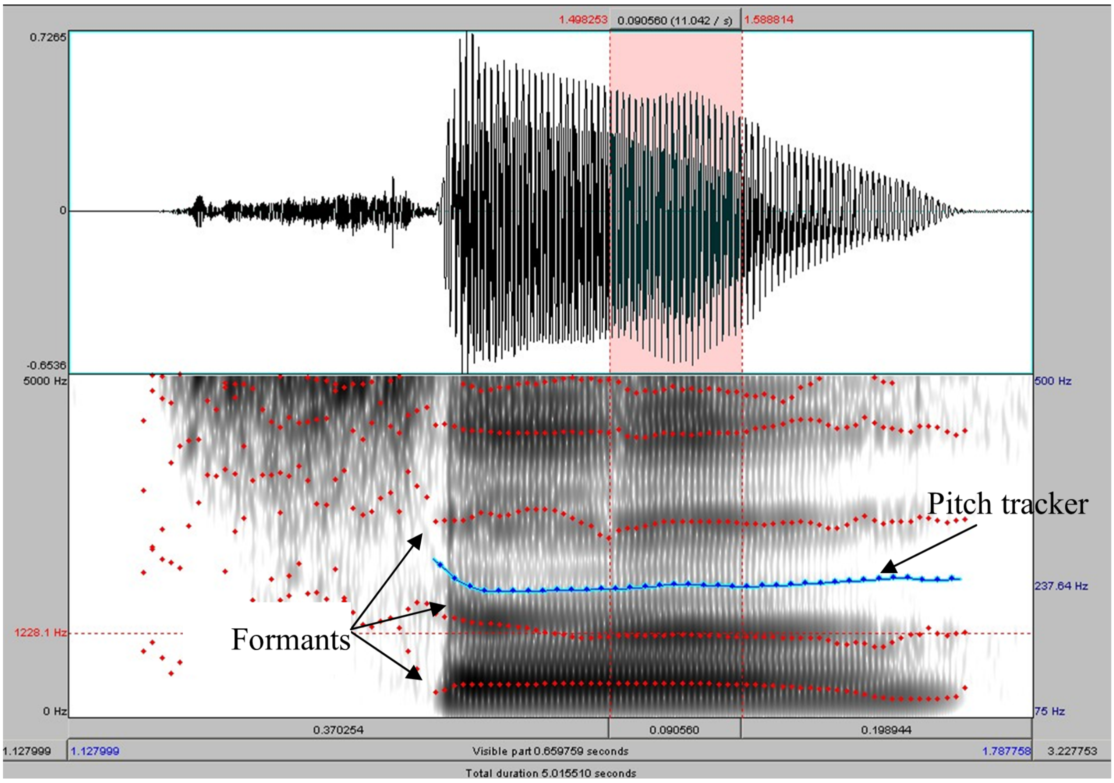


Figure 2.1: A sample measurement interval for the vowel [o] in Ifè so 'to attach', shown in the Praat graphical user interface's Sound window. Red dots show formant tracks, while the blue line/dots show pitch tracking/glottal pulses. Image reproduced from Starwalt (2008: 82).

Starwalt also took alternative measurements of F1, F2 and F3 at a single point within the steady state interval discussed here. These measurements were not directly used in her analyses, but are important for modelling values for B1, A1 and A2 as described in §2.4.1.2 and §2.4.1.3.

2.4.1.2 B1, ΔB1

Unlike with the formant frequency measurements, B1 was measured at a single point within the steady state of the vowel. Starwalt reports that the algorithm used was the LPC method, possibly using the in-built bandwidth measuring algorithm in Praat, but this is not explicitly stated.

Following Hess (1992), Starwalt compares these measured values for B1 with a model for B1 ($\hat{B}_1$) developed by Fant (1972: 47). Fant's formula is given in (11). Since B1 was measured at a single point in the vowel, the value F_1 used in this equation must also be the F1 value measured at that same point, to ensure that the model is as close to the sample data as possible.

(11) $$\hat{B}_1 = 15\left(\frac{500}{F_1}\right)^2 + 20\left(\frac{F_1}{500}\right)^{\frac{1}{2}} + 5\left(\frac{F_1}{500}\right)^2$$

Starwalt compares the measured values of B1 to the output of this formula ($\hat{B}_1$) by simply subtracting one from the other to give Delta B1 (ΔB1) – the displacement of measured B1 from its predicted value – as can be seen in (12).

(12) $$\Delta B_1 = B_1 - \hat{B}_1$$

This displacement is calculated on the basis of Hess (1992), who found that [–ATR] vowels generally had much higher measured values than those predicted by Fant's formula (ΔB1 $\approx$ 50Hz), while [+ATR] vowel bandwidths were similar or even lower than predicted values.

2.4.1.3 A1, A2, A1*–A2*

A1 and A2 are the amplitudes of the most prominent harmonic associated with the first two formants (F1 and F2) respectively. They can be measured by first analysing the vowel sound as a narrowband spectrogram, which is detailed in the frequency domain rather than the time domain, then taking a spectral slice at a single point during the steady state of the vowel – this would be the same point as the readings for bandwidth discussed before. Starwalt then measured on a case-by-case basis the amplitudes of the harmonic most closely associated with the peak of the formant band, presumably using the Praat graphical user interface's Spectrum window, as shown in Figure 2.2.

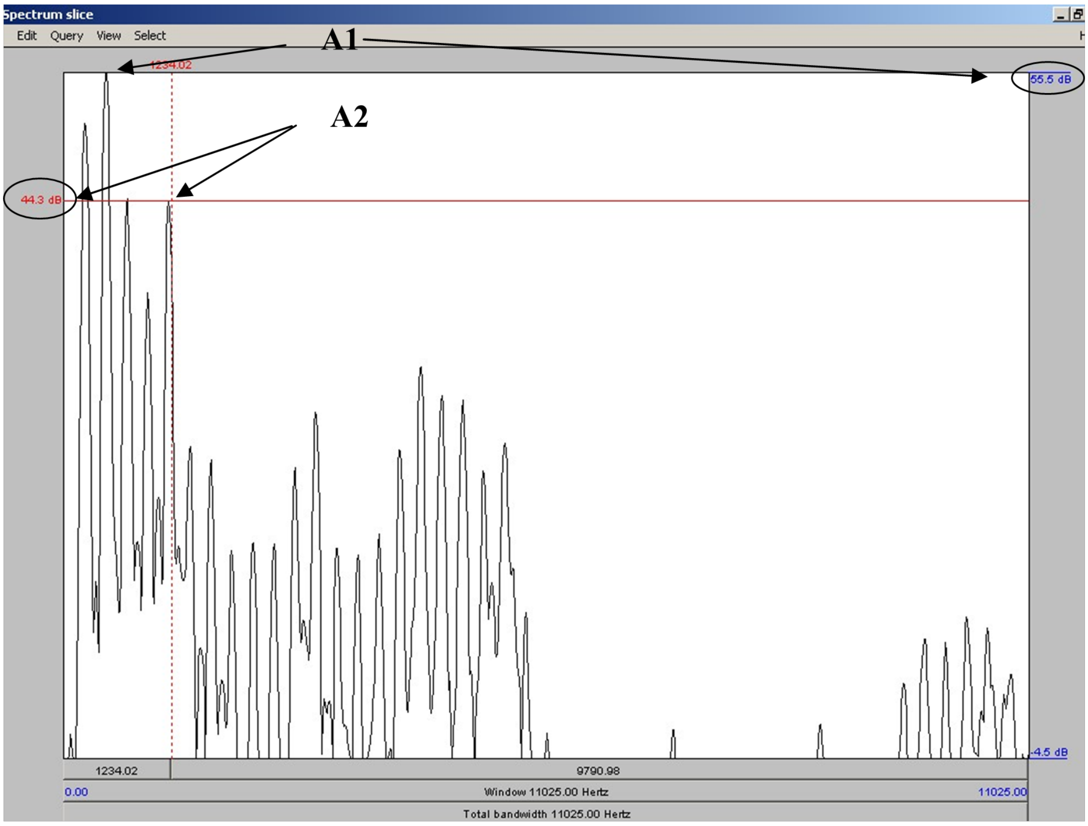


Figure 2.2: Measurement of A1 and A2 for the vowel [o] in Ifè `so` 'to attach', shown in the Praat graphical user interface's Spectrum window. A1 is measured as the amplitude of the 2nd harmonic (H2) and A2 is measured for the 5th harmonic (H5). Image reproduced from Starwalt (2008: 94).

The relevant measure for the discussion of distinguishing ATR vowels is not A1–A2, however, but a normalised value for A1–A2 taken by comparing this difference with a model developed by Fulop et al. (1998). This model is rather complex, taking into consideration the contributions to the spectrum from: (1) each of the first three formants individually ($\hat{A}_F(f)$), (2) the higher formants together ($\hat{A}_{upper}(f)$), and (3) a combined curve describing the losses from larynx pulses and lip radiation ($\hat{A}_{loss}(f)$). These three components together model the expected amplitude for a given frequency of the spectrum ($\hat{A}(f)$) up to around 3500Hz. The five curves and the final model are plotted in Figure 2.3, while the relevant equations are given in (13).

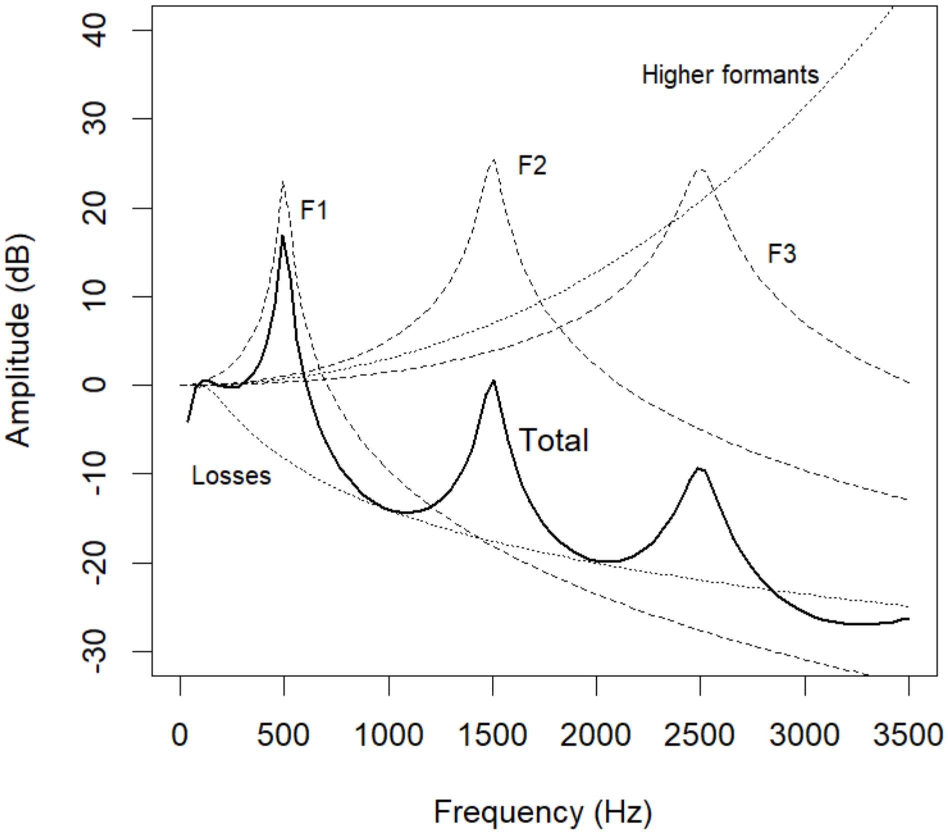


Figure 2.3: Plot of the amplitude model from Fulop et al. (1998) (solid line) used to calculate A1*–A2*, along with its component parts: F1, F2, F3 are shown with dashed lines, higher formants and losses from glottal pulses and lip radiation are shown with dotted lines.

(13) a. $$\hat{A}_F(f) = 20\log_{10}\left(\frac{F^2+\left(\frac{b}{2}\right)^2}{\sqrt{(f-F)^2+\left(\frac{b}{2}\right)^2}\sqrt{(f+F)^2+\left(\frac{b}{2}\right)^2}}\right)$$

b. $$\hat{A}_{upper}(f) = 0.72\left(\frac{f}{492}\right)^2 + 0.0033\left(\frac{f}{492}\right)^4$$

c. $$\hat{A}_{loss}(f) = g\left(-20\log_{10}\left(2\frac{\frac{f}{100}}{1+\left(\frac{f}{100}\right)^2}\right)\right)$$

d. $$\hat{A}(f) = \hat{A}_1(f) + \hat{A}_2(f) + \hat{A}_3(f) + \hat{A}_{upper}(f) + \hat{A}_{loss}(f)$$

In the equations, F is (observed) formant frequency, b is the expected bandwidth for that formant (assuming B1 = 30, B2 = 80, B3 = 150, following Fulop et al. (1998)), and g represents phonation type (in this case we are comparing against modal voice, so $g = -1$).[7]

Thus, to model A1 and A2, we just have to plug in the F1, F2 and F3 values (observed at that same point that the A1 and A2 measurements were taken) into equation (13d) to give $\hat{A}(f)$, the predicted amplitude at that frequency (which will be the peak of the formant). We can call these predicted values $\hat{A}_1$ and $\hat{A}_2$. Starwalt then calculated normalised A1–A2 by subtracting the predicted difference from the observed difference, as in (14).[8]

$$A_1^* - A_2^* = (A_1 - A_2) - (\hat{A}_1 - \hat{A}_2) \tag{14}$$

2.4.1.4 CoG

The final measure, centre of gravity, was measured automatically through an in-built Praat function with default settings (where the power $p = 2$).

2.4.2 Starwalt's results

The rest of this section will explore, and attempt to summarise, Starwalt's findings on the differences between the cross-height pairs [e o] and [ɪ ʊ]. For all measurements, she conducted ANOVA tests with post-hoc Tukey tests using the statistical software SPSS.

Of the 11 languages in the study, only 5 have vowel systems involving both [e o] and [ɪ ʊ] for direct comparison (Kinande, LuBwisi, Foodo, Ikposo and Ekiti-Yoruba). The remaining six languages all have 7-vowel /1IU/ vowel systems without allophonic [ɪ ʊ], and so they cannot be analysed for differences between the cross-height pairs.

2.4.2.1 Formant results (F1, F2, B1)

Like in previous studies, Starwalt found almost no differences in F1 between [e o] and [ɪ ʊ]. There were no significant F1 differences between these vowels in Kinande or Ekiti-Yoruba, while any cases of significance were greatly outnumbered by those without any significance in LuBwisi and Foodo. The only language in which there seemed to be a consistent significant difference across speakers was in Ikposo, where [e] had a significantly higher F1 than [ɪ] for all speakers. Even in this case, however, no effect was found for the back vowels.

F2, on the other hand, seems to be a better distinctor between the cross-height pairs across languages, particularly when distinguishing the back vowels. [o] had significantly higher F2 values than [ʊ] for all speakers in Kinande, LuBwisi and Foodo, while in Ekiti-Yoruba only one speaker did not have a significant difference. Even in Ikposo, 3 out of 5 speakers had significant F2 differences for both front and back vowels.

---

[7] Graphing the function for $\hat{A}_{loss}(f)$ as written in Fulop et al. (1998), Guion et al. (2004) and Starwalt (2008) gives the intended plot, but reflected along the x-axis (a rising pattern instead of the expected negative decay pattern). As such, while in all those works it is reported that $g = 1$, I will use the value –1 here, since I believe it fits the intent of the original model better. Alternatively, formula (13c) should have a coefficient of 20 instead of –20 (this is the option I chose for the Praat script in my study).

[8] Note that the notation Ax* refers to each value separately being normalised ($A_1^* = A_1 - \hat{A}_1$).

Starwalt's results suggest that F2 is less consistent at distinguishing between the front vowels. All Foodo speakers but one have significantly higher F2 for [ɪ] than [e], but the same is true only for 2 out of 4 Kinande speakers. In LuBwisi and Ekiti-Yoruba, almost all speakers have no significant F2 differences.

In general, the F1 and F2 results match up considerably well with the existing literature already discussed: F1 cannot reliably distinguish [e o] and [ɪ ʊ], while in some languages F2 is more peripheral for [–ATR] vowels. Starwalt's results also suggest that F2 is a better predictor of [ATR] class membership for back vowels than for front vowels.

There is considerably less literature looking at B1 for cross-height pairs since the original paper by Hess (1992), which found that Akan speakers significantly distinguished between [e o] and [ɪ ʊ] in terms of B1. However, Starwalt's results are far more mixed. On the one hand, in Foodo, all but one speaker had significantly higher B1 for [ɪ ʊ] than for [e o], corroborating Hess's findings for Akan. On the other hand, this difference was only significant half of the time in Ikposo and was negligably significant (for one or no speakers) for Kinande, LuBwisi back vowels and Ekiti-Yoruba.

2.4.2.2 Spectral tilt results (A1*–A2*, CoG)

Starting with A1*–A2*, Starwalt found a wide spectrum of different results, from a strong level of significance for all speakers in Foodo to barely any effect at all in Ekiti-Yoruba, with the other languages in the middle.

Almost all Foodo speakers consistently had significantly higher A1*–A2* values for [e o] than for [ɪ ʊ] (except for one speaker who had a significant difference in the opposite direction for the back vowels). LuBwisi also had a strong significant effect for most speakers: 4 out of 5 for front vowels and 3 out of 5 for back vowels. Ikposo had the same effect for front vowels (4 out of 5 speakers), but not for back vowels (where it was found for only one speaker).

On the less convincing end of the spectrum, only 1 out of 4 Kinande speakers had significantly higher A1*–A2* for [e] over [ɪ], and though 3 speakers had significant effects for the back vowels, one of these was in the opposite direction than expected. Finally, only one speaker of Ekiti-Yoruba had any significant effects for A1*–A2*, though it was at least in the predicted direction.[9] It is important to note that Starwalt's results, for the most part, are in the opposite direction to the results of Guion et al. (2004), who found that [ɪ ʊ] had higher values of A1*–A2* than [e o].

The only previous study on centre of gravity and these cross-height pairs is Starwalt's own previous study on Foodo (Anderson 2007), which is itself a precursor to Starwalt (2008). Indeed, Foodo is the most promising case for CoG distinguishing the cross-height pairs out of the five languages; for 3 out of 4 speakers, [ɪ] has significantly higher CoG than [e]. However, the same is not true for the back vowels, where just one speaker had significantly higher CoG for [o] than [ʊ] (the opposite direction to the front vowels).

The other four languages do not seem to suggest that CoG has much promise as a diagnostic for dis-

---

[9] The Ekiti-Yoruba data has particularly notable differences between the speakers for many measures. Speakers 1M-A and 2M-M have no significant differences for B1, A1*–A2* or CoG, while speaker 3F-F has significant differences in almost all those same cases. Whether this is due to random error or a gender-based difference remains unexplored.

tinguishing between [e o] and [ɪ ʊ]. In Ikposo, 3 out of 5 speakers had significantly higher CoG for [ɪ] than for [e] (and one of the other speakers had a significant difference in the opposite direction), and only 2 out of 5 speakers had the same significant effect for the back vowels. Similarly, in LuBwisi, 3 out of 5 speakers had significant effects for front vowels, but no speakers showed significant differences for back vowels. In both Kinande and Ekiti-Yoruba, any significant differences were limited to individual speakers, or speakers showed significant differences in contradictory directions to one another.

These results suggest that A1*–A2* differences vary a great deal between languages, but are sometimes quite indicative of whether a given vowel is [e o] or [ɪ ʊ], while centre of gravity is only rarely useful as a method of distinguishing between cross-height pairs (but tends to be better at doing so for front vowels than back vowels).

2.4.3 Starwalt on distinguishing 7-vowel systems

Towards the end of the dissertation, Starwalt (2008: 422–437) raises the question discussed so far in this section, of whether we can use acoustic data to reliably distinguish between /1IU/ and /2IU/ 7-vowel systems. She uses her data to informally explore a number of potential approaches on how these systems might differ acoustically; these come both from predictions in prior literature and her own post hoc observations, but are not formalised as hypotheses or predictions per se. These approaches are summarised in (15):

(15) Summary of potential ways to distinguish 7-vowel systems from Starwalt (2008)
  a. Comparing F1 with [i u]: /1IU/ systems tend not to have auditory confusion between adjacent vowel heights (i.e. between /i/ and /e/), while /2IU/ systems sometimes have auditory confusion between /i/ and /ɪ/. (from Casali 2003)
  b. Comparing ΔB1 means with [ɛ ɔ]:
    i. [e o] always have significantly lower ΔB1 means than [ɔ a], and also tend to be significantly lower than [ɛ].
    ii. [ɪ ʊ] tend not to have significantly different ΔB1 means to [ɛ] in some languages (in others, [ɪ ʊ] are more closely aligned with [i u]).
  c. Looking at the relative values of CoG and F1: While raw CoG means can be compared to investigate differences in [ATR], Starwalt suggests that these results would be better interpreted in terms of their displacement from F1 mean.

Regarding approach (15a), Starwalt found that there was a huge amount of variation in the difference between the F1 values for Level 1 and Level 2 vowels from language to language, as well as between front and back vowels, and that this variation was so great and so inconclusive that she claims that “F1 mean differentials for degree 1 [i.e. [i u]] and 2 vowels [i.e. [e o] or [ɪ ʊ]] do not necessarily help much in answering our question”.

Starwalt was more optimistic about the approaches involving ΔB1 in (15b), which were suggested based on the data of her study. However, as a diagnostic for 7-vowel system type, they are not particularly telling. As shown in Table 2.1, while it is true that the tendency for significant differences in

ΔB1 are more universal for [e o] cross-linguistically, still more than half show significant differences for ΔB1 for [ɪ] in all cases, and for [ʊ] before [ɔ] or [ɑ].

Table 2.1: Proportion of languages with significantly lower ΔB1 means for rows [e o ɪ ʊ] compared to columns [ɛ ɔ a] – grey cells indicate optimal point of comparison

	[ɛ]	[ɔ]	[a]
[e]	11/11 100%	11/11 100%	11/11 100%
[o]	9/11 82%	11/11 100%	11/11 100%
[ɪ]	3/5 60%	3/5 60%	4/5 80%
[ʊ]	1/5 20%	3/5 60%	3/5 60%

At best, for an individual language, we might be able to make a prediction based on a comparison of the back vowel in question ([o] or [ʊ]) and [ɛ] – the relevant cells are shaded grey in the table – since this comparison shows the greatest divergence between the cross-height pairs. Then if a language does have a significant lower ΔB1 mean for the unknown back vowel than [ɛ], it is more likely to be [o] (around 90% of cases with significant effects for pooled language data are [o]), while if it does not have a significant effect, it is marginally more likely to be [ʊ] (around 67% of cases without significant effects are [ʊ]). However, this requires making a prediction on the basis of a *lack* of significance, which does not alone have any statistical power (even if it could perhaps be indicative).

The final approach (15c) is mentioned, but not particularly expanded on in Starwalt (2008) specifically. Rather, she mentions this idea as a jumping point for a discussion about language- and speaker-specific articulatory realisations of ATR – particularly that individual speakers may differ in which articulatory mechanisms they use to produce the contrast between [+ATR] and [–ATR] vowels, and as such they may also have different acoustic profiles. As such, the difference between F1 and CoG also does not seem to be a particularly useful diagnostic for distinguishing the cross-height vowels.

Starwalt then goes on to conclude that it is possible that some speakers have neutralised the difference between such vowels, or at least that in a 7-vowel system, speakers might not make a distinction in terms of ATR – having a neutral pharynx volume and laryngeal settings – since there is already a concomitant change in height that can be used as a reliable cue instead.

2.5 Conclusions

In this chapter, I introduced the main linguistic concepts and debate involved in distinguishing the two types of 7-vowel system in the Niger-Congo and Nilo-Saharan languages (§2.3), particularly from an acoustic perspective. The final section of the chapter focused on one particularly broad study, Starwalt (2008), which forms the basis of this study, and detailed the methodology and results that were found.

In the remainder of this work, I will now focus specifically on my own study of the vowel system of Boa-Leboale. As such, I will introduce the Boa-Leboale language and its phonological system in Chapter 3, my study's methodology in Chapter 4 and its results in Chapter 5.

Chapter 3

Boa-Leboale

This chapter introduces Boa-Leboale, the language investigated by this thesis's main acoustic study. The information presented here about Boa-Leboale (and other Boa languages) is primarily derived from the work of Gerrit de Wit. In particular, the main sources are a paper presented at the 52nd Colloquium on African Languages and Linguistics (CALL) (de Wit 2022a) and the phonology chapter of an as-yet-unpublished book on Boa-Leboale (de Wit 2022b). In addition, I present information from the online version of Ethnologue (Eberhard et al. 2023), de Wit's work on the closely-related Boa-Yewu (de Wit 2020) and an earlier sketch of Boa-Leboale by André Motingea Mangulu (Motingea 2005).

3.1 Language introduction

Boa (also spelt Bwa) is a cluster of Northwest Bantu languages spoken in the Bas-Uélé Province of the Democratic Republic of the Congo. This thesis concerns Boa-Leboale, one of the major varieties of Boa, spoken by five chiefdoms in the Bambesa territory. De Wit (2022b) estimates on the basis of census figures for these chiefdoms that Boa-Leboale was spoken by around 95,000 speakers in 2021.

In the ISO 639-3 standard, Boa is given the identifier bww, while in the Guthrie classification system for Bantu languages, it is classified with the code C44.

The closest Boa variety to Boa-Leboale is Yewu, spoken in three chiefdoms in the Buta territory by around 40,000–45,000 people. The lexical similarity between Yewu and Leboale is 90%. The most closely related non-Boa languages include Pagibete, Kango, Ngelima and Liko (Eberhard et al. 2023; de Wit 2022b).

There have been a number of previous academic works on Boa. Védy (1904) is an early anthropological study of Boa people and society, much like Halkin (1910), which also gives an early sketch of Boa phonology and grammar. Several further linguistic works on Boa are cited on Glottolog, including a grammar sketch by A. Nkabuwakabili from 1986 and a dialectological study by B. Banyama from 2006, though these were not available to me for further verification. More recent works on Boa include the sources already mentioned: the grammar by Motingea (2005) and the upcoming book by de Wit.

3.2 Boa-Leboale vowel system

Boa-Leboale has a 7-vowel system, like those described in §2.2. Its phonological system makes it not entirely clear if Boa-Leboale is a /2IU/ or a /1IU/ system, however Motingea (2005) and de Wit (2022b) both report that the Level 2 vowels are /ɪ ʊ/, as shown in (16).

(16) The vowels in Boa-Leboale (Motingea 2005; de Wit 2022b)

Level 1	i	u
Level 2	ɪ	ʊ
Level 3	ɛ	ɔ
Level 4	a	

Though it is unclear whether the Level 2 vowels are indeed /ɪ ʊ/ or /e o/, I will follow previous authors in treating them as [–ATR] /ɪ ʊ/ when giving examples in Boa-Leboale, with the understanding that their precise status is unknown.

The main difficulty in ascertaining the status of the Level 2 vowels in Boa-Leboale is that there is no robust ATR harmony system, static or alternating. There are some static restrictions on which vowels cooccur within roots, shown in Table 3.1, but they cannot easily be explained in terms of ATR.

Table 3.1: Table of vowel combinations in noun roots. Rows represent the first vowels of the root, columns the second vowel. Attested combinations are marked with +, sparsely attested combinations are marked with (+) and unattested combinations are shaded and marked with *. (de Wit 2022b)

	/i/	/u/	/ɪ/	/ʊ/	/ɛ/	/ɔ/	/a/
/i/	+	*	+	+	+	*	+
/u/	+	+	+	*	+	*	+
/ɪ/	+	+	+	(+)	+	+	+
/ʊ/	+	(+)	+	+	+	*	+
/ɛ/	*	*	(+)	+	+	*	*
/ɔ/	*	*	+	+	+	+	*
/a/	+	+	+	+	+	*	+

In terms of their cooccurence in noun roots, the Level 2 vowels can be grouped with the Level 1 and Level 3 vowels in different ways. They behave like /i u/ (and unlike /ɛ ɔ/) in that they can occur before /i u/, but they behave like /ɛ ɔ/ (and unlike /i u/) in that they can occur after /ɛ ɔ/. One of these groupings will be on the basis of [ATR] and the other on the basis of [high], but it is impossible to know which is which from the alternations alone.

In addition to these static patterns, there are also a number of morphologically-conditioned harmonic alternations, however these are often ambiguous as to whether they are ATR or height harmony. These alternations are given and discussed further in the following section.

3.3 Vowel alternations

The alternations involved here are all long-distance vowel feature assimilations, i.e. vowel harmony, except for one case of apparent dissimilation. They are grouped here by their direction of application. Regressive assimilations are those cases where the source of the assimilated features (the conditioning environment) is found later in the word than the target of the assimilations (the changing segment). Progressive assimilations are the opposite: the source is found earlier in the word than the target.

In all cases, I will briefly touch on what a phonological analysis might look like, mostly in aid of explaining the difficulties in ascertaining the identities of the Level 2 vowels, but I will not give a full phonological account of them, since this would be beyond the scope of this thesis (though certainly an interesting study for the future). All example data is from the main study of this thesis unless otherwise indicated.

3.3.1 Regressive assimilations

3.3.1.1 Recent past -i and causative -is

The main regressive assimilation investigated by my study is the alternation in stem vowels before two Boa-Leboale suffixes: the recent past suffix -i and the causative suffix -is.[1] The general rules are given in (17), and examples are given in (18).

(17) a. /ʊ/ → [u] / __C {-i, -is}
b. /ɛ, ɔ/ → [ɪ, ʊ] / __C {-i, -is}

(18) a. i. na- b**ʊ**m -a =ndɛ́
1S- frapper -FV =FUT
'je frapperai'

ii. na- b**ú**m -i
1S- frapper -RECPAST
'j'ai frappé'

iii. na- ˜- b**ú**m -is -a =ndɛ́
1S- 3SO- frapper -CAUS -FV =FUT
'je le ferai frapper'

b. i. na- l**ɛ**ŋg -a =ndɛ́
1S- faire.attention -FV =FUT
'je ferai attention'

ii. na- l**ɪ**ŋg -ɪ́
1S- faire.attention -RECPAST
'j'ai fait attention'

c. i. na- s**ɔ**m -ɔ =ndɛ́
1S- cacher -FV =FUT
'je cacherai'

---

[1] This assimilation does not seem to be triggered by the plural addressee suffix -ni, so this is potentially morphologically conditioned.

ii. na- s**ʊ**m -í
1S- cacher -RECPAST
‘j’ai caché’

iii. na- ˜- s**ʊ**m -is -a =ndɛ́
1S- 3SO- cacher -CAUS -FV =FUT
‘je le ferai cacher’

There are occasional attested examples of /ɪ/ → [i] in the same environment but this is rare, especially compared to the rules given for /ʊ, ɛ, ɔ/, which apply in almost all cases.

This set of assimilations is surprising for a number of reasons, principally that the assimilating feature seems to be different for (17a) and for (17b). If the Level 2 vowels are /ɪ ʊ/, then (17a) is an assimilation of the feature [ATR], and (17b) of the feature [high]. If they are instead /e o/, the features involved are reversed. The only way that both phonological rules can be described with one commonality is if Boa-Leboale in fact has 2 sets of Level 2 vowels, one of which is purely allophonic variants of /ɛ, ɔ/. For example, we could postulate that both rules involve the spreading of [+ATR] from the /i/ in the suffix to the root. Then rule (17a) works perfectly, but rule (17b) has the output [e, o] instead of [ɪ, ʊ]. This is one of the main hypotheses tested by Hypothesis 2 of the acoustic study (see §4.3.2, §4.3.3.2 and §5.2.2).

3.3.1.2 ɪ-dentity avoidance

The other regressive pattern found in Boa-Leboale vowels is the occasional dissimilation of /ɪ/ before a suffix containing /ɪ/, such as the subjunctive suffix -ɪ, the future negative -ɪ or the applicative -ɪlɪ. The rule is given in (19) and examples in (20).

(19) /ɪ/ → [ɛ] / __C -ɪ

(20) a. i. na- t**ɪ́**b -a =ndɛ́
1S- rire -FV =FUT
‘je rirai’

ii. tɪ́- t**ɛ́**b -ɪ́ =ndɛ́
1S.N- rire -N.F =FUT
‘je ne rirai pas’

iii. na- ˜- t**ɛ́**b -ɪlɪ -a =ndɛ́
1S- 3SO- rire -APPL -FV =FUT
‘je rirai pour qqn’

b. i. na- l**ɪ́**ŋg -á =ndɛ́
1S- enfler -FV =FUT
‘j’enflerai’

ii. ná- l**ɛ́**ŋg -ɪ́
1S- enfler -SUBJ
‘que j’enfle’

There are potential cases where this rule also seems to apply to /ʊ/, but in most of these cases, it is difficult to tell if the underlying vowel is /ʊ/ or /ɔ/ in the first place. This rule is notable because it is the only example of vowel dissimilation in Boa-Leboale. If the Level 2 vowels are /ɪ ʊ/, then this is dissimilation of the feature [high]; if they are /e o/, this is dissimilation of the feature [ATR].

The fact that this rule applies only to some verbs and not others suggests that the verbs should be either morphologically or phonologically distinct. Morphologically, this would be into different conjugation classes, but phonologically, there would need to be different underlying representations. One possible explanation is that an earlier stage of the language had 9 vowels. Then the verbs containing /e/ might become [ɛ] (with spreading of [−ATR]) to distinguish them from the roots with /ɪ/. However, without historical evidence it is difficult to speculate on the precise origins of this process, though we can investigate the possibility that present-day Boa-Leboale has 9 underlying vowels. This hypothesis, that there are two sets of underlying Level 2 vowels, is tested in Hypothesis 1 of the acoustic study (see §4.3.2, §4.3.3.1 and §5.2.1).

3.3.2 Progressive assimilations

3.3.2.1 Assimilation of suffixes with /ɪ/

One of the main progressive assimilations involves the same suffixes just discussed: the ones including the vowel /ɪ/. After a root with a Level 1 vowel as the last vowel, the suffixes are pronounced with /i/ instead of /ɪ/. The rule is given in (21) and some examples in (22).

(21) /ɪ/ → [i] / {i, u} C -__

(22) a. i. tɪ́- ɓín **-í** =ndɛ́
1S.N- danser -N.F =FUT
'je ne danserai pas'

ii. ná- ɓín **-í**
1S- danser -SUBJ
'que je danse'

iii. na- ˜- ɓín **-ili** -a =ndɛ́
1S- 3SO- danser -APPL -FV =FUT
'je danserai pour qqn'

b. i. na- ˜- gúg **-ili** -a =ndɛ́
1S- 3SO- suivre -APPL -FV =FUT
'je suivrai pour qqn'

ii. tɪ́- gug **-i** =ndɛ́
1S.N- suivre -N.F =FUT
'je ne suivrai pas'

In a /2IU/ analysis, this is progressive assimilation of [ATR]. In a /1IU/ analysis, this would be a progressive assimilation of [high], much like the inverse of the regressive assimilation of /ʊ/ before suffixes with /i/ (§3.3.1.1). Since the target of the assimilation is in the suffix, this rule does not vary from root to root like it would with the regressive assimilations. In fact, this rule applies to all verbal suffixes with /ɪ/, so there is no reason to speculate about the possibility of a 9-vowel system on the basis of this rule.

3.3.2.2 Spreading of /ɛ ɔ/

Suffixes containing /a/ are the target of a total assimilation if the final vowel of a root is one of the Level 3 vowels /ɛ ɔ/. This process is phonologically iterative: if there is a long string of suffixes containing /a/, they all undergo assimilation.

(23) /a/ → [ɛ, ɔ] / {ɛ, ɔ} C -__

(24) a. na- tɪ́b **-a** =ndɛ́
1S- rire -FV =FUT
'je rirai'

b. na- dɔp **-ɔ** =ndɛ́
1S- frapper -FV =FUT
'je frapperai'

This phonological process is less directly related to the questions involved in this thesis since it does not involve the Level 2 vowels in any capacity. It does however explain two of the gaps in the vowel combinations table (Table 3.1), namely why we do not find */ɛ...a/ or */ɔ...a/. It also shows that the progressive harmony is strictly to the right and does not affect the prefix.

3.3.2.3 Noun class suffixes

The final phonological alternation to discuss is the noun class suffix -yV. This suffix is found on a minority of the nouns in Classes 1 and 7 (and a few nouns in Classes 3, 5 and 9). It has three forms, depending on the level of the final vowel. There are (as far as I know) no attested nouns with final /i/ or /ɪ/ and a class suffix, presumably because the vowel in the root and suffix would merge in these cases. The rule for the realisation of the suffix is given in (25) and some examples are given in (26).[2]

(25) /-yV/
→ [-yi] / {u} __
→ [-yɪ] / {ʊ} __
→ [-yɛ] / {ɛ, ɔ, a} __

(26) a. i. ɪ́-ŋgáŋgú-**yi** 'AUG-1a.coléoptère sp.-1'
ii. ɪ́-gbú-**yi** '7-pâte de maïs-7'
b. i. ɪ-mʊ̌-**yɪ** 'AUG-3a.tête-3'
ii. ɪ-bʊ́-**yɪ** '7-médicament-7'
c. i. ɪ́-sɔ́sɔ́-**yɛ** 'AUG-1a.perdrix-1'
ii. ɪ-nɔ́nɔŋgɛ́-**yɛ** 'AUG-1a.fainéant-1'
iii. ɪ-n-ká-**yɛ** 'AUG-1-femme-1'

Phonologically, this is similar to the alternation in stem vowels before suffixes with /i/ (see example (17)), since they both concern all the vowels in Levels 1–3. Here, the suffix -yV seems to agree with the final vowel of the noun for both [ATR] and [high]. This is the case regardless of whether we analyse the Level 2 vowels as /ɪ ʊ/ or /e o/.

---

[2] These examples were kindly provided to me by Gerrit de Wit.

3.3.3 Alternations overview

An overview of all 5 phonological alternations is given in (27).

(27) a. /ʊ/ → [u], /ɛ, ɔ/ → [ɪ, ʊ] before recent past -i or causative -is.
b. /ɪ/ → [ɛ] (sometimes) before suffixes containing /ɪ/.
c. /ɪ/ → [i] in suffixes after a root containing /i, u/.
d. /a/ → [ɛ, ɔ] in suffixes after a root containing /ɛ, ɔ/.
e. The noun class suffix /-yV/ found on a minority of Class 1 and 7 nouns agrees for ATR and height with the final vowel of the root. It has the forms -yi, -yɪ and -yɛ.

These alternations do not obviously point to a particular analysis of the Level 2 vowels. This is particularly the case because the Level 2 vowels alternate with both the Level 1 vowels (in (27a), (27c), (27e)) and the Level 3 vowels (in (27a), (27b), (27e)) in different alternations. There are also some asymmetries between the front and back vowels: (27a) mostly applies to the back vowel /ʊ/, while (27b) mostly applies to the front vowel /ɪ/.

A full in-depth phonological analysis of these patterns is beyond the scope of this thesis, but presents fertile ground for further research. For the purpose of this thesis, it suffices to note that these patterns cannot be described simply, nor does the phonology clearly indicate one analysis of the Level 2 vowels over another.

Finally, I want to note that several of the alternations described here look similar on the surface to patterns in other languages. In particular, alternation (27a) is reminiscent of the alternation found in Ethiopian Komo, as described briefly in §2.2 and repeated here in (28).

(28)	a.	[k**ɛ́**ʃ-ʊ́]	‘thresh-DD1’
		[d**ɔ̀**t’-ʊ́]	‘ask-DD1’
		[h**á**m-ʊ́]	‘yawn-DD1’
	b.	[k**é**ʃ-úk]	‘thresh-DD2’
		[d**ò**t’-úk]	‘ask-DD2’
		[h**ə́**m-úk]	‘yawn-DD2’

Komo, like Boa, has an underlying 7-vowel system with the Level 2 vowels /ɪ ʊ/. However, as shown in (28), there are still allophonic surface vowels [e o ə], which appear as [+ATR] alternants of /ɛ, ɔ, a/. This raises the interesting proposition that the same process could be occurring in Boa-Leboale as well: perhaps alternation (27a) involves just ATR harmony and not a mixture of ATR and height harmony.

3.4 Related vowel systems

Besides Boa-Yewu, which is so similar to Leboale that there is little in the way of grammatical differences, the most closely related languages to Boa-Leboale include Pagibete, Kango, Ngelima and

Liko. I will briefly compare the vowel system of Boa-Leboale to Pagibete and Liko in terms of the vowel inventory and any phonological alternations.

Pagibete has a high level of lexical similarity to both Boa-Yewu and Boa-Leboale. Reeder (1998) analyses it as a 7-vowel system with [+ATR] Level 2 vowels /e o/. She also reports a number of vowel alternations similar to those found in Boa-Leboale. Firstly, Reeder (1998: 25) reports a progressive spreading process, shown in (29) where /ɔ/ spreads to following /a/, much like alternation (27d) in Leboale, which also mostly occurs with the back vowels. Reeder (1998: 26) also reports raising of /e o/ in verb roots before the past tense suffix -i and the causative extension -is, as shown in (30).

(29)	a.		a-ɓún-**a**	‘s/he will crack’
	b.		a-ɓɔ́t-**ɔ**	‘s/he will give birth’
(30)	a.	i.	a-l**é**k-a	‘s/he will set a snare’
		ii.	a-l**í**k-í	‘s/he set a snare’
	b.	i.	a-s**o**s-a	‘s/he will wash’
		ii.	a-s**u**s-í	‘s/he washed’
		iii.	n-s**u**s-ís-ɛ́	‘to cause to wash’

The vowel raising in (30) is similar to Leboale alternation (27a), though it affects only the Level 2 vowels, not the Level 3 vowels /ɛ ɔ/.[3] It is also notable that the alternation applies equally to front /e/ and back /o/, where the equivalent alternation in Boa-Leboale mostly applies to /ʊ/ and only rarely to /ɪ/. De Wit (2020: 427–431) finds that in Boa-Yewu, the alternation is even more restricted: there is only one attested case where /ɪ/ is raised, and only around half of the verbs with /ʊ/ undergo the assimilation to /u/.

Liko is comparatively less close to Boa-Leboale in terms of lexical similarity, but also presents another useful point of comparison. Where the alternations in Pagibete can be accounted for without the inclusion of ATR, Liko's phonological alternations are almost all examples of ATR harmony. De Wit (2015: 44) describes Liko as having a 9-vowel system, with both /ɪ ʊ/ and /e o/. This system is best described not in terms of different levels, but in terms of the combination of 3 vowel heights and the values of ATR, since all vowels but /a/ can be split into harmonic pairs which are identical for all other feature values.[4]

There are comparatively more vowel alternations in Liko compared to Boa-Leboale or Pagibete, but there are some which cover the same ground as the alternations discussed so far and can be used to compare how each language works differently – or even to offer a new perspective on the other languages. Like Boa-Leboale and Pagibete, verb root vowels are affected by the recent past and causative suffixes in Liko, as can be seen in (31).

(31)	a.	i.	ká-ɓ**ɪ́**k-á	‘to despise’

---

[3] Reeder (1998: 27) notes that Level 3 /ɛ/ is also raised in a few frequent verbs, but it is raised to Level 1 /i/, not to Level 2 like in Boa-Leboale.

[4] /a/ does have a [+ATR] alternant, [o], in [+ATR] contexts, despite differing for the feature [+round]. Thus in Liko [o] can correspond with underlying /a/ or /ɔ/; this is often known as harmonic ‘re-pairing’ (see, for instance, Baković (2000: §2.5)).

		ii.	no-ɓ**í**k-i	‘I despised’
		iii.	ká-k**ɪ́**t**ɪ́**l-á	‘to block the road’
		iv.	kó-k**í**t**í**l-ís-ó	‘to cause to block the road’
	b.	i.	ká-kp**ʊ**ɗ-á	‘to approach’
		ii.	no-kp**u**ɗ-í	‘I approached’
		iii.	ká-b**ʊ**m-á	‘to hit’
		iv.	kó-b**u**m-ís-ó	‘to cause to hit’

The main difference between the Liko and Boa-Leboale alternations is that the alternation is not one of vowel raising, but just spreading of [+ATR]. The other difference is that there is no raising or spreading of [+ATR] to the [–ATR] mid vowels /ɛ ɔ/, because Liko has no verbs with underlying /ɛ ɔ/. We can imagine that there would be spreading of [+ATR] to give surface [e o], but there is no way to know for certain.

While there is no equivalent of the dissimilation process from Boa-Leboale, there is also an equivalent of the progressive assimilation of the suffixes with /ɪ/. In Liko, most suffixes are underlyingly [–ATR], including the benefactive suffix -ɪlɪ (which is cognate to the Boa-Leboale applicative suffix -ɪlɪ). When any [–ATR] affix appears with a [+ATR] root, there is spreading of [+ATR] to the affixes. This idea that the presence of any [+ATR] vowels will spread [+ATR] to all vowels in a word is known as ATR dominance. The example with the benefactive suffix can be seen in example (32) (de Wit 2015: 82).

(32)	a.	kā́-ꜜɓɪ́t-**ɪ́**ly-**a**	‘to slap for someone’
	b.	kā́-lut-**í**ly-**o**	‘to pull for someone’

These two examples show that how, in some ways, the messy system in Boa-Leboale is not particularly different from the regimented ATR-dominance system found in Liko. The regressive and progressive assimilation of [+ATR] between roots and suffixes in Boa-Leboale is reminiscent of an ATR-dominance system that has long since decayed through the erosion of the vowel system from 9 vowels to 7 and from a growing number of exceptions to each assimilation rule. Boa-Yewu and Pagibete also seem to be similar decayed systems, with Yewu having developed even more exceptions than Leboale and Pagibete having reanalysed one instance of ATR assimilation in terms of vowel raising.

3.5 Conclusions

In this chapter, I introduced Boa-Leboale and gave an overview of the vowel system: its inventory, static cooccurence restrictions and phonological patterns. I then briefly discussed two closely-related languages – Pagibete and Liko – to give an impression of how similar phonological patterns can be analysed in a number of ways. While this thesis will not focus on the phonological patterning of Boa-Leboale, it is clear that it is a particularly fertile ground for future research.

Instead, the remainder of this thesis will focus on an acoustic study conducted on Boa-Leboale with the aim of answering some of the main questions raised in this chapter. Does Boa-Leboale have a /2IU/ or a /1IU/ vowel system? Does Boa-Leboale have allophonic [+ATR] mid vowels [e o] in [+ATR] contexts?

Chapter 4

Methodology

In this chapter, I will outline the methodology for the principal study undertaken for this thesis. The data collection and annotation (§4.1) was undertaken in the Congo by my supervisor, Gerrit de Wit, so I will focus primarily on the analytic methods which I performed myself, both acoustic (§4.2) and statistical (§4.3). This methodology is mostly based on the one employed by Starwalt (2008), which is described in §2.4. One crucial difference between her methodology and my own is that hers was conducted on many speakers of many languages, and most of the measurements seem to have been taken manually. By contrast, this study focuses on just one speaker of Boa-Leboale and the measurements were taken automatically by a Praat script (given in full in Appendix A) running in Praat (v6.2.02; Boersma & Weenink 2021). I will include snippets of this Praat script at the relevant moments during the discussion of the methodology for the acoustic analysis (§4.2). Appendix B additionally provides the R script used in the statistical analysis. I include these scripts in the thesis to illustrate their use and to encourage their replicability, since this is one major advantage of the automatic data analysis method.

4.1 Data collection and annotation

The data involved in this acoustic study are a subset of a number of recordings made by Gerrit de Wit in Isiro, Haut-Uélé Province, in May 2022 with the help of the Boa team (consisting of the CITBA linguist Dominique Banotanea Bapokanzo, along with Patrick Alipanazanga Masanga, Rev. Sébastien Banale Aleumba, Julienne Etindepua Embatia and Dieudonné Tazanaba Kadite).

The recordings used in this study involved the speaker, Dieudonné Tazanaba, reading aloud words from two different wordlists. The first wordlist is a list of 295 nouns where each of the 7 vowels is found in the stressed syllable, mostly with the same vowel in the following syllable. The distribution of words across the vowels is as follows: /a/ 58, /ɪ/ 40, /ɛ/ 28, /i/ 37, /ʊ/ 43, /ɔ/ 41, /u/ 48.

The second wordlist is a list of 624 verbforms. These verbforms are in a number of different morphological contexts: 85 in the future tense, 105 future applicatives, 104 subjunctives, 84 recent past, 123 future causatives and 123 future negatives. For each morphological context, the verbs used include 6 of the 7 vowels in Boa-Leboale as their main vowel (all excluding /a/).

De Wit used a Tascam DR-05 Linear PCM Recorder to make the recordings with a sample rate of 44,100Hz and a bit depth of 16 bits. The recordings were saved as uncompressed WAV files.

Once the recording was complete, de Wit annotated the audio files in Praat (Boersma & Weenink 2021) with TextGrids, where each TextGrid had two interval tiers: one to demarcate the vowels to be analysed acoustically and one to demarcate and gloss the words themselves. These TextGrid files, along with the WAV files of the recordings and Microsoft Word documents of the wordlists, constituted the primary materials that I was provided by de Wit for the purposes of my study.

4.2 Acoustic analysis

The first portion of analysis involved taking the data collected and annotated as described in §4.1 and using a Praat script to automatically measure a number of acoustic variables. These variables are the same as those from Starwalt (2008), plus cepstral peak prominence (CPP), which was measured following Hillenbrand et al. (1994) and Olejarczuk et al. (2019).

4.2.1 Automatic measurement

As mentioned earlier, one of the principal differences between the methodology of this study and the one in Starwalt (2008) is the automatisation of the measurement collection using Praat scripting.

4.2.1.1 Praat scripts

I want to first introduce some of the primary concepts of how the Praat scripting language functions before describing how it was used for collecting the relevant acoustic measurements. Praat scripts primarily involve the manipulation and analysis of Objects – these Objects are any files imported into Praat or any analysis files produced by Praat. The files imported into Praat will tend to be audio files and annotation files, which become **Sound** Objects and **TextGrid** Objects respectively. These Objects can be analysed into other Objects; for instance, if a Sound Object is analysed with a Fourier transform, the output is a **Spectrum** Object.

The Praat scripting language is, generally speaking, used to perform three types of function: modify a given Object, convert between Object types, and take measurements from a given Object. Many important acoustic measurements cannot be measured directly from a Sound Object, which is why intermediate Objects must be created first. Many Object types can also be exported in a graphical format, though this feature is not frequently employed in this study.

By way of demonstration, one of the measurements taken in this study is F1, the frequency of the first formant. Since we are measuring the average formant frequency over the steady state of the vowel, we must first extract this portion of the vowel into its own Sound Object. Then we must analyse this Sound Object into a Formant Object, which we can then query for the mean frequency of the first formant. The script for these actions is given in (33), showing the three main functions being used in purple: `Extract part`, `To Formant (burg)` and `Get mean`.

(33)
```
    selectObject: idSnd
    idVowelSnd = Extract part: t1[i], t3[i], "rectangular", 1, "yes"
```

```
idVowelForm = To Formant (burg): 0, nFormants, freq_range, 0.025,
    50
```

```
f1[i] = Get mean: 1, t1[i], t3[i], "hertz"
```

Manipulations of Objects in this way form the foundation of the Praat script used to measure all the acoustic correlates involved in this script.

4.2.1.2 Pros of automatic measurement

But why use a Praat script at all? After all, these measurements can be taken using the Praat graphical user interface without using any scripts (henceforth ‘manual measurement’). The advantages of automatic measurement are, broadly speaking, speed and consistency.

Automatic measurement is much faster than manual measurement for large amounts of data, which are needed for such studies. Starwalt’s study involved the measurement of 6 correlates per token, 20 tokens per vowel, 7-9 vowels per speaker, 3-5 speakers per language, and 11 languages in total. This gives a ballpark estimate of 27,000 to 60,000 separate measurements that were taken. Doing all of these measurements in the graphical user interface would not only take a very long time, but would also be extremely tiresome for the researcher. Using automatic measurement, studies can analyse more tokens in the same amount of time and the more tokens are taken, the more statistically powerful any findings will be.

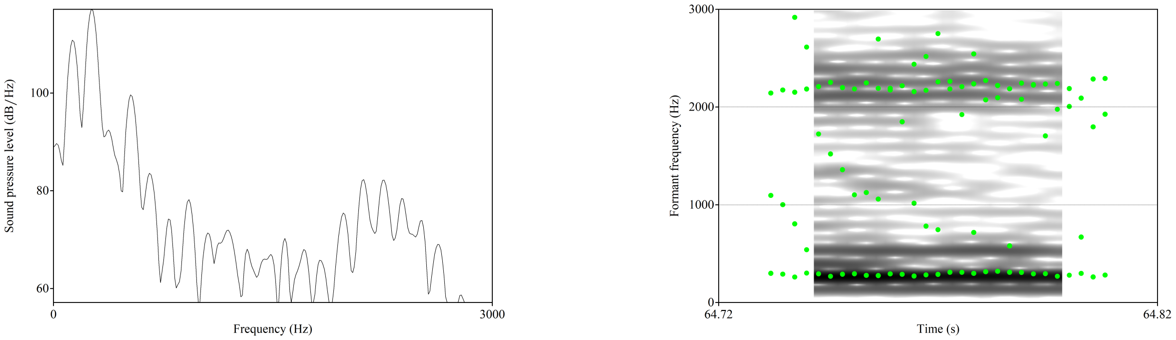
The other advantage to automatic measurement is that it can be applied systematically and consistently: anyone with the same data and script could get the same measurements, which is less likely to be the case for manual measurement where humans might introduce rounding or copying errors, or might be inconsistent in where a given measurement is taken. This consistency of application also means that studies with automatic measurements are verifiable, more easily replicable, and that the same script can be applied to new data by other researchers with increased comparability between studies.

Finally, in this particular study, automatic measurement is particularly useful as the Praat script can also perform the modelling of B1, A1 and A2 directly, rather than being performed in another program with the output from the Praat graphical user interface.

4.2.1.3 Cons of automatic measurement

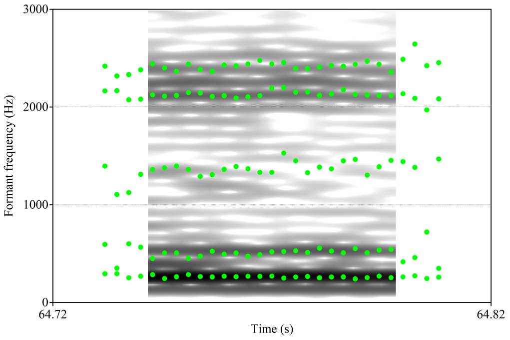
However, automatic measurement is by no means perfect. The main disadvantage of automatic measurement compared to manual measurement is that it is particularly prone to measurement errors. Since manual measurement is human-mediated, obvious measurement errors can and will be resolved on a case-by-case basis.

An example from the Boa study can show what types of measurement errors are most common in formant tracking. Formant tracking in Praat is usually reliable, but can be quite sensitive to the parameters input by the researcher: the maximum formant frequency and the maximum number of formants being measured (Barreda 2021). This difficulty is compounded when two formants are particularly close to one another, as with F1 and F2 for rounded back vowels like /o u/. In this case, the formants overlap greatly, meaning it is often difficult to distinguish them from one another if the maximum formant frequency is set too high or the maximum number of formants is set too low. An example is shown in Figure 4.1.



(a) Spectrum of vowel midpoint.

(b) Spectrogram and formant tracking with default settings.



(c) Spectrogram and formant tracking with manually adjusted settings.

Figure 4.1: Demonstration of a F2 tracking error in the vowel [ũ] in Boa-Leboale nãlʊ́ŋgisandɛ́ ‘je le ferai téter’.

In the lower frequencies of Figure 4.1a, the first two formants can clearly be identified by eye: F1 includes harmonics 2 and 3 around 200Hz, while F2 is most associated with harmonic 5 at around 500Hz. This can also be seen on the spectrogram in Figures 4.1b and 4.1c, where F2 is the second dark band from the bottom. However, using the automatic formant tracking with default settings (maximum formant frequency = 5000Hz, maximum number of formants = 5.5[1]) only correctly identifies F1 and F3, with values for F2 varying sporadically in between the two, as can be seen in Figure 4.1b. It is only possible to correctly track F2 following human intuition by changing one of these two values, as shown in Figure 4.1c, which instead has a maximum formant frequency of 3000Hz. This solution is not without its own problems: it is now tracking an phantom formant between F2 and F3. Setting the maximum formant frequency to 3000Hz would also give entirely incorrect readings for other vowels. Since the formant tracking in Praat is blind to the vowel being analysed, it cannot take these factors into account, where a human researcher easily can.

There are several tools which have been developed to improve Praat’s formant tracking, including

---

[1] Praat’s formant tracking algorithm computes a number of ‘poles’ equal to twice the number of formants. This is why the maximum number of formants can be given to the nearest 0.5.

Fast Track (Barreda 2021), a suite of Praat scripts which runs multiple formant analyses with a range of maximum formant frequencies and selects an optimal analysis based on the smoothness of the formant track. Fast Track also involves a number of heuristics to rule out formant values that are considered impossible or unrealistic, several of which I adopted in my own script to avoid some of these measurement error outliers.

Regardless of these errors, I still believe that automatic measurement is clearly superior to manual measurement for larger scale data analyses such as these. The best solution would be to implement a system such as Fast Track, or in its absence to involve a number of heuristics directly in the Praat script that can mitigate some of the most common measurement errors (as is done in this study). Finally, one of the advantages of automatic measurement can also help against its disadvantage: the more data is collected, the less impact outliers have on the statistical power of the data. As long as the errors are not systematic and as long as there are many more valid measurements than erroneous ones, the presence of a few outliers will be more than compensated for by a larger dataset.

4.2.2 Points of measurement

Like with Starwalt's methodology (§2.4.1), some measurements were best averaged over the steady state of the vowel, others taken at a single point within that steady state. To avoid transition effects at the edge of the vowel, I approximated the steady state of the vowel as the section between 30% and 70% of the annotated interval. The single point was taken as the midpoint of the interval. This can be seen in (34), which shows three extracts of different parts of the Praat script. (34a) shows how the duration was measured, along with `t1` (the start of the steady state), `t2` (the midpoint of the steady state) and `t3` (the end of the steady state). (34b) shows which analysis Objects were created at the midpoint `t2`: a spectral slice and an LTAS (long-term average spectrum), while (34c) shows the steady state of the vowel (between `t1` and `t3`) being extracted for further analyses.

(34) a.

```
starttime[i] = Get start time of interval: 1, interval
endtime[i] = Get end time of interval: 1, interval
duration[i] = endtime[i] - starttime[i]
t1[i] = 0.3 * duration[i] + starttime[i]
t2[i] = 0.5 * duration[i] + starttime[i]
t3[i] = 0.7 * duration[i] + starttime[i]
```

b.

```
selectObject: idVowelSpecgram
idVowelSlice = To Spectrum (slice): t2[i]
idVoweltas = To Ltas (1-to-1)
```

c.

```
selectObject: idSnd
idVowelSnd = Extract part: t1[i], t3[i], "rectangular", 1,
    "yes"
```

4.2.3 Formant frequency and bandwidth (F1, F2, F3, B1, ΔB1)

Regarding formants, the script takes five measurements of formant frequency and one of formant bandwidth: `f1`, `f2`, `f1mid`, `f2mid`, `f3mid` and `b1`. The first two (`f1`, `f2`) are averaged over the steady state of the vowel and are the values which will be referred to as F1 and F2 throughout this methodology and in the results. The remaining four (`f1mid`, `f2mid`, `f3mid`, `b1`) are measurements taken specifically at the midpoint of the vowel – these measurements of formant frequency (`f1mid`, `f2mid`, `f3mid`) are used only as part of the modelling of B1, A1 and A2.

For all measurements, the extracted Sound Object of the steady state of the vowel is analysed into a Formant Object, using the Burg algorithm for LPC analysis (35a). The input parameters are generally set to Praat's default values unless they are input directly by the user. The time step is set to the default (25% of the window length), the maximum number of formants and maximum formant frequency are both input by the user when the script first runs. I used the standard values of maximum 5 formants with a maximum frequency of 5000Hz, since the speaker is male. The window length is set to be 25ms (giving a time step of 6.25ms) and the pre-emphasis 50Hz, both the default values.

The six measurements are then taken directly from this Formant Object, either taking the average across the steady state or the value at the midpoint (35b). The output is given in Hz and values taken at the midpoint are measured with linear interpolation. This is because the Formant Object will not have a measurement that precisely lines up with the midpoint of the vowel, so it interpolates the track of the formant between the two closest measurements and returns the value in-between.

(35) a.

```
selectObject: idVowelSnd
idVowelForm = To Formant (burg): 0, nFormants, freq_range,
    0.025, 50
```

b.

```
selectObject: idVowelForm
f1[i] = Get mean: 1, t1[i], t3[i], "hertz"
f2[i] = Get mean: 2, t1[i], t3[i], "hertz"
b1[i] = Get bandwidth at time: 1, t2[i], "hertz", "linear"
```

```
f1mid = Get value at time: 1, t2[i], "hertz", "linear"
f2mid = Get value at time: 2, t2[i], "hertz", "linear"
f3mid = Get value at time: 3, t2[i], "hertz", "linear"
```

Early stages of the data analysis showed that there were a number of measurement errors where F1 and B1 were measured as being much higher than what could be seen by eye. As such, I implemented two of the automatic error-detection heuristics from Fast Track (Barreda 2021), given in (36).

(36) a. F1 should not be higher than 1000Hz, especially for a male speaker.[2]

b. B1 should not be higher than 500Hz.

These heuristics were implemented relatively crudely: if any condition is not met with the measurements from the Formant Object, an alternative Formant Object is created with a lower maximum formant frequency (4900Hz) and new values for all formant measurements are remeasured. If any conditions are still not met, then another Formant Object is created with an even lower maximum

---

[2] Barreda (2021) uses 1200Hz as a cut-off, but the outliers in my data were all clearly above 1000Hz, so I opted for this instead.

formant frequency (4800Hz) and all formant measurements retaken. There is no application of the heuristic if the heuristics are still not met after two iterations, so outliers are still possible in the final measurements. The script for this process is shown in (37).

(37)

```
if f1[i] > 1000 or b1[i] > 500 or f1mid > 1000
    # appendInfoLine: "Barreda heuristic condition met: ", words$[i], " f1: ", f1[i], " b1: ", b1[i]

    selectObject: idVowelSnd
    idVowelFormAlt = To Formant (burg): 0, nFormants, 4900, 0.025, 50

    selectObject: idVowelFormAlt
    f1[i] = Get mean: 1, t1[i], t3[i], "hertz"
    f2[i] = Get mean: 2, t1[i], t3[i], "hertz"
    b1[i] = Get bandwidth at time: 1, t2[i], "hertz", "linear"
    f1mid = Get value at time: 1, t2[i], "hertz", "linear"
    f2mid = Get value at time: 2, t2[i], "hertz", "linear"
    f3mid = Get value at time: 3, t2[i], "hertz", "linear"

    if f1[i] > 1000 or b1[i] > 500 or f1mid > 1000
        # appendInfoLine: "Barreda heuristic still met: ", words$[i], " f1: ", f1[i], " b1: ", b1[i]

        selectObject: idVowelSnd
        idVowelFormAlt2 = To Formant (burg): 0, nFormants, 4800, 0.025, 50

        selectObject: idVowelFormAlt2
        f1[i] = Get mean: 1, t1[i], t3[i], "hertz"
        f2[i] = Get mean: 2, t1[i], t3[i], "hertz"
        b1[i] = Get bandwidth at time: 1, t2[i], "hertz", "linear"
        f1mid = Get value at time: 1, t2[i], "hertz", "linear"
        f2mid = Get value at time: 2, t2[i], "hertz", "linear"
        f3mid = Get value at time: 3, t2[i], "hertz", "linear"

        removeObject: idVowelFormAlt2
    endif
    # appendInfoLine: "New values: ", words$[i], " f1: ", f1[i], " b1: ", b1[i]
    removeObject: idVowelFormAlt
endif
```

The displacement of B1 from predicted values, ΔB1, is also calculated directly in the Praat script by implementing Fant's (1972) formula (see §2.4.1.2). Fant's formula and the Praat script extract are compared in (38). Note also that the F1 value used here is `f1mid` rather than `f1`, so that all measurements entered into the formula are measured in the same way.

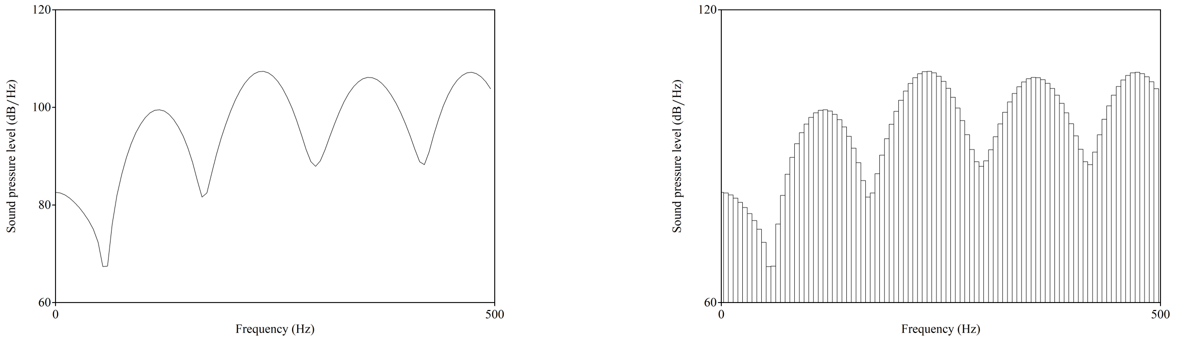
(38) a.
$$\Delta B_1 = B_1 - \left(15\left(\frac{500}{F_1}\right)^2 + 20\left(\frac{F_1}{500}\right)^{\frac{1}{2}} + 5\left(\frac{F_1}{500}\right)^2\right)$$

b.
```
deltab1[i] = b1[i] - (15 * (500 / f1mid) ^ 2 + 20 * (f1mid / 500) ^ 0.5 + 5 * (f1mid / 500) ^ 2)
```

4.2.4 Formant amplitude (A1, A2, A1*–A2*)

The formant amplitude measures `a1` and `a2` involve a different path of Praat Objects than those of formant frequencies. This is because Praat does not have the function to measure the maximum amplitude in a given frequency range from a Formant or Spectrum Object.[3]

The closest workaround (shown in (39)) is instead to first convert the original Sound Object into a Spectrogram Object. Then you create a Spectrum Object from a single point in that Spectrogram Object, which you can then convert into a Long-Term Average Spectrum (or LTAS) Object. This effectively converts the samples in the Spectrum Object into a number of narrow frequency bands (see Figure 4.2) giving the power spectral density (informally, the amplitude) for each frequency band. By choosing a 1-to-1 conversion, the frequency resolution can be kept the same, such that both Objects contain the same information in different formats. The LTAS Object can then be queried for the maximum value in a given range (with parabolic interpolation, as recommended in the Praat manual when an LTAS Object is created with a 1-to-1 conversion).



(a) Example of a spectrum.

(b) Example of an LTAS (long-term average spectrum).

Figure 4.2: Comparison of a spectrum and LTAS, showing the 0–500Hz range of the spectrum taken at the midpoint of the vowel [ɪ] from Boa-Leboale nabɪlandɛ́ 'je refuserai'.

(39)
```
selectObject: idSnd
idVowelSpecgram = To Spectrogram: 0.03, 11025, 0.005, 5, "Gaussian"
```

```
    selectObject: idVowelSpecgram
    idVowelSlice = To Spectrum (slice): t2[i]
    idVoweltas = To Ltas (1-to-1)
```

```
    selectObject: idVoweltas
    a1[i] = Get maximum: f1mid - 150, f1mid + 150, "parabolic"
    a2[i] = Get maximum: f2mid - 150, f2mid + 150, "parabolic"
```

---

[3] There is a way to find the closest peak in energy to a given frequency, but this will not necessarily respond to the highest energy of *any* harmonic in the formant range. As such, it is better to use a measure that can search within a range.

The settings for the Spectrogram Object are those for a narrow-band spectrogram with a window length of 30ms, maximum frequency of 11025Hz, time step of 5ms and frequency step of 5Hz. The spectrogram must be narrow-band in order to have enough details on the precise intensity of each harmonic.

The spectrogram is also taken for the entire sound file, with spectral slices taken at the midpoints of each relevant vowel, instead of taking spectrograms for individual vowels. This is for two reasons: firstly, some vowels are too short to be analysed as their own spectrograms, and secondly, it is marginally faster to have the spectrogram analysis performed just once over the entire sound file than to extract a larger portion around each vowel and take spectrograms of each one.[4]

Which range should be queried to find the intensity maxima for A1 and A2? I opted for the measured values of F1 and F2 ± 150Hz, since the first few datapoints I looked at suggested that the appropriate size of the range should be 300Hz, even for F1. Other linguists have been more conservative with their frequency ranges: in other scripts measuring formant amplitudes, the frequency range is instead ± 10% of the measured value of F1. This would give ranges of only around 60–100Hz for F1, and 100–500Hz for F2. My experience suggested that these ranges were too narrow compared to what I would have chosen from visual inspection, hence my choice of a 300Hz range.

Like with ΔB1, the modelling of A1 and A2 is also performed directly within the Praat script, following the formulae given by Fulop et al. (1998) (see §2.4.1.3). Examples (40) and (41) show the comparison between the formulae and the Praat script respectively.

(40) a. $$\hat{A}_F(f) = 20\log_{10}\left(\frac{F^2+\left(\frac{b}{2}\right)^2}{\sqrt{(f-F)^2+\left(\frac{b}{2}\right)^2}\sqrt{(f+F)^2+\left(\frac{b}{2}\right)^2}}\right)$$

b. $$\hat{A}_{upper}(f) = 0.72\left(\frac{f}{492}\right)^2 + 0.0033\left(\frac{f}{492}\right)^4$$

c. $$\hat{A}_{loss}(f) = g\left(-20\log_{10}\left(2\frac{\frac{f}{100}}{1+\left(\frac{f}{100}\right)^2}\right)\right)$$

d. $$\hat{A}(f) = \hat{A}_1(f) + \hat{A}_2(f) + \hat{A}_3(f) + \hat{A}_{upper}(f) + \hat{A}_{loss}(f)$$

e. $$A_1^* - A_2^* = (A_1 - A_2) - (\hat{A}_1 - \hat{A}_2)$$

(41)

```
modelb1 = 30
modelb2 = 80
modelb3 = 150
f = f1mid
modela1_f1 = 20 * log10 ((f1mid ^ 2 + (modelb1 / 2) ^ 2)/(sqrt((f
    - f1mid) ^ 2 + (modelb1 / 2) ^ 2) * sqrt((f + f1mid) ^ 2 +
    (modelb1 / 2) ^ 2)))
modela1_f2 = 20 * log10 ((f2mid ^ 2 + (modelb2 / 2) ^ 2)/(sqrt((f
    - f2mid) ^ 2 + (modelb2 / 2) ^ 2) * sqrt((f + f2mid) ^ 2 +
    (modelb2 / 2) ^ 2)))
```

---

[4] In some cases, the TextGrid annotation only covers a portion of a given sound file. With the current set-up, the entire sound file is still converted to a Spectrogram Object, which would be time-inefficient. There is, therefore, room to make the Praat script run more efficiently in these situations as well.

```
modela1_f3 = 20 * log10 ((f3mid ^ 2 + (modelb3 / 2) ^ 2)/(sqrt((f
    - f3mid) ^ 2 + (modelb3 / 2) ^ 2) * sqrt((f + f3mid) ^ 2 +
    (modelb3 / 2) ^ 2)))
modela1_fN = 0.72 * (f / 492) ^ 2 + 0.0033 * (f / 492) ^ 4
modela1_glot = 20 * log10 (2 * ((f / 100) / (1 + (f / 100) ^ 2)))
modela1 = modela1_f1 + modela1_f2 + modela1_f3 + modela1_fN +
    modela1_glot
f = f2mid
modela2_f1 = 20 * log10 ((f1mid ^ 2 + (modelb1 / 2) ^ 2)/(sqrt((f
    - f1mid) ^ 2 + (modelb1 / 2) ^ 2) * sqrt((f + f1mid) ^ 2 +
    (modelb1 / 2) ^ 2)))
modela2_f2 = 20 * log10 ((f2mid ^ 2 + (modelb2 / 2) ^ 2)/(sqrt((f
    - f2mid) ^ 2 + (modelb2 / 2) ^ 2) * sqrt((f + f2mid) ^ 2 +
    (modelb2 / 2) ^ 2)))
modela2_f3 = 20 * log10 ((f3mid ^ 2 + (modelb3 / 2) ^ 2)/(sqrt((f
    - f3mid) ^ 2 + (modelb3 / 2) ^ 2) * sqrt((f + f3mid) ^ 2 +
    (modelb3 / 2) ^ 2)))
modela2_fN = 0.72 * (f / 492) ^ 2 + 0.0033 * (f / 492) ^ 4
modela2_glot = 20 * log10 (2 * ((f / 100) / (1 + (f / 100) ^ 2)))
modela2 = modela2_f1 + modela2_f2 + modela2_f3 + modela2_fN +
    modela2_glot

modela1_a2 = modela1 - modela2
```

```
a1_a2[i] = a1[i] - a2[i]

norma1_a2[i] = a1_a2[i] - modela1_a2
```

4.2.5 Centre of gravity

The measurement of centre of gravity is the most simple of all the acoustic measurements involved in this study. It can be queried directly from a Sound Object, and I used the default Praat settings with the power set to 2, as shown in (42). The centre of gravity is measured over the entire steady state of the vowel, as per Starwalt (2008).

(42)

```
selectObject: idVowelSpecRange
cog[i] = Get centre of gravity: 2.0
```

4.2.6 Cepstral peak prominence

CPP (cepstral peak prominence) is the only acoustic measure in this study not present in Starwalt (2008). Instead, the methodology is broadly taken from Hillenbrand et al. (1994), though in practice simply follows the default parameters suggested by the Praat manual.

CPP is a measure of periodicity: how much of the sound can be analysed as periodic. It involves analysing a cepstrum, an inverse spectrum, in terms of how prominent the largest cepstral peak is. A cepstrum with a particularly prominent peak is typical of a periodic signal (for the human voice, this will usually be the fundamental frequency of modal voicing), while a cepstrum without a prominent peak is more aperiodic. CPP is in effect a normalised measure of cepstral peak amplitude – since amplitude is typically higher at lower quefrencies (inverse frequencies), a regression line is fitted to the cepstrum, and prominence simply refers to the displacement of the amplitude from the regression line. This can be seen in Figure 4.3, which shows a cepstrum and its regression line. The most prominent cepstral peak at a quefrency of 8.45ms (around a third of the way from the left) can be seen clearly.

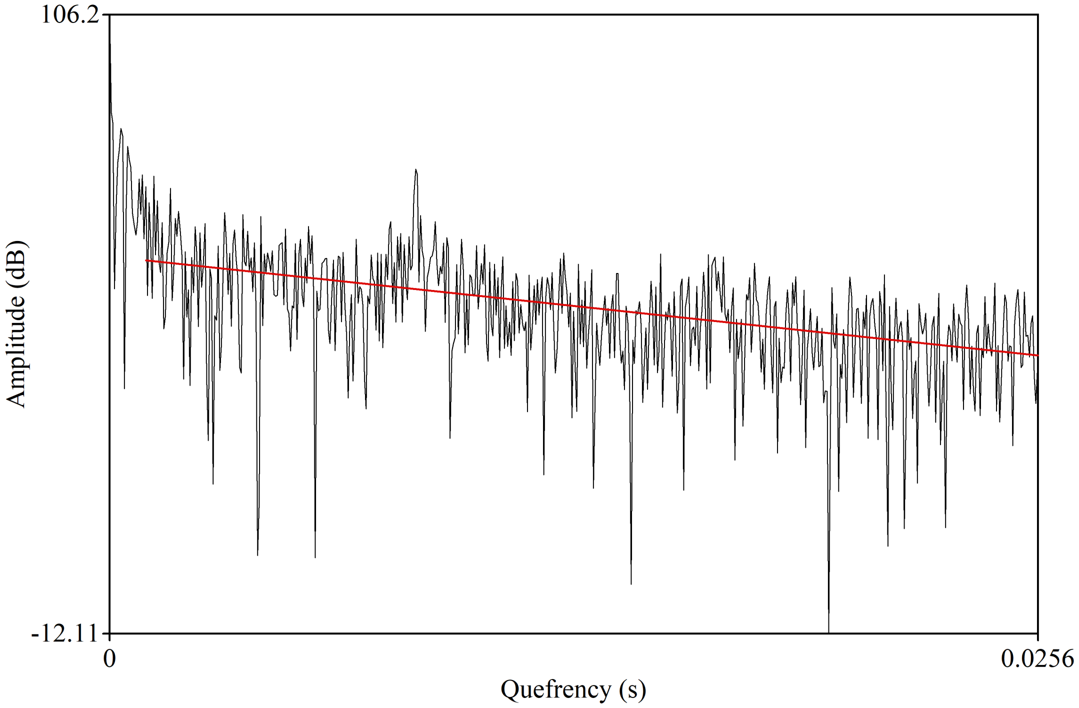


Figure 4.3: Cepstrum of the midpoint of the vowel [ɪ] from Boa-Leboale nabɪlandɛ́ 'je refuserai', showing the straight regression line in red.

Following Hillenbrand et al. (1994), CPP is measured without filters, but also with a 2500–3500Hz bandpass filter and a 2500Hz highpass filter, since the lower frequencies are often more periodic than their higher counterparts, even in breathy voiced speech. The non-filtered CPP and the average of all three CPP measurements are both reported. The creation of each of the different filtered Sound Objects is shown in (43).

(43)

```
selectObject: idVowelSnd
idVowelSndBandPass = Filter (pass Hann band): 2500, 3500, 100
selectObject: idVowelSnd
idVowelSndHighPass = Filter (pass Hann band): 2500, 0, 100
```

Each filtered version of the steady state of the vowel is then converted from a Sound Object to a PowerCepstrogram Object (effectively, an inverse of a spectrogram). Then a cepstral slice is taken at the midpoint of the vowel, giving a PowerCepstrum Object which is then queried for CPP directly.

All parameter values given are the Praat default settings. This process is shown in (44) for all three types, plus the average calculation at the end.

(44)

```
#NoFilter
selectObject: idVowelSnd
idCepsgramNF = To PowerCepstrogram: 60, 0.002, 5000, 50
idCepsRangeNF = To PowerCepstrum (slice): t2[i]
cpp_nofilt[i] = Get peak prominence: 60, 333.3, "parabolic",
    0.001, 0.05, "Straight", "Robust slow"
removeObject: idCepsgramNF, idCepsRangeNF

#BandPass
selectObject: idVowelSndBandPass
idCepsgramBP = To PowerCepstrogram: 60, 0.002, 5000, 50
idCepsRangeBP = To PowerCepstrum (slice): t2[i]
cpp_band[i] = Get peak prominence: 60, 333.3, "parabolic", 0.001,
    0.05, "Straight", "Robust slow"
removeObject: idCepsgramBP, idCepsRangeBP

#HighPass
selectObject: idVowelSndHighPass
idCepsgramHP = To PowerCepstrogram: 60, 0.002, 5000, 50
idCepsRangeHP = To PowerCepstrum (slice): t2[i]
cpp_high[i] = Get peak prominence: 60, 333.3, "parabolic", 0.001,
    0.05, "Straight", "Robust slow"
removeObject: idCepsgramHP, idCepsRangeHP

cpp_avg[i] = (cpp_nofilt[i] + cpp_band[i] + cpp_high[i] ) / 3
```

4.2.7 Data output

The Praat script saves the value for each relevant measurement as a value in an array.[5] At the end of the script, all the measurements are appended as a new line in the output file, the format of which is determined by the user. I exported all the data in a comma-separated format (CSV). The line of the script that produces the output file is shown in (45).

(45)

```
appendFileLine: "'outputfile$''ext$'", filename$, sep$,
    words$[i], sep$, t2[i], sep$, sep$, sep$, sep$, sep$,
    duration[i], sep$, f1[i], sep$, f2[i], sep$, b1[i], sep$,
    deltab1[i], sep$, normal_a2[i], sep$, cog[i], sep$,
    cpp_nofilt[i], sep$, cpp_avg[i]
```

To prepare the data for statistical analysis, I then imported the resulting CSV file into Microsoft Excel, and manually filled in the columns for surface vowel, underlying vowel, ATR value (Y for

---

[5] This is not strictly necessary, but given that the measurements are taken for each vowel one after the other, it gives the flexibility to refer to any of these values outside of the for-loop in which they were measured.

[+ATR], N for [−ATR]and O for uncertain /ɪ ʊ/ or /e o/) and morphological context. The resulting Excel file was then ready for import into RStudio for statistical analysis.

4.3 Statistical analysis

The statistical analysis consists of two main parts: the descriptive statistics (§4.3.1), which describe the general shape and tendencies of the data without any testing, and inferential statistics (§4.3.2–4.3.3), which test specific hypotheses for significance. Both types of analysis were performed using R (v4.2.3; R Core Team 2023) in RStudio (v2023.3.0.386; Posit team 2023).

4.3.1 Descriptive statistics

The first part of the data analysis involves looking at the general shape of the data, which is particularly important in a study such as this where there are a large number of variables and tokens. To get a good overview of the data, we need to examine the distribution of measurements for each intersection of the independent variables (vowel and underlying vowel) and dependent variables (the acoustic measurements themselves). As an example, I might want to see the trends in the measurements of F1 just for the vowel /i/ across all my data. This would involve taking a subset of the total dataset of just the rows of measurements of the vowel /i/, and the column of F1 measurements. I might then want to know a number of different statistics about that data:

- Are there any outliers or anomalous datapoints?
- What is the central point of the data?
- How dispersed is the data?
- Is the data skewed in a particular direction?

The definition of outliers and anomalies is controversial. There are many schools of thought on what constitutes an outlier and when it is appropriate to exclude outliers from an analysis. Given the automatic measurement, I very much expect to find outliers in the data, though the heuristics in the Praat script should avoid the most likely offenders. Since the presence of outliers is a natural part of using data that was measured automatically, I mostly will not exclude outliers from the data unless they are clearly anomalous for recognisable reasons that could not be foreseen. There are several tools that can be employed to reduce the impact of outliers, such as data normalisation and logarithmic transformation. It is not possible to predict in advance what outliers and outlier-reduction methods will be needed without looking at the data itself. As such, I will finish the rest of my discussion on outliers while discussing the descriptive statistics in the Results chapter (§5.1.3).

The remaining statistics can all be computed simply. Given the continuous nature of all the dependent variables in this study, computing the mean or median is most appropriate to describe the central point of the data. In general, the median is more appropriate than the mean when dealing with data with significant numbers of outliers which will have a greater effect on the mean. The dispersion of the data can be shown with a number of common measures: the variance, standard deviation and interquartile range. There is also a relatively simple statistic (simply known as ‘skewness') which measures how skewed data is to the left or right.

All of these features of the data can also be seen graphically. A box plot shows the median and interquartile range, along with giving one way of interpreting outliers as those points more than 1.5 times the interquartile range away from the lower or upper quartile. It can also indicate skewness: in a skewed distribution, one of the lower or upper quartile will be closer to the median than the other. As such, the box plot is a very useful tool for initial inspection of the data. Other alternatives to the box plot include the violin plot, which shows the distribution of data as a curve, and the diamond plot, which shows the mean and a confidence interval of 95%.[6] When looking at the skew of the data, it is also useful to look at histograms, since these also give a visual indication of how normal the data is distributed, much like a violin or ridgeline plot.

4.3.2 Research questions and hypotheses

The second part of the statistical analysis involves testing hypotheses about the language in general using the data from this specific sample. In all cases, these tests rely on calculating the probability that the sample of data taken just for this study looks how it does, given certain hypotheses about the language in general (a.k.a. the statistical 'population'). It is called 'inferential statistics', because the statistics of the sample are used to 'infer' information about the wider population. It is through this testing of hypotheses that I hope to provide answers to my research questions from the first chapter (§1.2). In this section, I will restate my research questions, describe the related research hypotheses and where they came from, then in the next section, the statistical tests I will perform and how to interpret their results.

This study has 2 main research questions, given in (46).

(46) I. Do the Level 2 vowels in Boa-Leboale include just one pair of vowels or two?
II. If Boa-Leboale has just one pair of Level 2 vowels, are they /e o/ or /ɪ ʊ/?
If it has two pairs of Level 2 vowels, which pair appears in which contexts?

Research question (I) seeks to investigate if the category currently described as the Level 2 vowels, written in Chapter 3 as /ɪ/ and /ʊ/, is truly uniform. In particular, I wish to determine if the vowels transcribed as [ɪ] and [ʊ] seem to be unimodal across all acoustic measures and if there is any significant acoustic difference between /ɪ ʊ/ in derived and underived environments. The hypotheses for research question (I) are given in (47)–(48).

(47) **Hypothesis 1:** Each Level 2 vowel's distribution of acoustic measurements is...
a. H_0: ...ambiguous regarding its unimodality.
b. H_1: ...strongly unimodal or strongly multimodal.

(48) **Hypothesis 2:** For each measure, derived and underived Level 2 vowels have...
a. H_0: ...neither equivalent nor different means.

---

[6] Note that a diamond plot is not a purely descriptive graph; the computation of a confidence interval is an inferential method, since it is inferring where the mean of the population is likely to be based on the sample data, rather than simply describing a statistic of the sample data alone.

b. H_1: ...equivalent or different means.

If the Level 2 vowels do indeed comprise two sets of vowels, then we would expect at least one of the acoustic measures to have two different means (and thus be multimodal). Thus, Hypothesis 1 generally tests for unimodality or multimodality across all acoustic measures.

I also want to more specifically test the hypothesis that Boa-Leboale could be a 9-surface-vowel language on the basis of its phonology. As described in §3.3, in some morphological contexts, [ɪ ʊ] alternate with /i u/ and in others they alternate with /ɛ ɔ/. We can construct two categories of [ɪ ʊ] in the dataset: **derived** [ɪ ʊ] is any case where the original verb has underlying /ɛ ɔ/ and has undergone raising before the recent past suffix -i or the causative suffix -is, while **underived** [ɪ ʊ] is all other cases (where the underlying form is /ɪ ʊ/). Given that there are many languages with underlying 7-vowel systems with additional allophonic vowels in certain contexts, Hypothesis 2 tests for equivalence and difference between these different occurrences of [ɪ ʊ].[7]

While research question (I) focuses on *how many* vowels there are in Boa-Leboale, research question (II) instead looks at *which* vowels are present in Boa-Leboale. This question obviously depends on the results of the tests performed for research question (I). The current assumption in previous literature is that there is only one set of Level 2 vowels underlyingly, in which case, research question (II) simply looks to determine if that set of vowels is /ɪ ʊ/ or /e o/. This hypothesis has not been specifically tested before, since all prior studies have been on languages with both sets of vowels or with a clear idea of which Level 2 vowels are present. As such, I looked at the results of past studies of 9-vowel languages (particularly the results of Starwalt (2008); Olejarczuk et al. (2019)) to determine a number of trends regarding the relationship between the Level 2 vowels and their neighbours, as shown in (49).

(49) Tendencies of relative acoustic measurement values
a. For ΔB1: ε/ɔ > **ɪ/ʊ** > i/u > **e/o**
b. For A1*–A2*: **e/o** > i/u > **ɪ/ʊ** > ε/ɔ [8]
c. For centre of gravity: ε > **e** > **ɪ** > i [9]
d. For CPP: ε/ɔ > **ɪ/ʊ** > **e/o** > i/u

I did not include the measurements for F1 and F2 because there was no indication that there was a consistent cross-linguistic tendency towards a certain relationship between the vowels. There is no consistent evidence that [ɪ ʊ] and [e o] have different values for F1, and there is no consistent trend for the relative position of [i u] compared to [ɪ ʊ] for F2. I also chose to only include ΔB1 not raw B1, since the trends in previous studies were more consistent for ΔB1.

---

[7] Note that this assumes that Boa-Leboale is underlyingly a 7-vowel language to begin with. It would also be worthwhile to test for any acoustic differences between vowels which lower to [ε] before suffixes containing /ɪ/ and those which do not. This way, one could determine if the variation in the application of that rule comes form an underlying phonemic or morphological difference. Unfortunately, the scope of this thesis is too limited to include this test as well.

[8] This is the trend found by Starwalt (2008), but not Guion et al. (2004). I follow Starwalt's results because they come from the analysis of multiple languages.

[9] There is no recognisable trend for the back vowels, since there is no consistent difference previously reported for CoG for the Level 2 back vowels.

These are all just trends that the data tended to follow, rather than being statistically significant hierarchies of values. As such, they should be taken with a heavy pinch of statistical salt: while I will form testable hypotheses from these trends and infer information about the underlying vowel system of Boa-Leboale, each of these trends is not, in and of itself, a reliable indicator of ATR class in a given language. The purpose of analysing these trends with statistical methods is twofold: (1) to provide a new set of testable hypotheses that can be applied more rigorously to different languages in future research and (2) to give multiple independent indicators of possible ATR class for the vowels in Boa-Leboale. While each test by itself is not necessarily a surefire method of determining the identity of a vowel in Boa-Leboale, if the combination of these tests gives the same outcome each time, then this provides solid (if not irrefutable) evidence to support a given conclusion. By applying many of these tests, the chances of an individual false positive increases, but the chances of *all* tests agreeing in which direction they fall decreases.

These trends can be systematised into hypotheses in two ways. Note that I use Lx as shorthand to indicate Level x vowels, tested both on the front and back vowels, e.g. if a hypothesis involves comparing L1 and L2, then one test compares [i] and [ɪ~e] and another compares [u] and [ʊ~o].

For the first two trends involving ΔB1 and A1*–A2*, the values for [ɪ ʊ] and [e o] tend to be on either side of the values for [i u]. As such, if the mean values for the Level 2 vowels are statistically different from the mean Level 1 values, then we can use their position relative to the Level 1 vowels as an indicator of which vowel they are likely to be. Take for instance an unknown Level 2 back vowel [B], and we want to know if it is [ʊ] or [o]. If the mean value of ΔB1(B) is significantly different from the mean value for ΔB1(u), then if mean ΔB1(B) > mean ΔB1(u), B is likely to be [ʊ] (since the trend is that ʊ > u). If mean ΔB1(B) < mean ΔB1(u), B is likely to be [o] (since the trend is that u > o). The hypotheses, along with their interpretations, are given in (50)–(51).

(50) **Hypothesis 3:**
  a. H_0: $\Delta B1(L1) = \Delta B1(L2)$
  b. H_1: $\Delta B1(L1) \neq \Delta B1(L2)$
    i. *Case*: $\Delta B1(L1) > \Delta B1(L2)$, *Interpretation*: L2 is [+ATR]
    ii. *Case*: $\Delta B1(L1) < \Delta B1(L2)$, *Interpretation*: L2 is [–ATR]

(51) **Hypothesis 4:**
  a. H_0: $A1^*–A2^*(L1) = A1^*–A2^*(L2)$
  b. H_1: $A1^*–A2^*(L1) \neq A1^*–A2^*(L2)$
    i. *Case*: $A1^*–A2^*(L1) > A1^*–A2^*(L2)$, *Interpretation*: L2 is [–ATR]
    ii. *Case*: $A1^*–A2^*(L1) < A1^*–A2^*(L2)$, *Interpretation*: L2 is [+ATR]

For the last two trends involving centre of gravity and CPP, this approach does not work, because there is no intermediate vowel that can be compared against. Instead, we can compare the distances between the means of the Level 2 vowels and the means of the Level 1 and 3 vowels. The idea is that whichever external vowel the Level 2 vowel is closer to might indicate which vowel it is. Take for instance an unknown Level 2 front vowel [F], and we want to know if it is [ɪ] or [e]. If the mean value for CPP is closer to the mean value for [i] than for [ɛ], then it is more likely to be [e], according to the tendencies given. These tests are particularly weak, since it is very possible that both [e] and [ɪ] might be closer to [i] than to [ɛ], for example. Then this test would mistakenly suggest that both vowels are [e]. As such, I propose this test *only* alongside other tests for the sake of seeing to what extent they line up with one another, and with the caveat that any results are likely to be unreliable taken by themselves.

These hypotheses are given in (52)–(53), along with another hypothesis (54), taken from my interpretation of Starwalt's (2008) comments on distinguishing the Level 2 vowels (see §2.4.3).

(52) **Hypothesis 5:**

a. H_0: $\text{CoG}(\text{ɛ}) - \text{CoG}(\text{ɪ}) = \text{CoG}(\text{ɪ}) - \text{CoG}(\text{i})$

b. H_1: $\text{CoG}(\text{ɛ}) - \text{CoG}(\text{ɪ}) \neq \text{CoG}(\text{ɪ}) - \text{CoG}(\text{i})$

   i. *Case*: $\text{CoG}(\text{ɛ}) - \text{CoG}(\text{ɪ}) > \text{CoG}(\text{ɪ}) - \text{CoG}(\text{i})$, *Interpretation*: L2 is [+ATR]

   ii. *Case*: $\text{CoG}(\text{ɛ}) - \text{CoG}(\text{ɪ}) < \text{CoG}(\text{ɪ}) - \text{CoG}(\text{i})$, *Interpretation*: L2 is [–ATR]

(53) **Hypothesis 6:**

a. H_0: $\text{CPP}(\text{L3}) - \text{CPP}(\text{L2}) = \text{CPP}(\text{L2}) - \text{CPP}(\text{L1})$

b. H_1: $\text{CPP}(\text{L3}) - \text{CPP}(\text{L2}) \neq \text{CPP}(\text{L2}) - \text{CPP}(\text{L1})$

   i. *Case*: $\text{CPP}(\text{L3}) - \text{CPP}(\text{L2}) > \text{CPP}(\text{L2}) - \text{CPP}(\text{L1})$, *Interpretation*: L2 is [–ATR]

   ii. *Case*: $\text{CPP}(\text{L3}) - \text{CPP}(\text{L2}) < \text{CPP}(\text{L2}) - \text{CPP}(\text{L1})$, *Interpretation*: L2 is [+ATR]

(54) **Hypothesis 7:**

a. H_0: $\Delta\text{B1}(\text{ɛ}) \leq \Delta\text{B1}(\text{ʊ})$

b. H_1: $\Delta\text{B1}(\text{ɛ}) > \Delta\text{B1}(\text{ʊ})$

   i. *Case*: $\Delta\text{B1}(\text{ɛ}) > \Delta\text{B1}(\text{ʊ})$, *Interpretation*: L2 is [+ATR]

In the next section I will outline all the statistical tests I will perform to test these 7 hypotheses.

4.3.3 Significance tests

In general, there are 3 different types of statistical test used to test these hypotheses. These are the folding test for unimodality, equivalence testing for equivalence of means and a generalised linear hypothesis testing for difference of means. All statistical tests will be performed in R using RStudio (R Core Team 2023; Posit team 2023).

4.3.3.1 Hypothesis 1

For Hypothesis 1, I will test for unimodality or multimodality with the folding test (Siffer et al. 2018). The intuition behind the folding test is that multimodal data has a pivot point where, if the data is reflected in it, the total variance of the data goes down, while for unimodal data, reflection at any pivot point would increase the total variance. This test was chosen because it has unimodality as one of the alternative hypotheses rather than the null hypothesis, and because it can be applied to multivariate data. The folding test statistic Φ can be interpreted depending on its value relative to

1. If $\Phi \geq 1$, then the distribution is unimodal, while if $\Phi < 1$, then the distribution is multimodal. I will conduct the test with the R package *rfolding*. There is currently no obvious way to determine p-values for a given number of observations and dimensions[10], but Siffer et al. (2018: 6) provides a table of quantiles based on Monte-Carlo simulations at a significance level of $p = 0.05$. I will use this table to interpret the significance of the folding test.

(55) gives the testing hypothesis for the folding statistic Φ, which will be calculated separately for each of the Level 2 vowels. The interpretation table only has a maximum of 5 dimensions, so I will include only F1, F2, ΔB1, A1*–A2*, centre of gravity and average CPP to reduce the number of dimensions to just 6 (instead of 8).

(55) Hypothesis 1:
  a. H_0: $\Phi = 1$
  b. H_1: $\Phi < 1$ (data is multimodal) or $\Phi > 1$ (data is unimodal)

4.3.3.2 Hypothesis 2

For Hypothesis 2, I will test both the equivalence and difference of the means of most measurements for the derived and underived Level 2 vowels. In order to test equivalence, I need to use an equivalence test, as outlined in Lakens et al. (2018). Since statistical tests cannot ever categorically prove that the means of two variables are entirely identical (since even the smallest of differences *could* be shown to be significant with sufficient data), equivalence testing involves instead testing whether the difference of means is smaller than a smallest effect size of interest (SESOI). If the effect is significantly smaller than the SESOI, then the variables are said to be equivalent to one another. Equivalence is therefore not entirely complementary to difference: two variables could have a significant difference between them, but this difference could be small enough that they are still considered equivalent by an equivalence test.

Lakens et al. (2018) discuss a number of justifications for how to choose the SESOI. One possible objective justification for the SESOI would be a just-noticeable difference (JND), when the fact that the difference is perceived by a person is relevant. Alternatively, a subjective justification for the SESOI could be based on the critical effect sizes from an earlier study – this way the SESOI would be the smallest effect size which would be considered statistically significant in that earlier study.

Choosing the SESOI on the basis of JNDs and critical effect sizes both have benefits and drawbacks. One way in which JNDs are more appropriate than critical effect sizes is that JNDs rely on raw data, while critical effect sizes are dependent on the standard deviation of the sample, and thus can be affected by different study designs (Baguley 2009: 607). Though this study took inspiration from Starwalt (2008), it differs in many important aspect; particularly, that the measurements are taken automatically with a script instead of manually, which could increase the standard deviation. However, JNDs are also prone to the context of the sounds listeners are listening to; Ghitza & Goldstein (1986) found that JNDs for formant frequency were up to 4 times as large for natural connected speech than for unnatural stationary speech. The other problem is that JNDs are not reported for

---

[10] The *rfolding* package gives an output for the p-value, but this value is clearly erroneous since it is always the same regardless of the input.

all of the measurements in this study (e.g. with centre of gravity for vowels), especially those measurements which are normalised with reference to a model (ΔB1 and A1*–A2*). This last problem is the most troublesome of all: if JNDs were known for all these acoustic correlates, then that would be the best option, but in their absence, the critical effect size at least applies for all measures, even if it is not necessarily comparable between studies.

Starwalt (2008) took 20 measurements per vowel per speaker, with a significance level of 0.01 (having applied Bonferroni's correction to the original significance level of 0.05). Given this, the critical value for F is 7.35, which can be transformed into a Cohen's *d* of 0.880, using the methodology of Thalheimer & Cook (2002) for calculating Cohen's *d* from published research. This is shown in (56), where n_1 and n_2 are equal to 20.

(56) $d = \sqrt{F\left(\frac{n_1+n_2}{n_1 n_2}\right)\left(\frac{n_1+n_2}{n_1+n_2-2}\right)}$

This typically corresponds to a large effect size (greater than 0.8), which suggests that Starwalt's study design only found effects significant if they were relatively large, which makes sense given the low significance level. Conversely, an equivalence test with the SESOI set to such a large effect size is more likely to find significant equivalence, since any smaller effects would be deemed equivalent.

The equivalence tests will be run to compare the derived and underived Level 2 vowels using the R package *TOSTER*, which runs two one-sided t-tests to test if the means of the two samples are equivalent at the given effect size boundaries (Δ_U and Δ_L are the upper and lower bounds for a t-test based on the given effect size), with the larger of the two p-values determining whether the test is significant or not. It also runs a two-sided t-test to determine if the means of the samples are significantly different from one another. Hypothesis 2 can therefore be split into two statistical tests, with different hypotheses for equivalence and difference (57).

(57) a. Hypothesis 2a (equivalence):
    i. H_0: $t > \Delta_U$ or $t < \Delta_L$
    ii. H_1: $\Delta_L \leq t \leq \Delta_U$
  b. Hypothesis 2b (difference):
    i. H_0: $t = 0$
    ii. H_1: $t \neq 0$

4.3.3.3 Hypotheses 3–6

For Hypotheses 3 and 4, I will test the difference of means of the Level 1 and Level 2 vowels for ΔB1 and A1*–A2*. In both cases, the null hypothesis is that their difference equals 0, and the alternative hypothesis is that their difference does not equal 0, with the interpretation depending on whether the difference is positive or negative (i.e. which value is higher than the other).

For Hypotheses 5 and 6, I will instead test a linear combination of the means of the Level 1, 2 and 3 vowels for centre of gravity and CPP. In both cases, the linear combination is $L1 + L3 - 2 \cdot L2$, which is equivalent to the expressions given earlier (§4.3.2). The null hypothesis is that this value equals 0, and the alternative hypothesis is that it does not equal0, with the interpretation once again depending on whether the difference is positive or negative (i.e. which other level the Level 2 vowels are closer to).

To test both types of hypotheses, I will model the data with multivariate generalised linear models (GLM), since they tend to be more robust with violated assumptions compared to, say, ANOVA. In particular, given how different vowels are more prone to measurement errors than others, it is likely that there will not be equal variances between different vowels for several measurements, especially because the measurements are taken automatically. By contrast, Starwalt (2008) performed her statistical tests with ANOVA modelling.

I will then use the R package *multcomp* to perform post-hoc testing of the contrasts of interest (see Bretz et al. (2016: 53–67) for more details about how this package functions). I have constructed a contrast matrix corresponding to the linear combinations of variables for Hypotheses 3–6, so only the required tests are performed to avoid performing a number of unnecessary tests which would have to be corrected for to avoid the issues arising from multiple testing.

The test hypotheses are given in (58)–(61).

(58) **Hypothesis 3:** $D_B = \Delta\mathrm{B1(L1)} - \Delta\mathrm{B1(L2)}$
  a. H$_0$: $D_B = 0$
  b. H$_1$: $D_B < 0$ (L2 is [–ATR]) or $D_B > 0$ (L2 is [+ATR])

(59) **Hypothesis 4:** $D_{SF} = \mathrm{A1^*–A2^*(L1)} - \mathrm{A1^*–A2^*(L2)}$
  a. H$_0$: $D_{SF} = 0$
  b. H$_1$: $D_{SF} < 0$ (L2 is [+ATR]) or $D_{SF} > 0$ (L2 is [–ATR])

(60) **Hypothesis 5:** $D_{CoG} = \mathrm{CoG(i)} + \mathrm{CoG(ɛ)} - 2\mathrm{CoG(ɪ)}$
  a. H$_0$: $D_{CoG} = 0$
  b. H$_1$: $D_{CoG} < 0$ (L2 is [–ATR]) or $D_{SF} > 0$ (L2 is [+ATR])

(61) **Hypothesis 6:** $D_{CPP} = \mathrm{CPP(L1)} + \mathrm{CPP(L3)} - 2\mathrm{CPP(L2)}$
  a. H$_0$: $D_{CPP} = 0$
  b. H$_1$: $D_{CPP} < 0$ (L2 is [+ATR]) or $D_{SF} > 0$ (L2 is [–ATR])

4.3.3.4 Hypothesis 7

Finally, for Hypothesis 7, I am testing if the mean ΔB1 for /ɛ/ is greater than for /ʊ/. Unlike Hypotheses 3–6, this is a one-tailed test, which can also be performed using the R package *multcomp* to do a post-hoc test on the GLM model. The only difference is that the alternative hypothesis H_1 is one-sided ($>$) instead of two-sided ($\neq$), as shown in (62).

(62) **Hypothesis 7:** $D = \Delta\mathrm{B1(ɛ)} - \Delta\mathrm{B1(ʊ)}$
  a. H$_0$: $D \leq 0$
  b. H$_1$: $D > 0$ (L2 is [+ATR])

Chapter 5

Results

In this chapter, I present the results of the study described in Chapter 4. First, I will describe the general trends and patterns (e.g. mean, variance) on the data for each vowel/measure pair (§5.1) and how I deal with outliers and skewed distributions. I will then go through each of the hypotheses described in the previous chapter (§4.3.2) one by one, giving the outcome of all of the statistical tests performed, along with any notes on how to interpret the results. After giving the results for all hypotheses associated with a given research question, I will discuss how the results answer (or fail to answer) the research question. At the end of the chapter, I will give an overview of all the results (§5.4).

5.1 Descriptive statistics

5.1.1 Vowel chart

Before presenting the data for each acoustic measure, Figure 5.1 shows the vowel chart for F1 and F2 for all vowels. Some trends are visible just from this chart: tokens of Level 1 vowels /[i u]/ are relatively dispersed along the F2 axis, while the other vowels tend to be more spread out along the F1 axis. There is also a reasonable degree of overlap between neighbouring vowels, though the vowels do form noticeable clusters, usually close to their mean values. The differences in frequency between vowels can also be seen clearly: /[i]/ and /[ɪ]/ clearly have the most tokens, while /[a]/ has the fewest. The next section examines each acoustic measure in more detail by vowel.

5.1.2 Breakdown by acoustic measure

For each acoustic measure, I provide a table with the mean, standard deviation, minimum, median, maximum and skewness of each vowel for that measure. The skewness of each distribution is calculated with the skewness function from the R package *moments*, which returns the third standardised moment.[1] I also provide diamond plots, which show the distribution of all the measurements taken for each vowel, along with a diamond which shows the mean and an approximate 95% confidence

---

[1] This measure of skewness is a natural extension of the methods used to calculate mean and variance. The first (raw) moment of a distribution is its mean, then the second moment of a distribution – once centralised by mean – is its variance, so the third moment of distribution – once standardised for variance – is its skewness.

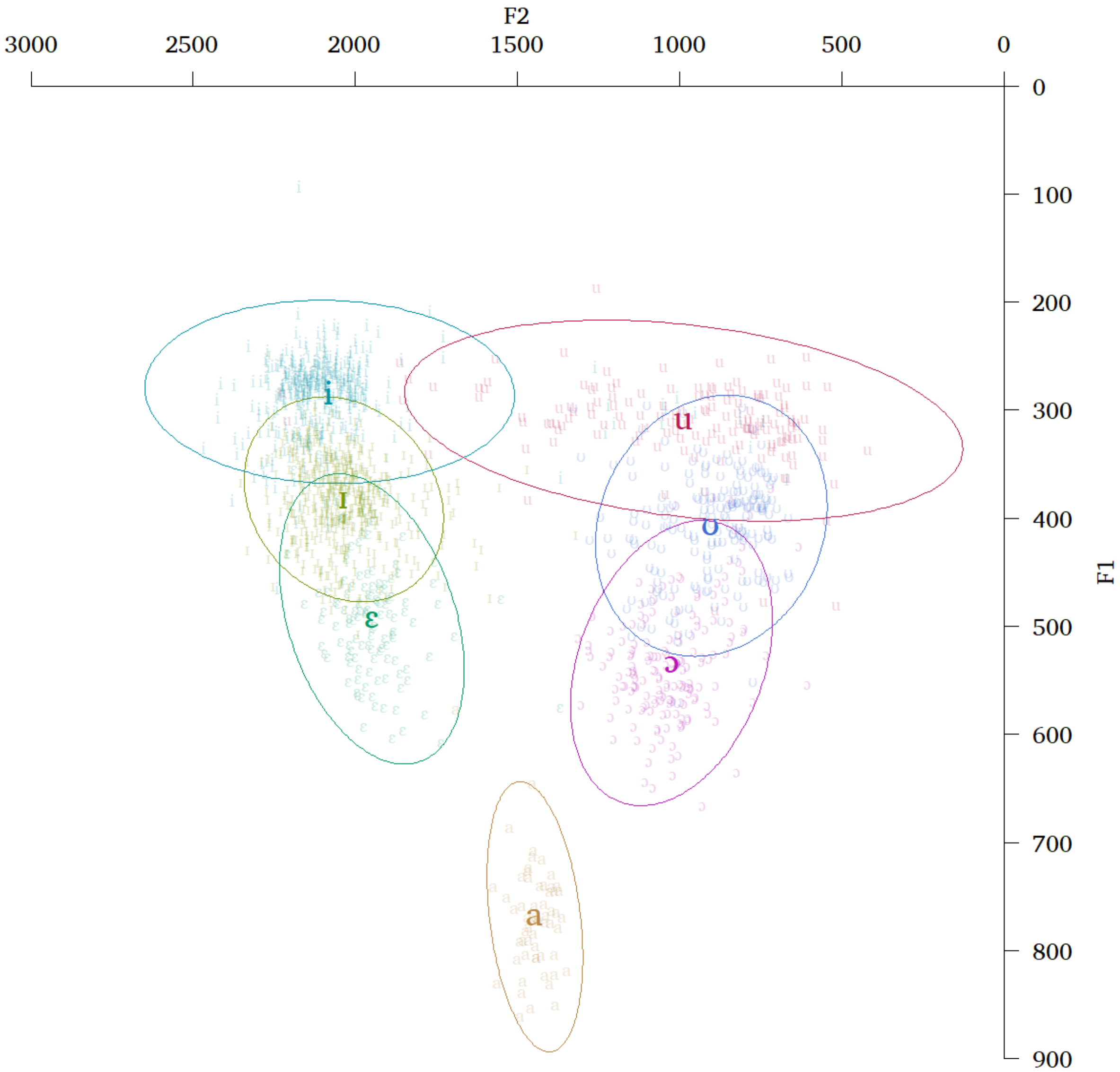


Figure 5.1: Vowel chart plotting F1 and F2 for all tokens of each vowel, along with means and ellipses with a confidence interval of 95%

interval for the mean.

5.1.2.1 F1

The descriptive statistics of the F1 measurements (in Hz) for each vowel, across all morphological groups, are given in Table 5.1. They can also be visualised in the diamond plot in Figure 5.2. As expected, the mean corresponds closely to the height/level of the vowel: Level 1 vowels have the lowest F1 values on average, while the highest mean is the Level 4 vowel /[a]/. There are also no clear anomalies in terms of the minima or maxima for any of the vowels, except for /[i]/ 92Hz.

In terms of skewness, there is a trend for lower values of F1 to be skewed positively, while higher values for F1 are skewed negatively. This likely results from the automatic measurement process employed, which will tend to produce erroneous measurements that fall towards the midrange of

Table 5.1: Descriptive statistics of F1 (Hz) measurements for each vowel

Vowel	Mean	StDev	Min	Median	Max	Skew
i	283	34	92	277	463	0.65
ɪ	382	39	248	380	507	0.29
ɛ	493	54	322	491	608	−0.57
a	768	49	576	769	861	−0.99
ɔ	534	53	372	538	666	−0.59
ʊ	407	49	295	398	570	0.43
u	310	37	186	309	485	1.71

F1 rather than measurements at each extreme.

5.1.2.2 F2

The descriptive statistics of the F2 measurements (in Hz) for each vowel, across all morphological groups, are given in Table 5.2. They can also be visualised in the diamond plot in Figure 5.3. As expected, mean F2 corresponds with the frontness of the vowel, with the front vowels having mean values around 2000Hz, back vowels 1000Hz and /[a]/ in the middle (approx. 1500Hz). The standard deviations are roughly on par with those for F1 in that they are all around 10% of the value itself. The major exceptions to this are /[i, a, u]/. /[a]/ has a low standard deviation presumably by virtue of the lower number of tokens (which can be seen visually in Figure 5.3 as /[a]/ having fewer dots), while /[i, u]/ also have a great deal of anomalous points, as can be clearly seen with their minimum and maximum values respectively, which are considerably out of the expected range, and their skewness values.

Table 5.2: Descriptive statistics of F2 (Hz) measurements for each vowel

Vowel	Mean	StDev	Min	Median	Max	Skew
i	2,080	231	670	2,114	2,467	−3.66
ɪ	2,036	125	1,318	2,045	2,351	−1.11
ɛ	1,950	114	1,370	1,953	2,210	−1.32
a	1,447	58	1,348	1,445	1,689	1.31
ɔ	1,026	126	607	1,028	1,312	−0.43
ʊ	904	144	664	879	1,327	0.63
u	989	347	423	900	2,225	1.11

In general, the measurements for F2 are, like with F1, skewed towards the central values, so front vowels have negative skew and back vowels positive skew. Regardless of this skew, most vowels have a clear concentration of tokens at one point except for /[u]/, which has tokens spread evenly across the range. This contributes to /[u]/'s unexpectedly high standard deviation. This presumably comes from the difficulties of measuring F2, as described earlier in the acoustic methodology (§4.2.1.3). For rounded back vowels like /[u]/, the F1 and F2 values are very close to one another, so can easily be confused for one larger formant instead of two smaller ones. This means that Praat's formant tracker, which does not have any information about which kind of vowel it is measuring, often misidentifies

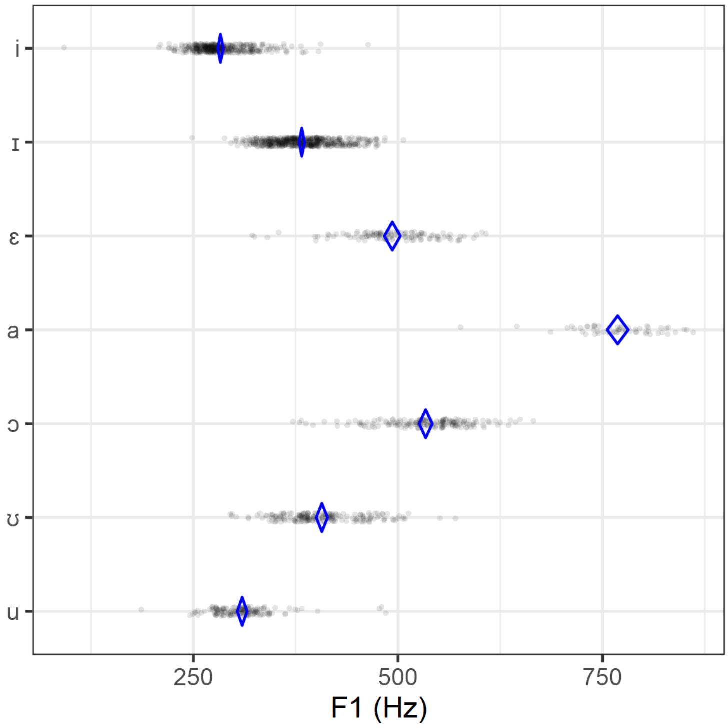


Figure 5.2: Diamond plot showing F1 measurements for each vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

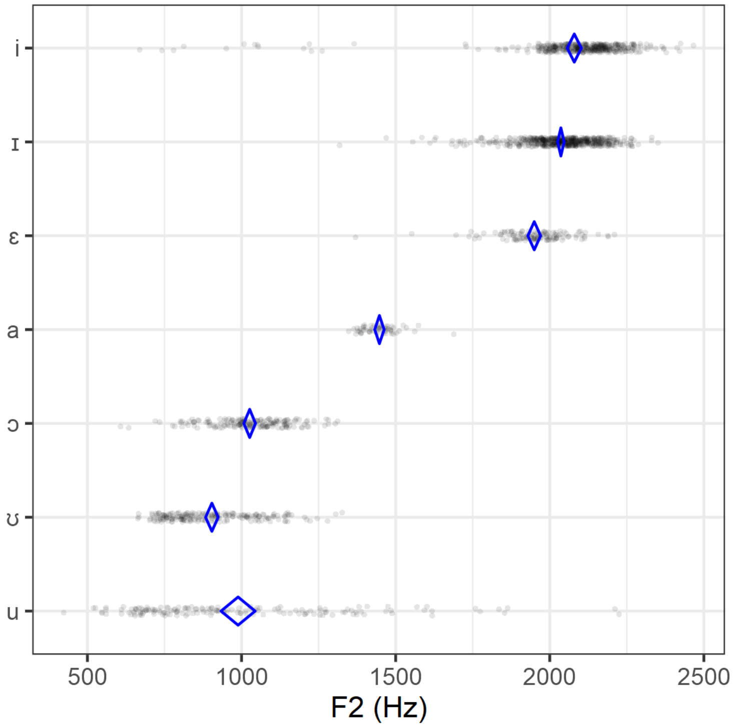


Figure 5.3: Diamond plot showing F2 measurements for each vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

the value of F2 for these values (either measuring a phantom formant or F3), and also gives a higher value for F1 and B1 since it believes there is one wider formant at a middle frequency instead of two narrower formants, one high and one low.

5.1.2.3 B1, ΔB1

The descriptive statistics of the B1 measurements (in Hz) for each vowel, across all morphological groups, are given in Table 5.3. They can also be visualised in the diamond plot in Figure 5.4. The mean and median values are lowest for /[ɪ, ɛ]/ and highest for /[u]/. As described for F2, /[u]/ has a much higher standard deviation than the other vowels because of the inherent difficulties in accurately identifying F1 and F2 separately. In many cases, the automatic measurement incorrectly identified both formants as a single formant with a higher bandwidth, which leads to the unexpectedly high mean value and standard deviation of /[u]/.

Table 5.3: Descriptive statistics of B1 (Hz) measurements for each vowel

Vowel	Mean	StDev	Min	Median	Max	Skew
i	59.9	54.6	8.8	42.4	431.3	3.06
ɪ	41.7	23.7	7.1	36.9	184.6	2.17
ɛ	42.1	21.9	15.4	34.2	126.2	1.45
a	99.1	45.2	36.7	93.2	247.7	1.14
ɔ	93.1	53.4	26.1	80.2	317.7	1.79
ʊ	90.0	51.7	25.1	77.9	444.8	2.72
u	134.2	82.7	24.8	115.3	489.7	1.63

It is also very clear that all measurements are strongly skewed in the positive direction. It seems as though the automatic measurement as implemented here is prone to measuring high values for F1 bandwidth. Since all vowels are skewed in the same direction, it is possible to apply a transformation to the data to make it more normal before testing (see §5.1.3), which would reduce the impact of the skewed distribution of data on the results.

The descriptive statistics of the ΔB1 values (in Hz) for each vowel, across all morphological groups, are given in Table 5.4. They can also be visualised in the diamond plot in Figure 5.5. The trends are very similar to those for B1, though the vowel with the lowest mean and median is now /[i]/. The impact of the skewed B1 measurements is also very clear here, with /[i]/ having a median value of −23.9Hz and a maximum value of 366.2Hz. As such, a similar transformation as to B1 can be performed on the ΔB1 data to make it more normal and reduce the impact of outliers on the outcome of the statistical tests (see §5.1.3).

5.1.2.4 A1*–A2*

The descriptive statistics of the A1*–A2* values (in dB) for each vowel, across all morphological groups, are given in Table 5.5. They can also be visualised in the diamond plot in Figure 5.6. There is a general trend that higher level vowels have higher A1*–A2* than lower level vowels: indeed the lowest mean and median value is Level 4 /[a]/ and the highest is Level 1 /[i]/. This is complicated by a few factors: the back vowels have lower measurements than their front counterparts, so

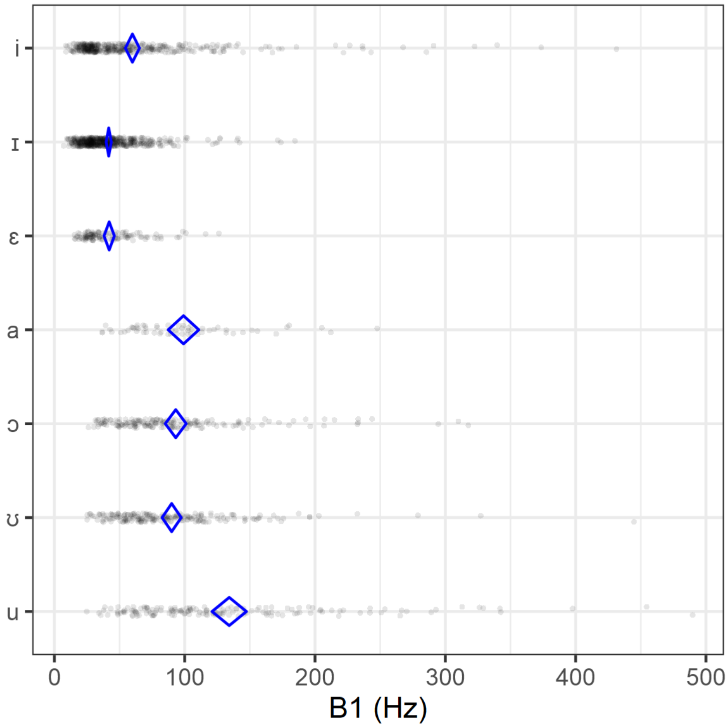


Figure 5.4: Diamond plot showing B1 measurements for each vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

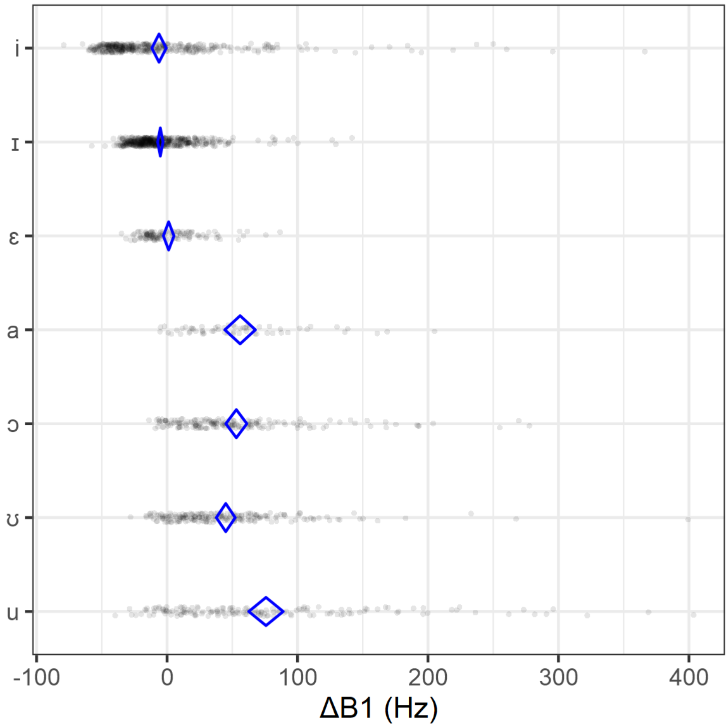


Figure 5.5: Diamond plot showing ΔB1 measurements for each vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

Table 5.4: Descriptive statistics of ΔB1 (Hz) values for each vowel

Vowel	Mean	StDev	Min	Median	Max	Skew
i	−6.1	55.6	−79.2	−23.9	366.2	2.82
ɪ	−5.2	24.3	−57.7	−10.6	141.9	2.09
ɛ	1.2	22.4	−35.0	−6.0	86.6	1.36
a	56.0	45.0	−5.4	50.6	205.1	1.17
ɔ	53.1	53.5	−13.9	39.4	277.7	1.78
ʊ	45.0	51.0	−27.9	33.7	399.3	2.72
u	75.8	82.6	−39.8	55.1	403.5	1.45

this measurement does not directly map onto vowel height/level across different vowel backnesses. Also, mean /[ɪ]/ is higher than mean /[ɛ]/, but the inverse is true for their medians. The minima and maxima generally follow this trend too, and there are no clear anomalous points. However, the standard deviations are quite high across the board, showing that the measurements are highly dispersed. Once again, the highest standard deviation is found for /[u]/, presumably due to the difficulty of accurately measuring its F1 and F2. The data are not very skewed, though as before the most skewed vowel is /[u]/. Where there is skew, it tends to be positive, though /[a]/ and /[ʊ]/ are both very weakly skewed negatively (though with values of –0.19 and –0.18, they both have effectively symmetrical distributions).

Table 5.5: Descriptive statistics of A1*–A2* (dB) values for each vowel

Vowel	Mean	StDev	Min	Median	Max	Skew
i	4.94	5.40	−5.69	3.99	33.36	1.19
ɪ	−0.54	4.03	−10.08	−1.06	17.16	0.86
ɛ	−1.04	4.13	−14.27	−0.81	18.45	0.75
a	−7.02	2.96	−14.41	−6.85	−0.55	−0.19
ɔ	−5.42	3.23	−13.06	−5.38	9.07	0.60
ʊ	−3.94	3.99	−13.98	−3.45	5.26	−0.18
u	0.24	7.10	−10.99	−1.12	29.20	1.45

5.1.2.5 Centre of gravity

The descriptive statistics of the centre of gravity measurements (in Hz) for each vowel, across all morphological groups, are given in Table 5.6. They can also be visualised in the diamond plot in Figure 5.7. Like A1*–A2*, the mean and median measurements correspond closely to the height/level of the vowel, in part because the centre of gravity tends to hover near or just above F1, so /[a]/ has the highest mean and median CoG and /[u]/ the lowest. CoG is also influenced by F2, so back vowels all have lower CoG than their front counterparts.

Visually it is clear that /[ɛ]/ has a particularly dispersed distribution compared to the other vowels, which is reflected in its high standard deviation (194Hz). It is not clear to me why this might be the case, especially since the other vowels seem relatively bunched. Given that CoG is not dependent on formant tracking like the other measurements discussed so far, any outliers or skew cannot be

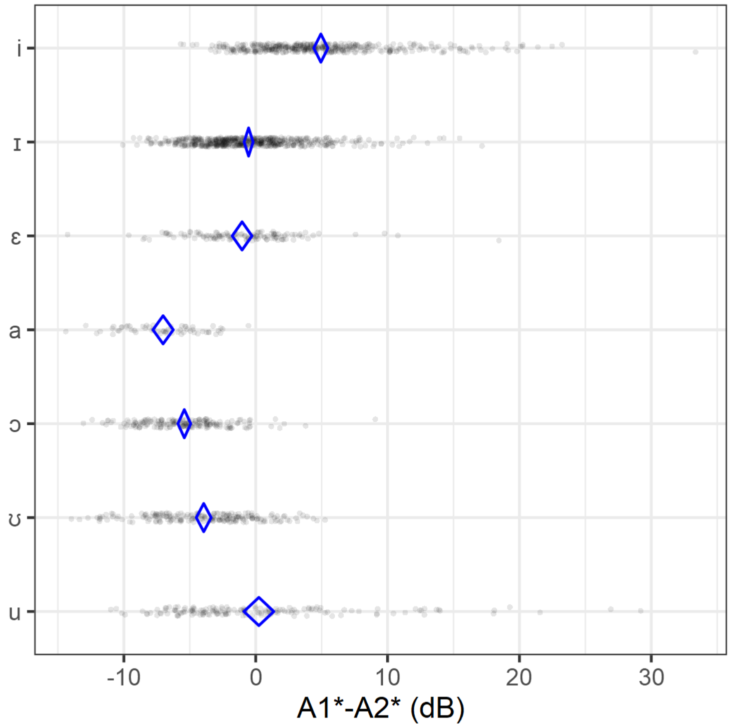


Figure 5.6: Diamond plot showing A1*–A2* measurements for each vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

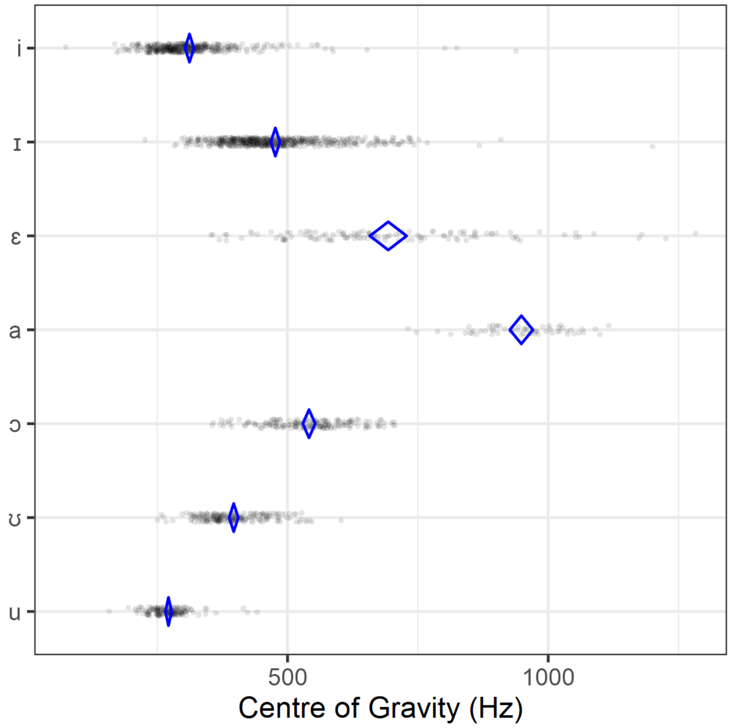


Figure 5.7: Diamond plot showing centre of gravity measurements for each vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

Table 5.6: Descriptive statistics of centre of gravity (Hz) measurements for each vowel

Vowel	Mean	StDev	Min	Median	Max	Skew
i	312	86	74	297	938	2.66
ɪ	477	104	226	458	1,200	1.37
ɛ	693	194	353	658	1,283	0.68
a	949	85	731	954	1,117	−0.31
ɔ	541	79	355	547	706	−0.14
ʊ	397	61	251	388	603	0.51
u	272	37	158	272	442	1.27

explained through inadequacies of the automatic measurement.

Indeed, there are some measurements, particularly for /[i]/ and /[ɛ]/, that are noticeably far away from the mean. Early explorations of the data found even more anomalous readings for /[i]/, several as high as 3000–4000Hz, which is well outside of the norm for a vowel. However, on further inspection, all of these cases came not from a mismeasurement, but a misannotation: they were all causative verbs (with the suffix `-is`) with verb roots ending in /[t]/ or /[s]/. The front high vowel /[i]/ between these consonants was difficult to identify, and was indeed pronounced with some degree of sibilance on account of the upcoming /[s]/. Sibilant fricatives are notable for having particularly high centres of gravity compared to other fricatives, never mind compared to vowels. In the end, I was able to relabel the vowels so that the annotation covered only the portion with periodic behaviour, where the vocalic qualities would outweigh the fricative ones. Even after this adjustment, the tokens for /[i]/ above 750Hz all come from such verbs, so it is clear that there was still some anticipatory sibilance during the vowel which gave rise to the anomalous measurements.

These unexpectedly high measurements contribute to the high skewness for /[i]/ compared to the other vowels. These outliers can be mitigated through transformation, like for B1 and ΔB1, though this would make the distributions for /[a]/ and /[ɔ]/ more negatively skewed (see §5.1.3 for further discussion).

5.1.2.6 CPP

The descriptive statistics of the unfiltered CPP measurements (in dB) for each vowel, across all morphological groups, are given in Table 5.7. They can also be visualised in the diamond plot in Figure 5.8. The means and medians are approximately related to the height/level of the vowels: the highest mean/median is the Level 4 vowel /[a]/, the lowest is Level 1 /[i]/. This indicates that the lower level vowels have greater prominence of the cepstral peak, which can be interpreted as the vowel being more modally voiced.

All vowels have very dispersed distributions. While there are different central tendencies for each vowel, they all have a wide range of measured results. This can be seen graphically in Figure 5.8 in how the tokens are spread along the entire horizontal axis for almost all vowels, and in Table 5.7 by the high standard deviations compared to the differences between the means and the fact that all vowels have similarly low minima and similarly high maxima. All vowels have unskewed

Table 5.7: Descriptive statistics of unfiltered CPP (dB) measurements for each vowel

Vowel	Mean	StDev	Min	Median	Max	Skew
i	12.50	3.23	4.01	12.29	26.29	0.49
ɪ	14.94	4.46	5.02	14.22	28.81	0.48
ɛ	17.25	4.80	5.87	16.71	27.76	0.20
a	19.74	5.83	6.55	20.36	30.18	−0.21
ɔ	15.81	4.30	7.50	15.50	28.88	0.47
ʊ	15.43	4.22	4.21	14.88	27.69	0.29
u	14.28	3.82	5.44	13.76	25.66	0.49

distributions, with skewness under an absolute value of 0.5 in all cases.

The descriptive statistics of the averaged CPP measurements (in dB) for each vowel, across all morphological groups, are given in Table 5.8. They can also be visualised in the diamond plot in Figure 5.9. The results are similar to those for the unfiltered CPP, though the distributions are somewhat less dispersed; the standard deviations are lower and the minima/maxima are less extreme. The data are very slightly more positively skewed, and might also benefit from transformation to reduce skewness (see discussion in §5.1.3).

Table 5.8: Descriptive statistics of averaged CPP (dB) measurements for each vowel

Vowel	Mean	StDev	Min	Median	Max	Skew
i	12.41	2.37	5.45	12.38	21.58	0.37
ɪ	13.42	2.41	7.03	13.30	23.92	0.80
ɛ	14.47	2.68	7.81	13.91	22.31	0.89
a	15.95	3.59	10.09	15.27	24.93	0.75
ɔ	13.88	2.53	7.24	13.51	22.22	0.82
ʊ	13.92	2.61	7.40	13.69	21.86	0.47
u	13.94	2.88	7.02	13.40	21.83	0.51

5.1.3 Skewness, outliers and transformation

As discussed in the previous section, there are several measures where the results are significantly skewed in a particular direction for all vowels. Many of the statistical tests involved in testing the hypotheses to answer the research questions of this study assume that the data is normal, or at least close enough to normality to not have a large impact on the results. Thus in those cases where there is a significant skew, one way to make the data more normal is to apply a transformation to the data.

In this case, there are three cases where I think it would be appropriate to apply a transformation to the data: B1, ΔB1 and centre of gravity. For B1, the data is very positively skewed, so this can be adjusted for by considering the logarithm of the measurement instead of the measurement itself. This would also reduce the impact of high outliers, since under logarithmic transformation, these are transformed to less outlying values.

Consider the violin plots in Figure 5.10, showing the shape of the distribution of the B1 measure-

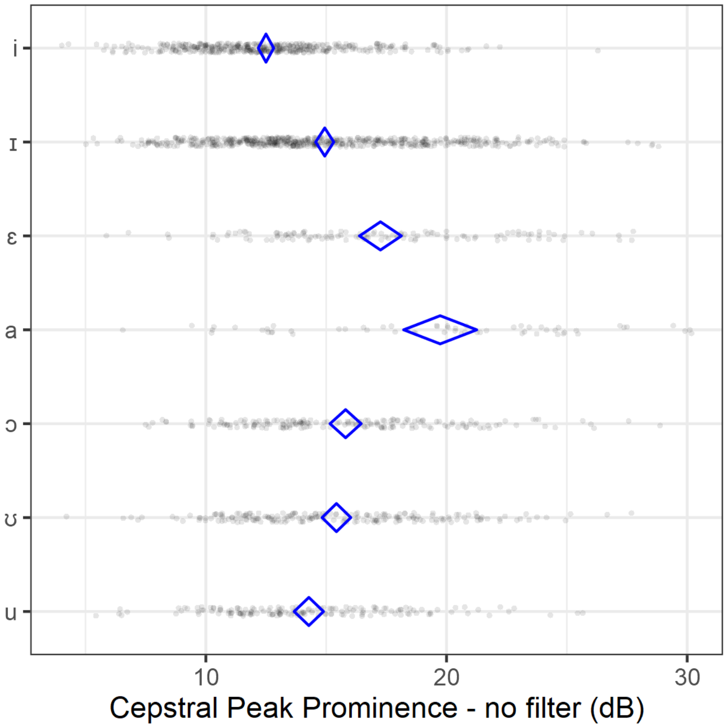


Figure 5.8: Diamond plot showing unfiltered CPP measurements for each vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

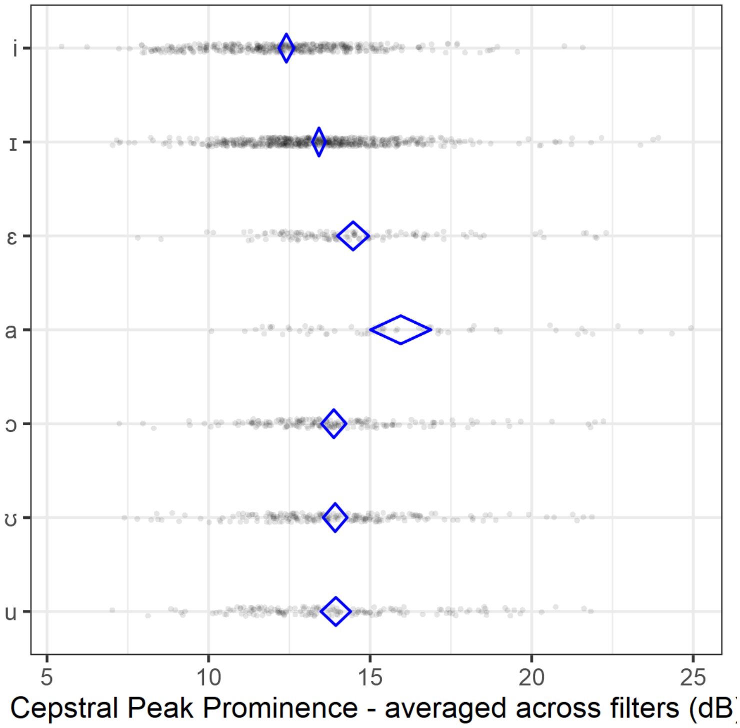
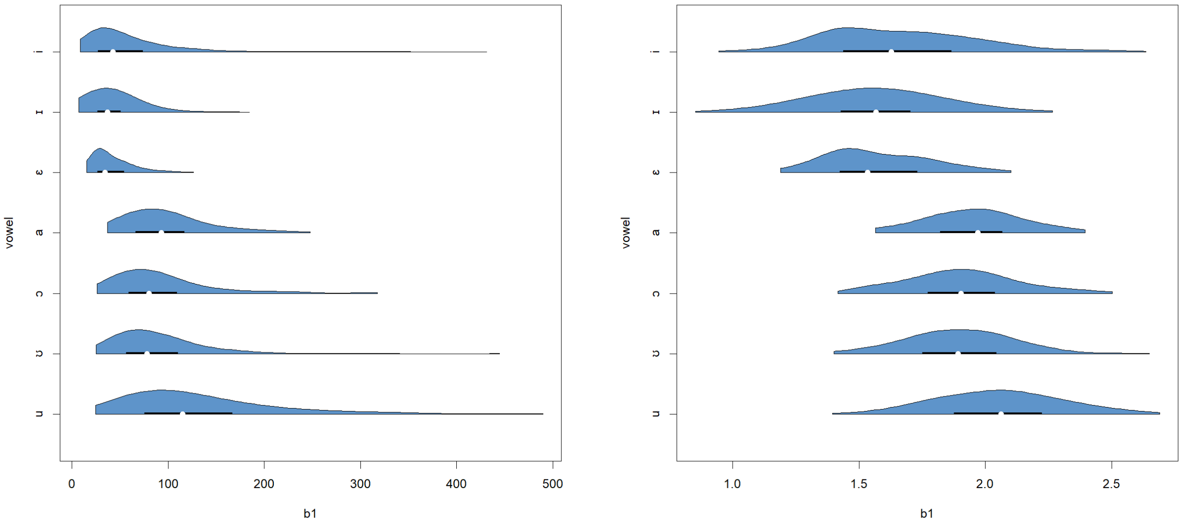


Figure 5.9: Diamond plot showing averaged CPP measurements for each vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

ments before and after logarithmic transformation. While the untransformed data (in Figure 5.10a) is clearly skewed and not very normal, the log-transformed data (in Figure 5.10b) is much more symmetrical and looks much more normally distributed (though there are clear differences in variance from one vowel to another).



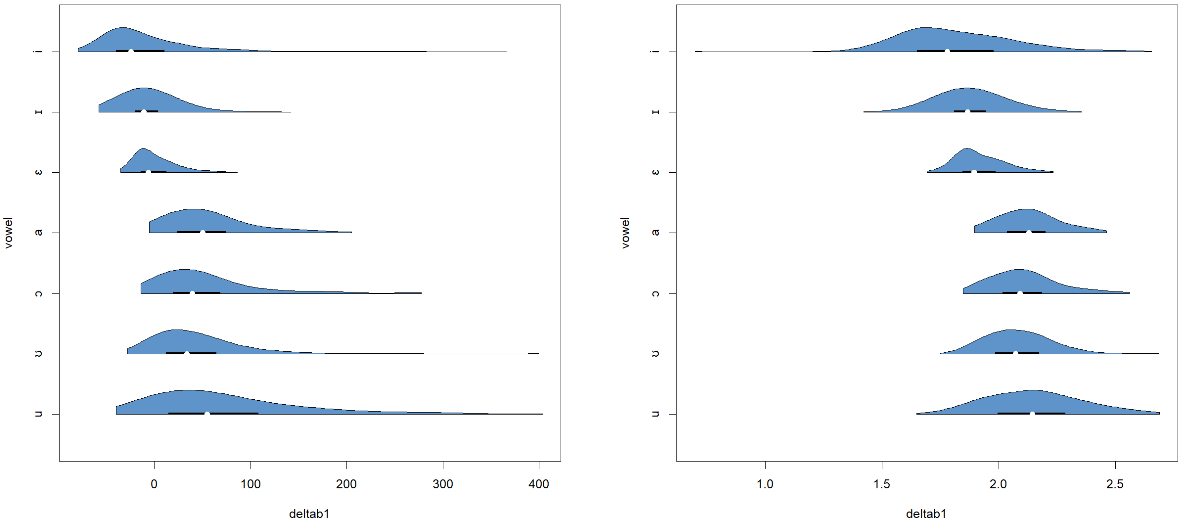
(a) Violin plot of untransformed B1 measurements (Hz)

(b) Violin plot of log-transformed B1 measurements (log(Hz))

Figure 5.10: Violin plots showing the effect of logarithmic transformation on the distribution of B1 measurements

A similar transformation can be applied to the ΔB1 measurements, with the additional complication that there are negative values of ΔB1 which cannot have a direct logarithm taken. I considered two options to apply a logarithm-like transformation to the ΔB1 measurements. The first was to apply a version of the logarithm function around 0, which would effectively calculate $\log{(b+1)}$ for values greater than or equal to 0 and $-\log{(-b+1)}$ for values less than 0. This would preserve the polarity of all the measurements and produce a logarithmic transformation for both positive and negative inputs. The problem with this option is that, though it correctly bunches higher measurements closer together, it also spreads out the measurements close to 0 apart from one another. Given that many measurements were close to 0, this gave undue importance to the difference between positive and negative values (especially since so many vowels had distributions which bridged the positive and negative realms). Thus I went for the second option, which was to shift all measurements up by the addition of a constant, such that the minimum measurement became 5. I chose 5 because it avoided the stretching near 1 while maintaining an appropriate level of skew reduction. The resulting distributions from the shifted log-transformation can be seen in Figure 5.11. Note the stretching of the distribution for the minimum measurement in 5.11b which, though unideal, is preferable to the untransformed skew.

Applying the same transformation to the centre of gravity measurements also gives a degree of skewing of the lower values, but also reduces the impact of the high outliers for measurements of /[i]/, as can be seen in Figure 5.12. In particular, the front vowels become much more normal-looking in



(a) Violin plot of untransformed ΔB1 measurements (Hz)

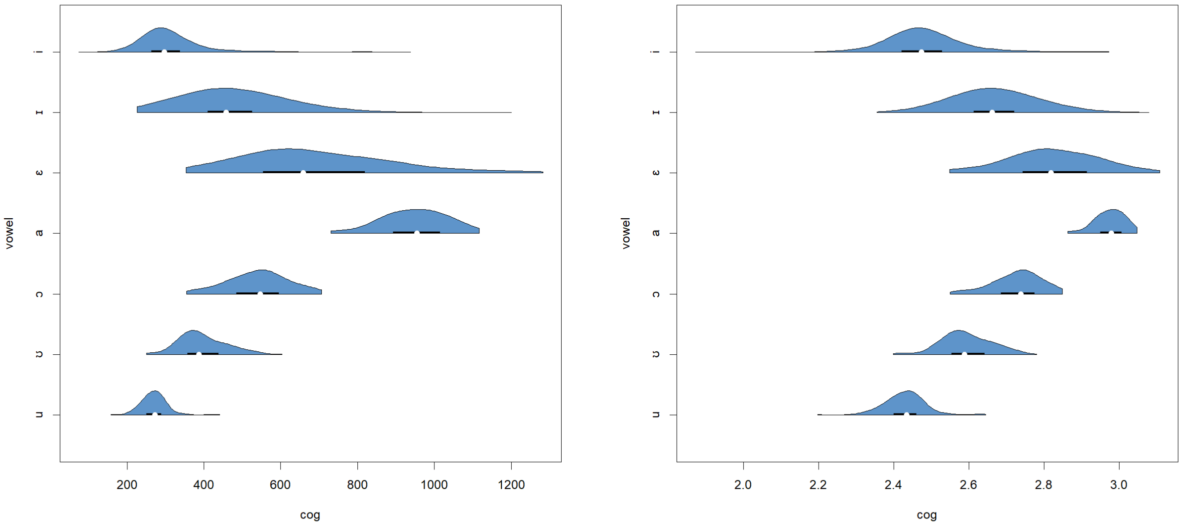
(b) Violin plot of shifted log-transformed ΔB1 measurements (log(Hz))

Figure 5.11: Violin plots showing the effect of (shifted) logarithmic transformation on the distribution of ΔB1 measurements

their distributions after transformation, while /[a]/ and /[ɔ]/ have slightly more negatively skewed distributions after transformation (though by no means to the same extent that the front vowels were positively skewed beforehand). I argue that the benefits of transformation outweigh the downsides, so centre of gravity is transformed for all the following statistical analyses.

Arguably, averaged CPP could also be a candidate for log-transformation: all of the vowels are skewed positively, and though applying the transformation does make some vowels have negatively skewed distributions, their absolute skewness is still smaller than the untransformed skewness (with the exception of /[i]/ which would go from a skewness of 0.37 to –0.38). On the other hand, the original skewness is quite minor in the first place, with no vowels having a skewness value above 1. For B1 and ΔB1, on the flipside, all skewness values are greater than 1, so log-transformation is very obviously useful. In the end, I argue that it is better to work with untransformed data where it is appropriate, and I do not find the untransformed skewness sufficiently worrisome to necessitate transformation.

One potential concern with the methodology employed with transformations here is that only some variables have been transformed, while others have been left as they were. This is relevant because both the folding test used to test Hypothesis 1 and the generalised linear model applied for Hypotheses 3–7 are multivariate, taking into account all variables at once. If this is indeed a concern, the tests could be performed without any transformations, with transformations of all variables or with non-parametric tests which make no assumptions about the underlying distribution of the data. I argue that, given that there is no direct comparison of the *values* of any measurement with another (e.g. I never directly compare the value of F1 to the value of CoG), it is only important that for each measurement the transformation is applied to all values measured (e.g. for all vowels).



(a) Violin plot of untransformed CoG measurements (Hz)

(b) Violin plot of log-transformed CoG measurements (log(Hz))

Figure 5.12: Violin plots showing the effect of logarithmic transformation on the distribution of CoG measurements

Another concern is that the transformations were applied to reduce skewness to make the distributions more normal-like, though in theory the most relevant factor here would be whether the *residuals* of each variable were normally distributed. Given the limited nature of this study, I only opted to check that skewness was reduced, but in future studies it might be good to be even more meticulous in determining how and where to apply these kinds of transformation.

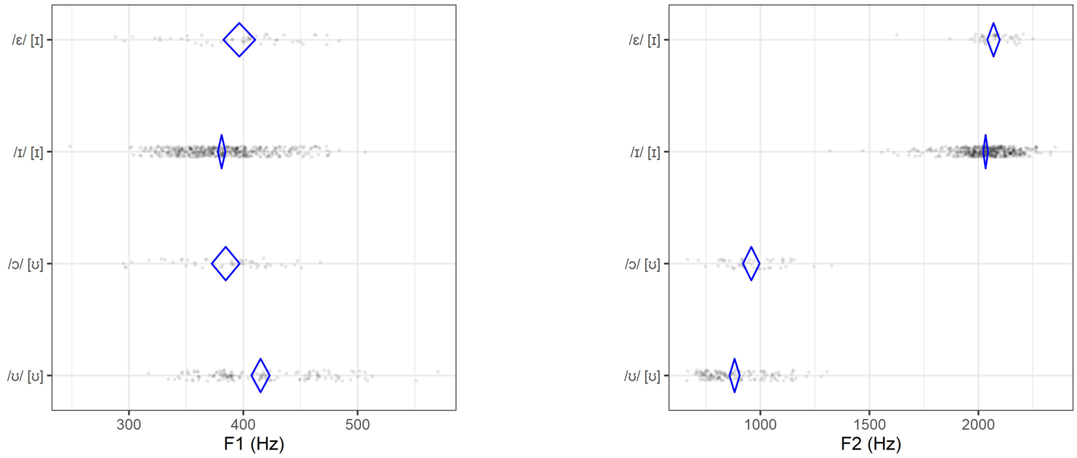
5.1.4 Comparing derived and underived vowels

Ahead of the discussion of the results of the tests for Hypothesis 2, it is worth looking briefly into what the data looks like when comparing derived and underived Level 2 vowels. The raw datafile has annotations for the transcribed surface vowel (named *vowel*) and the proposed underlying vowel (named *urvowel*). Thus the vowels of interest for this question are those where the *vowel* is transcribed [ɪ ʊ] with the *urvowel* being either /ɪ ʊ/ (underived) or /ɛ ɔ/ (derived). The hypothesis is testing the idea that the examples of raised /ɛ ɔ/ might represent surface [e o] rather than [ɪ ʊ], in which case, we would expect to see differences in F2, B1, ΔB1, A1*–A2* or CoG.[2]

The means and standard deviations of all measures of derived and underived [ɪ] and [ʊ] are given in Table 5.9. Diamond plots are shown for these measurements in Figure 5.13. The most immediately noticeable fact is that there is a big difference in frequency of each *vowel*/*urvowel* combination. There are many tokens of underived [ɪ], fewer of underived [ʊ] and relatively few derived vowels in general. This makes sense given that [ɪ] is also found in a number of affixes and the derived vowels only appear in a few specific morphological environments. Looking at individual measures, there is no clear visual distinction between derived and underived tokens for B1, ΔB1 or A1*–A2*. There are

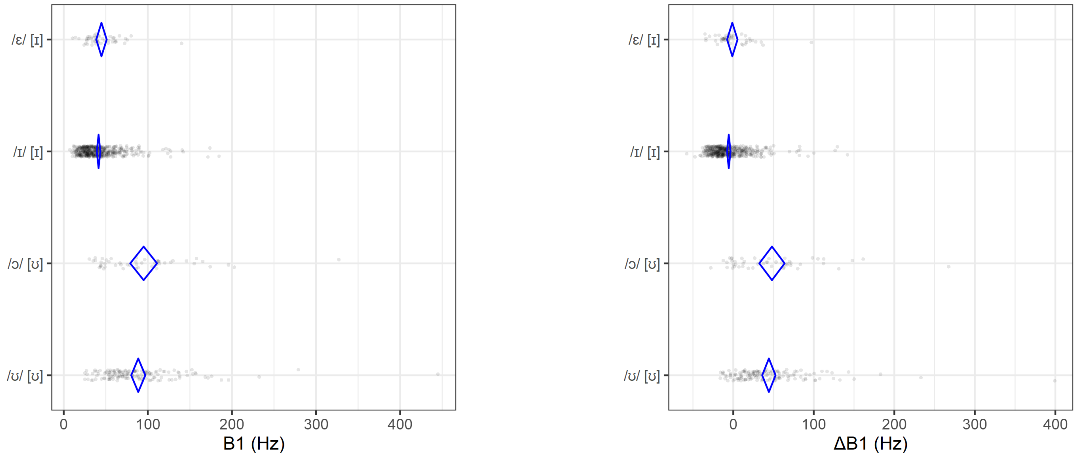
---

[2] There is no evidence from prior studies that F1 or CPP significantly distinguishes between [e o] and [ɪ ʊ], though Hypothesis 6 does test CPP as a possible distinguishing correlate on the basis of trends that were not tested statistically.



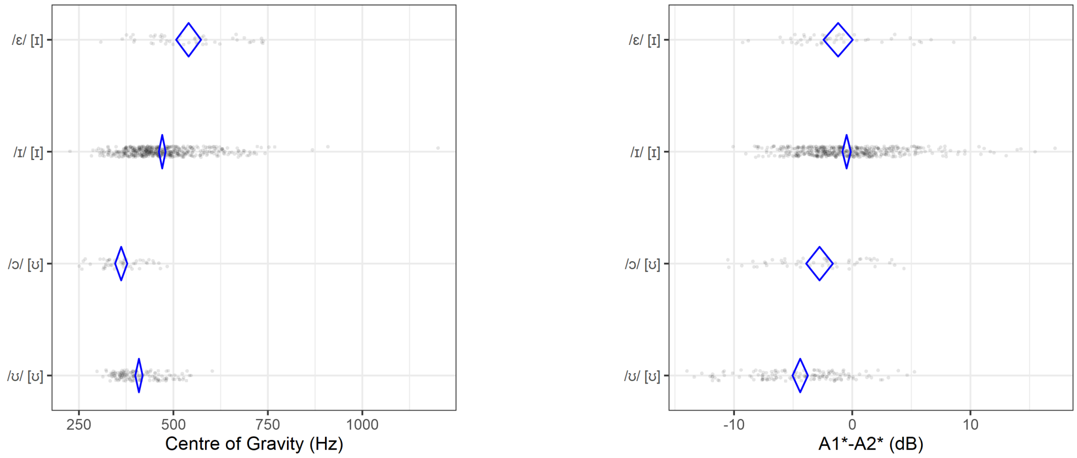
(a) F1 of derived vs underived vowels

(b) F2 of derived vs underived vowels



(c) B1 of derived vs underived vowels

(d) ΔB1 of derived vs underived vowels



(e) Centre of gravity of derived vs underived vowels

(f) A1*–A2* of derived vs underived vowels

Figure 5.13: Diamond plots showing measurements for surface [ɪ ʊ], separated by the underlying vowel. Dots show individual tokens, the blue diamond shows the mean and a 95% confidence interval on either side.

Table 5.9: Means and standard deviations of derived and underived Level 2 vowels

		Derived		Underived	
Vowel	Measure	Mean	SD	Mean	SD
ɪ	F1	396	48	381	37
ɪ	F2	2,069	99	2,033	127
ɪ	$\log_{10}$(B1)	1.61	0.20	1.56	0.22
ɪ	$\log_{10}$(ΔB1)	1.91	0.11	1.88	0.12
ɪ	A1*–A2*	−1.19	4.26	−0.48	4.01
ɪ	$\log_{10}$(CoG)	2.72	0.09	2.66	0.09
ɪ	CPPnf	14.56	4.89	14.98	4.42
ɪ	CPPavg	13.65	2.30	13.39	2.42
ʊ	F1	384	42	415	49
ʊ	F2	965	134	882	142
ʊ	$\log_{10}$(B1)	1.91	0.23	1.90	0.21
ʊ	$\log_{10}$(ΔB1)	2.09	0.16	2.09	0.14
ʊ	A1*–A2*	−2.67	3.89	−4.40	3.93
ʊ	$\log_{10}$(CoG)	2.55	0.07	2.61	0.06
ʊ	CPPnf	14.46	4.11	15.77	4.22
ʊ	CPPavg	13.37	2.88	14.11	2.49

means that seem different for F2, though it is hard to tell by eye what is a true difference and what is just an artefact of the number of tokens. For F1 and centre of gravity, there are also means that look different to the eye, but the differences are more extreme for [ʊ] than they are for [ɪ]. What this means in terms of equivalence and difference will be investigated when we look at Hypothesis 2, but it is worthwhile to bear in mind the differences in numbers of tokens between the derived and underived vowels when interpreting how they differ from one another.

5.2 Research question (I): One or two sets of Level 2 vowels?

5.2.1 Hypothesis 1: Unimodality

As described earlier (§4.3.3.1), Hypothesis 1 tests the unimodality of the Level 2 vowels using the multivariate folding test developed by Siffer et al. (2018). The output of the test is known as the folding statistic Φ, which indicates unimodality (that Boa-Leboale has just one set of Level 2 vowels) at values above 1 and multimodality (that Boa-Leboale has two sets of Level 2 vowels) at values below 1. The minimum possible value of Φ is 0. Instead of interpreting the value of Φ in terms of a p-value, Siffer et al. (2018: 6) gives a table of values showing values of Φ that are significantly different from 1 ($p < 0.05$) depending on the number of variables (‘dimensions’) and the number of data points (‘n’). In the case of this study, the number of variables is 6 (F1, F2, ΔB1, A1*–A2*, CoG, averaged CPP) and the number of data points is 543 for [ɪ] and 194 for [ʊ]. The table has a maximum number of dimensions of 5, and values for $n = 500$ and $n = 200$, which give approximate analogues for the true values of n for [ɪ] and [ʊ] respectively.

The relevant figures from Siffer's interpretation table are as follows. For [ɪ] ($n \approx 500$), the test is significant if the value of Φ is more than 0.17 away from 1 (if $\Phi(\text{ɪ}) < 0.83$ or > 1.17). For [ʊ] ($n \approx 200$), the test is significant if the value of Φ is more than 0.27 away from 1 (if $\Phi(\text{ʊ}) < 0.73$ or > 1.27).

Given this, the results of the multivariate test is given in Table 5.10. The first row gives the output values of the test statistic Φ and the second gives the number of tokens of each vowel included (n).

Table 5.10: Values of the folding statistic Φ for the reduced set of acoustic measures

Measure	ɪ	ʊ
all	0.000	0.001
n	543	194

Since the values of Φ(ɪ) and Φ(ʊ) are further away from 1 than the value given in Siffer's table ($\Phi(\text{ɪ}) < 0.83$ and $\Phi(\text{ʊ}) < 0.73$), both results suggest a significant result. And since both results are significantly lower than 1, this indicates that the vowels transcribed as [ɪ] and [ʊ] are significantly **multimodal** according to the multivariate folding test. This means that, taking all 6 measurements into consideration, both [ɪ] and [ʊ] were better described as having multiple modes rather than just one modal value, as we might expect if there is just one set of Level 2 vowels.

The next question is whether this multimodality measured is found in any particular measure or if it only arises when looking at all the measurements together. To investigate this, I also calculate the folding statistic for each measure separately ('unidimensional' tests), as shown in Table 5.11. These tests are exploratory and performed after having run other tests ('post-hoc'), so the aim is not to make inferential statements, just to observe if there is any measure in particular which is more multimodal.

Table 5.11: Values of the folding statistic Φ for each measure separately

Measure	ɪ	ʊ
F1	1.449	1.431
F2	1.894	1.146
$\log_{10}(\Delta B1)$	1.586	1.607
A1*–A2*	1.651	1.305
$\log_{10}(\text{CoG})$	1.552	1.498
CPPavg	1.829	1.693
n	543	194

The unidimensional statistics are all greater than 1, so they all indicate that each measure is unimodal when investigated separately from one another. Thus the multimodality found earlier for [ɪ] and [ʊ] must arise from the interaction of multiple measurements rather than just one. See §5.2.3 for a discussion on how these results can be interpreted in light of research question (I).

5.2.2 Hypothesis 2: Derived vs underived vowels

As described earlier (§4.3.3.2), Hypothesis 2 tests the equivalence of the derived and underived vowels for each measure separately using the TOST (Two One-Sided Tests) procedure. This involves conducting two one-sided t-tests, one for the region less than the upper equivalence bound for the mean and the other for the region greater than the lower equivalence bound for the mean (see §4.3.3.2 for justifications on how the equivalence bounds were chosen). The greater of these two p-values is considered the probability that the mean is lower than the lower bound or greater than the upper bound (the null hypothesis being that the two values are different from one another by a certain amount). The implementation in the R package *TOSTER* also conducts a two-sided t-test for whether the variables have significantly different means.

The results for [ɪ] and [ʊ] are reported in Tables 5.12 and 5.13 respectively. In both tables, the t-values and p-values for each of the three t-tests is reported.

Table 5.12: Results of the TOST tests for derived vs underived [ɪ]

		t			p		
Vowel	Measure	Lower	Upper	2-Sided	Lower	Upper	2-Sided
ɪ	F1	7.47	−3.14	2.17	0.000	0.001	0.035
ɪ	F2	8.87	−4.15	2.36	0.000	0.000	0.021
ɪ	$\log_{10}$(B1)	7.70	−4.46	1.62	0.000	0.000	0.110
ɪ	$\log_{10}$(ΔB1)	7.77	−4.39	1.69	0.000	0.000	0.097
ɪ	A1*–A2*	4.56	−6.80	−1.12	0.000	0.000	0.267
ɪ	$\log_{10}$(CoG)	9.89	−1.41	4.24	0.000	0.082	0.000
ɪ	CPPnf	5.03	−6.16	−0.57	0.000	0.000	0.572
ɪ	CPPavg	6.67	−5.22	0.73	0.000	0.000	0.470

Table 5.13: Results of the TOST tests for derived vs underived [ʊ]

		t			p		
Vowel	Measure	Lower	Upper	2-Sided	Lower	Upper	2-Sided
ʊ	F1	1.26	−9.92	−4.33	0.105	0.000	0.000
ʊ	F2	9.20	−1.73	3.74	0.000	0.044	0.000
ʊ	$\log_{10}$(B1)	5.72	−4.83	0.45	0.000	0.000	0.657
ʊ	$\log_{10}$(ΔB1)	5.30	−5.10	0.10	0.000	0.000	0.922
ʊ	A1*–A2*	8.14	−2.68	2.73	0.000	0.004	0.008
ʊ	$\log_{10}$(CoG)	0.33	−10.20	−4.93	0.369	0.000	0.000
ʊ	CPPnf	3.50	−7.36	−1.93	0.000	0.000	0.057
ʊ	CPPavg	3.59	−6.85	−1.63	0.000	0.000	0.107

The p-values can be interpreted as follows: for each measure, the higher p-value of the Lower and Upper tests is the p-value for the equivalence test (with the null hypothesis that the derived and underived vowels have non-equivalent means), and the p-value of the 2-Sided test is the p-value for the regular t-test (with the null hypothesis that the derived and underived vowels have non-different means). The significance level is set at $p = 0.025$, after applying a Bonferroni correction to correct for the fact that the tests were performed for both [ɪ] and [ʊ].[3]

---

[3] From my understanding, this is just one of many possible methods of correcting for the multiple testing problem, which is particularly relevant in this study, where multiple levels of the vowel factor and many different dependent vari-

The equivalence and difference of the means can be interpreted separately; the means could be statistically equivalent and statistically different, such as for [ʊ]~A1*–A2* in Table 5.13 (for the equivalence test, $p = 0.004$ and for the two-sided test, $p = 0.008$. Both of these fall below the threshold of the significance level of $p = 0.025$, so both results are statistically significant. This means that the means of the A1*–A2* are statistically different from one another, but by a small enough degree that they are effectively equivalent to one another as well.

For [ɪ], all measures but centre of gravity have significantly equivalent means for derived and underived vowels. They also do not show significant differences in the two-sided tests. The exception here is centre of gravity, where the derived vowels have significantly higher centre of gravity than underived vowels, as shown by the two-sided test. Moreover, the equivalence test at the current significance level does not rule out the possibility that they have different means ($p = 0.082$). This particular finding is consistent with the theory that the vowel derived from raising underlying /ɛ/ is in fact [e] instead of [ɪ], since previous studies found that [e] had higher measurements of centre of gravity than [ɪ].

For [ʊ], B1, ΔB1, A1*–A2* and both CPP measures have significantly equivalent means for derived and underived vowels. B1, ΔB1 and the CPP measures are also not significantly different in the two-sided test, though A1*–A2* is significantly equivalent and significantly different, as discussed above. The remaining measures (F1, F2 and CoG) are all significantly different in the two-sided test and the equivalent test does not rule out that they could have different means ($p > 0.025$).

Derived [ʊ] has lower F1, higher F2 and lower CoG means than underived [ʊ]. The F1 and CoG measurements cannot be interpreted as evidence for hypothesis 2 because there are no previously reported consistent differences between [ʊ] and [o] with regards to F1 or CoG.[4] Previous studies found that [o] typically had a higher mean F2 than [ʊ], so the F2 result is also consistent with the theory that the vowel derived from raising /ɔ/ is [o] instead of [ʊ].

5.2.3 Discussion of RQ(I) results

Research question (I) asks about whether there is more than one set of Level 2 vowels. On the basis of the results just described, we can make some progress towards answering this question.

Regarding Hypothesis 1 about the unimodality of the Level 2 vowels, the multidimensional folding test suggested a significant effect ($p < 0.05$) that [ɪ] and [ʊ] had a folding statistic Φ below 1, indicating multimodality. I also conducted post-hoc unidimensional folding tests for each measurement. These tests indicated that each acoustic measurement, for both [ɪ] and [ʊ], was more unimodal when taken in isolation. Thus, the reported multimodality will most likely come from the interaction of multiple acoustic measurements rather than from any given individual measure.

It is difficult to interpret this result without further research beyond the scope of this thesis; a cluster analysis could potentially shed some light on how the reported multimodality presents itself in the data, and if there are any connections between the different clusters and the phonological behaviour

---

ables are being tested. The Bonferroni correction is particularly conservative, and often less statistically powerful than other methods, but I have applied it here due to the simplicity of its use.

[4] Though there is a reported difference in CoG between the different Level 2 *front* vowels (see §2.4.2.2).

of the vowel(s) in question. For instance, some of the phonological processes in Boa-Leboale only apply to some verb roots and not others; this could potentially indicate that there are different vowels in these roots. It is however important to note that no such differences between roots have been reported impressionistically by any linguists working on Boa-Leboale, so any suggestion that this might be the case would need extremely solid evidence in support of it to be a reasonable conclusion.

I also have some doubts that the finding of multimodality here is completely reliable. Firstly, though the folding test developed by Siffer et al. (2018) is the only option I could find without unimodality as the null hypothesis and that is applicable for multivariate data, it is not a widely used statistical test; even the significance of the test is only available from extrapolating from a table of quantiles, and the output of the test is difficult to interpret. Perhaps if another such a test (either one where unimodality is not the null hypothesis or one which can looking for multivariate multimodality) is developed and proven to be useful in statistical contexts, it can be used more effectively on this data. Secondly, in all previous studies looking at differences between the Level 2 vowels, such differences have been observed in individual measurements – the fact that all the individual measures appear to be unimodal by themselves might perhaps be a better indicator than the test taking all of the measurements into consideration at once.

Regarding Hypothesis 2 about differences between derived and underived Level 2 vowels, the TOST equivalence and two-sided t-tests found that derived and underived Level 2 vowels had statistically equivalent means for the majority of the acoustic measurements investigated. All cases that were not shown to be significantly equivalent in the TOST tests were found to be significantly different in the two-sided t-tests, though some could be discounted as relevant to this specific hypothesis since there was no prior evidence that they had any predictive differences between the Level 2 vowels.

Thus the only cases which suggested that the derived and underived vowels are different Level 2 vowels were [ɪ]~CoG and [ʊ]~F2. In both cases, the direction in which these measures were significantly different for derived and underived vowels was consistent with the theory that the derived vowels might be [+ATR] [e o] and the underived vowels [−ATR] [ɪ ʊ]. The two-sided t-test for /ʊ/~A1*–A2* also found a significant difference between means consistent with this theory, though the equivalence tests found that the means were also significantly within the bounds of equivalence.

These tests therefore provide some tentative evidence that there could be a surface distinction between [ɪ ʊ] and [e o], with the latter only appearing in derived contexts. However, given that this difference was only shown to be non-equivalent for two relevant acoustic measures – CoG for [ɪ] and F2 for [ʊ] – this finding is by no means conclusive.

In summary, research question (I) sought to investigate whether each of the vowels transcribed as [ɪ ʊ] form a uniform category, or if there was evidence that Boa-Leboale had 9 vowels, either underlyingly or on the surface. The multidimensional folding tests for unimodality found that both [ɪ] and [ʊ] were significantly multimodal, suggesting that they might best be described as more than one group. However unidimensional tests for each acoustic measure found that they all seemed more unimodal in isolation. Further research, perhaps involving a cluster analysis, might be able to provide more insight into exactly how the categories might be divided (and if this corresponds with any phonological properties), and might be able to replicate this experiment with a more reliable and easily interpretable statistical test.

The TOST equivalence and two-sided t-tests found that for most relevant acoustic measures, underived [ɪ ʊ] and vowels derived from /ɛ ɔ/ in a raising environment are equivalent to one another. There were, however, two cases where the TOST tests could not rule out that the means were different from one another, and the two-sided tests found that the means were significantly different. Both cases were consistent with the theory that raised /ɛ ɔ/ is more like [+ATR] [e o] compared to underived [–ATR] [ɪ ʊ].

We can tentatively say the following as an answer for research question (I). If Boa-Leboale has 9 underlying vowels (with both [+ATR] and [–ATR] Level 2 vowels), this remains to be proven by further research into the clustering of each of the measurements. For now, it seems that there are only 7 underlying vowels. Raised /ɛ ɔ/ was shown to be equivalent to other underived cases of [ɪ ʊ] for almost all acoustic measures, aside from CoG (for [ɪ]) and F2 (for [ʊ]). These two cases had means which were significantly different and not significantly equivalent, and consistent with an interpretation that raised /ɛ ɔ/ is [+ATR] [e o], while the underived Level 2 vowels are [–ATR] [ɪ ʊ]. Thus it is possible (though awaiting further investigation) that Boa-Leboale has 9 surface vowels.

5.3 Research question (II): /1IU/ or /2IU/?

Research question (II) hopes to distinguish whether a given Level 2 vowel is [+ATR] or [–ATR] on the basis of a batch of different tests. Each test looks at a given measure and indicates one result or the other based on the Level 2 vowel's mean value relative to Level 1 or Level 3 mean values. Given the findings from research question (I), the tests for Hypotheses 3–7, which rely on generalised linear modelling and subsequent linear hypothesis testing using the R package *multcomp*, are conducted on the underived and derived Level 2 vowels separately. Both sets of results will be reported and interpreted in terms of whether they suggest that the given vowel is [+ATR] or [–ATR].

5.3.1 Hypothesis 3: ΔB1 diagnostic

As described earlier (§4.3.3.3), Hypothesis 3 tests whether the mean ΔB1 value for [ɪ ʊ] is greater or less than the mean ΔB1 value for [i u] (with both sets of values having undergone log-transformation, see §5.1.3). If a Level 2 vowel has a significantly greater mean ΔB1 value than its Level 1 counterpart, then this test indicates that the Level 2 vowel is [–ATR]. If the Level 2 vowel's ΔB1 is significantly less than its Level 1 counterpart, this test indicates that the Level 2 vowel is [+ATR].

The results of these tests – for the front vowels, back vowels, derived Level 2 vowels and underived Level 2 vowels – are shown in Table 5.14. The H_0 column denotes the linear hypothesis being tested, with the z and p columns showing the z- and p-values respectively.

The p-values given in this table are adjusted for the front/back cases using a single-step method. Thus, the only aspect which needs further adjustment is the fact that the underived and derived cases were tested separately, so I set the significance level here as $p = 0.025$.

For the front vowels, both derived and underived [ɪ] had a greater ΔB1 mean than [i] (as shown by the negative z-value). In both cases, this difference was also significant at the chosen significance

Table 5.14: Results of the linear hypothesis tests for ΔB1

H_0		z	p
$\text{i} - \text{ɪ} = 0$	underived	-5.24	0.000
$\text{u} - \text{ʊ} = 0$	underived	3.48	0.001
$\text{i} - \text{ɪ} = 0$	derived	-2.97	0.006
$\text{u} - \text{ʊ} = 0$	derived	2.13	0.065

level. This indicates that both the derived and underived vowels are [–ATR] [ɪ] rather than [+ATR] [e].

For the back vowels, both derived and underived [ʊ] had a smaller ΔB1 mean than [u] (as shown by the positive z-value). However, only the underived case found a significant difference (p = 0.001) at the chosen significance level. This indicates that the underived vowel is [+ATR] [o] rather than [–ATR] [ʊ].

Comparing the derived and underived vowels, there is no difference in terms of whether the z-value is greater than or less than 0. However, the z-values for both derived vowels are closer to 0 than their underived equivalents. I imagine this is because there are fewer tokens of derived vowels than underived vowels, rather than because there is some meaningful difference between the derived and underived vowels. In any case, the most important factor is whether the z-value is greater than or less than 0, which does not vary according to derivedness.

Taking this test at face value, there is a clear asymmetry in ATR between the front and back Level 2 vowels [ɪ o]. This explanation is unlikely if this is the only piece of evidence in its favour, so I suggest the following alternative. In the discussion of the descriptive statistics (§5.1.2.3), I noted that [u] had an unexpectedly high mean and standard deviation for both B1 and ΔB1. This is likely due to the automatic formant measurement incorrectly identifying the two narrow formants (F1, F2) of [u] as one much broader F1 formant, inflating the mean value and giving a wide dispersal of measurements. This would also lead to a much higher mean than expected for [u], which in turn gives a positive z-value in this test. As such, it is possible that the effect shown here is not evidence that the back Level 2 vowel is [+ATR] [o], but instead that the mean ΔB1 of [u] was measured inaccurately.

The same argumentation can be given for [i] as well, which has a mean much higher than its median and also has many anomalous outlying measurements. However, in this case, since the mean has been increased by the erroneous measurement, we would expect the true mean to be lower, which would simply indicate even more strongly that the Level 2 vowel is [–ATR] [ɪ]. In other words, it would reinforce the conclusion already reached rather than contradict it.

5.3.2 Hypothesis 4: A1*–A2* diagnostic

As described earlier (§4.3.3.3), Hypothesis 4 tests whether the mean A1*–A2* measurement for [ɪ ʊ] is greater or less than the mean A1*–A2* measurement for [i u]. If a Level 2 vowel has a significantly greater mean A1*–A2* value than its Level 1 counterpart, then this test indicates that the Level 2 vowel is [+ATR]. If the Level 2 vowel's A1*–A2* is significantly less than its Level 1 counterpart, this test indicates that the Level 2 vowel is [–ATR].

The results of these tests – for the front vowels, back vowels, derived Level 2 vowels and underived Level 2 vowels – are shown in Table 5.15. The H_0 column denotes the linear hypothesis being tested, with the z and p columns showing the z- and p-values respectively.

Table 5.15: Results of the linear hypothesis tests for A1*–A2*

H_0		z	p
$\text{i} - \text{ɪ} = 0$	underived	17.11	0.000
$\text{u} - \text{ʊ} = 0$	underived	8.51	0.000
$\text{i} - \text{ɪ} = 0$	derived	7.98	0.000
$\text{u} - \text{ʊ} = 0$	derived	3.57	0.001

Like with the results of Hypothesis 3, the p-values given in this table are adjusted for the front/back cases using a single-step method. Thus, the only aspect which needs further adjustment is the fact that the underived and derived cases were tested separately, so the significance level here is also $p = 0.025$.

Both derived and underived [ɪ ʊ] had a smaller A1*–A2* mean than [i u] (as shown by their positive z-values). In all cases, this difference was also significant at the chosen significance level. This indicates that both the derived and underived Level 2 vowels are [–ATR] [ɪ ʊ] rather than [+ATR] [e o].

Like with Hypothesis 3, the z-values for both derived vowels are slightly closer to 0 than their underived equivalents. This is also, like before, likely due to the relatively few tokens of derived vowels measured in this study, rather than any meaningful difference according to derivedness.

5.3.3 Hypothesis 5: Centre of gravity diagnostic

As described earlier (§4.3.3.3), Hypothesis 5 tests whether the mean centre of gravity measurement for [ɪ] is closer to the mean CoG measurement for [i] or [ɛ]. If [ɪ] is closer to [i] than it is to [ɛ] ($z > 0$), then this test indicates that the Level 2 vowel is [+ATR] [e]. If [ɪ] is closer to [ɛ] than it is to [i] ($z < 0$), this test indicates that the Level 2 vowel is [–ATR] [ɪ].[5]

The results of these tests – for the derived and underived Level 2 front vowel – are shown in Table 5.17. The H_0 column denotes the linear hypothesis being tested, with the z and p columns showing the z- and p-values respectively. The underived and derived cases were tested separately, so the significance level here is $p = 0.025$.

Both derived and underived [ɪ] gave negative z-values, but this was only significantly different from 0 for the derived vowels ($p < 0.001$), not for the underived vowels ($p = 0.065$). This indicates that derived [ɪ] is [–ATR] [ɪ] rather than [+ATR] [e].

---

[5] This interpretation relies on the mean values for centre of gravity for the three vowels being as predicted (ɛ > ɪ > i), which is indeed the case (see §5.1.2.5). If that hierarchy was reversed, the interpretation of the test would also be reversed; the distances would be negative and as such the direction of the inequality would be flipped (cf. 23 > 13 vs. -23 < -13).

Table 5.16: Results of the linear hypothesis tests for centre of gravity

H_0		z	p
$\text{i} + \text{ɛ} - 2\text{ɪ} = 0$	underived	-1.85	0.065
$\text{i} + \text{ɛ} - 2\text{ɪ} = 0$	derived	-5.11	0.000

5.3.4 Hypothesis 6: CPP diagnostic

As described earlier (§4.3.3.3), Hypothesis 6 tests whether the mean averaged CPP measurements for the Level 2 vowels are closer to the mean averaged CPP measurements for the Level 1 or Level 3 vowels. If [ɪ ʊ] is closer to [i u] than it is to [ɛ ɔ] ($z > 0$), then this test indicates that the Level 2 vowel is [–ATR] [ɪ ʊ]. If [ɪ ʊ] is closer to [ɛ ɔ] than it is to [i u] ($z < 0$), this test indicates that the Level 2 vowel is [+ATR] [e o].[6]

The results of these tests – for the front vowels, back vowels, derived Level 2 vowels and underived Level 2 vowels – are shown in Table 5.16. The H_0 column denotes the linear hypothesis being tested, with the z and p columns showing the z- and p-values respectively.

Table 5.17: Results of the linear hypothesis tests for averaged CPP

H_0		z	p
$\text{i} + \text{ɛ} - 2\text{ɪ} = 0$	underived	0.26	0.957
$\text{u} + \text{ɔ} - 2\text{ʊ} = 0$	underived	-0.81	0.662
$\text{i} + \text{ɛ} - 2\text{ɪ} = 0$	derived	-0.52	0.845
$\text{u} + \text{ɔ} - 2\text{ʊ} = 0$	derived	1.35	0.322

Like with the results of Hypotheses 3 and 4, the p-values given in this table are adjusted for the front/back cases using a single-step method. Thus, the only aspect which needs further adjustment is the fact that the underived and derived cases were tested separately, so the significance level here is also $p = 0.025$.

There are no trends for the polarity of z on the basis of backness or derived-ness, and none of the tests were close to being significant, so the averaged CPP measurements for the Level 2 vowels were all indistinguishable from falling halfway between the Level 1 and Level 3 vowels. As such, these tests give no indications about the identities of the Level 2 vowels.

5.3.5 Hypothesis 7: Diagnostic using ΔB1 of /ɛ ʊ/

As described earlier (§4.3.3.4), Hypothesis 7 uses a hypothesis inspired by a suggestion from Starwalt (2008) to tests whether the mean ΔB1 measurement for [ɛ] is greater than the mean ΔB1 measurement for [ʊ]. If [ɛ] is greater than [ʊ], this indicates that the Level 2 back vowel is [+ATR] [o], else the test does not indicate anything.

The results of these tests – for derived and underived [ʊ] – are shown in Table 5.18. The H_0 column denotes the linear hypothesis being tested, with the z and p columns showing the z- and p-values

---

[6] Like with Hypothesis 5, this interpretation relies on the mean values for CPP for the three vowels being as predicted (ɛ > ɪ > i), which is indeed the case (see §5.1.2.6). If that hierarchy was reversed, the interpretation of the test would also be reversed.

respectively. The underived and derived cases were tested separately, so the significance level here is $p = 0.025$.

Table 5.18: Results of the linear hypothesis test for ΔB1 of /ɛ ʊ/

H_0		z	p
$ɛ - ʊ \leq 0$	underived	-7.84	1.000
$ɛ - ʊ \leq 0$	derived	-5.21	1.000

Since the means were actually contrary to the alternative hypothesis being tested by a large margin (mean ΔB1(ɛ) = 1.2Hz, mean ΔB1(ʊ) = 45.0Hz), the test had a (strongly) non-significant result ($p =$ 1.000) in both cases. This test therefore also gives no indication about the identities of the Level 2 vowels.

5.3.6 Discussion of RQ(II) results

Research question (II) asks about whether we can use the trends observed for a number of acoustic measures to determine if a given Level 2 vowel is [+ATR] or [–ATR]. Given that the outcome of research question (I) was inconclusive on whether there was a difference in ATR between derived and underived Level 2 vowels, the tests discussed here also suggest some answers to that research question as well. There were 5 total hypotheses being tested through 16 linear hypothesis tests, whose results are shown in Table 5.19.

Table 5.19: Summary of indications from Hypotheses 3–7 (– indicates a test was not run, ? that a test did not give a significant result)

		Underived		Derived	
Hypothesis	Measure	Front	Back	Front	Back
H3	$\log_{10}$(ΔB1)	ɪ	o	ɪ	?
H4	A1*–A2*	ɪ	ʊ	ɪ	ʊ
H5	$\log_{10}$(CoG)	?	–	ɪ	–
H6	CPPavg	?	?	?	?
H7	$\log_{10}$(ΔB1)	–	?	–	?

Since all these hypotheses were, until now, untested, it would be premature to say that any individual given result confirms, or even makes it likely, that a given vowel is [+ATR] or [–ATR]. However, the hope is that the *combination* of each of these tests, which by themselves might not be reliable, gives a greater indication of which ATR value is more likely. Suppose that each individual test has a moderate chance of a Type I error (false positive). Since most tests are two-sided, if the null hypothesis is actually true, then we would expect any false positives to be equally likely to indicate a [+ATR] or [–ATR] interpretation. Thus if we apply many tests and find that their results all match one another, this is considerably less likely to arise by chance.

With that in mind, Table 5.19 shows that all but one significant result indicated that the Level 2 vowels were [–ATR]. The only case where this was not found was for the test from Hypothesis 3 with the underived back vowel, which was significantly different in the opposite direction, which indicated that the vowel was [+ATR]. However, as discussed in the section on Hypothesis 3 (§5.3.1), this might actually be an artefact of the automatic measurement system; particularly, the difficulties that Praat has with distinguishing F1 and F2 for rounded back vowels like [u]. This led to a very high dispersal of B1 measurements for [u], with an unexpectedly high mean value, which would contribute to the indication of [+ATR]-ness. Perhaps with a better formant tracking software, like Fast Track (Barreda 2021), it would be more clear if this was a measurement error or an underlying difference between means.

These tests indicate quite clearly that the Level 2 vowels in Boa-Leboale are [–ATR]. Furthermore, they also seem to indicate that this is just as true for the derived vowels as it is for the underived vowels, so these tests suggest that Boa-Leboale does indeed have 7 underlying and 7 surface vowels (giving another answer to RQ(I)).

In terms of assessing the hypotheses themselves, it is clear that Hypothesis 6 was not a useful criterion for distinguishing [+ATR] and [–ATR] vowels in Boa-Leboale, since all the tests had non-significant results. Indeed, no existing studies have linked CPP to meaningfully distinguishing between Level 2 vowels, so this result is perhaps not unexpected. Hypothesis 7 also gave no indication of ATR value for the underived or derived vowels, though this could be because the test is only set up to indicate [+ATR] Level 2 vowels, not [–ATR] ones. Perhaps through future research this criterion could be adjusted to indicate both possible values of [ATR].

Hypotheses 3–5 seem to have been useful diagnostics, with the exception of the B1 measurement problems, though it remains to be seen whether the findings here can be replicated for other speakers and other languages.

5.4 Summary of results and discussions

In this chapter, I presented and discussed the results of the acoustic study outlined in Chapter 4, and how these results answer my research questions: (I) Do the Level 2 vowels each comprise one vowel category or two? and (II) Are the Level 2 vowels [+ATR] or [–ATR]?

Regarding research question (I), the tests performed to test Hypotheses 1 and 2 were either difficult to interpret or inconclusive. For Hypothesis 1 (which tested uni-/multimodality), the multidimensional folding test indicated multimodality for both [ɪ] and [ʊ], but this was not found for any individual acoustic measurement, which all seemed to indicate unimodality. For Hypothesis 2 (which compared derived and underived Level 2 vowels), almost all acoustic measurements were found to be statistically equivalent regardless of whether the vowel was derived or not. There were, however, two exceptions: one for [ɪ] and one for [ʊ]. Both of these exceptions found that the derived vowels were significantly different and not significantly equivalent to their underived counterparts, and that this difference was consistent with the analysis that the derived vowels were [+ATR] [e o] and the underived vowels [–ATR] [ɪ ʊ].

As a result of this inconclusive finding, research question (II) was investigated separately for the derived and underived vowels. In both cases, the tests for Hypotheses 3–5 overwhelmingly indicated that the Level 2 vowels were [–ATR] [ɪ ʊ], while Hypotheses 6–7 found no significant results in either direction. The only exception to this trend was the test run on the underived back vowel in

Hypothesis 3 (involving ΔB1), which instead indicated that that vowel was [+ATR]. However, this outlier is likely to be a result of inaccurate B1 measurement for [u], due to Praat's difficulties in tracking F1 and F2 for rounded back vowels.

To conclude, based on the evidence presented by this study, I argue that the Level 2 vowels in Boa-Leboale are [–ATR] [ɪ ʊ], as suggested by prior work on the language, and so it has a 7-vowel /2IU/ system. I also argue that Boa-Leboale also only has 7 surface vowels, that there is no [e o] in derived contexts.

From a methodological perspective, I believe that the statistical tests used in this study had mixed successes. On the less successful side, Hypothesis 1 (the folding test for unimodality) seemed rather unreliable, or at least difficult to interpret without performing follow-up cluster analyses beyond the scope of this thesis. Hypotheses 6 (CPP) and 7 (ΔB1 of [ɛ] and [ʊ]) also did not give any indication about the ATR value of the vowel in question, so they may not be relevant as acoustic diagnostics for ATR. Hypothesis 2 (equivalence testing) had some success, but its difficulty also lay in its interpretation. I would be keen to see future research determine whether Hypothesis 2 gives an appropriate result when applied to a 7-vowel /2IU/ language known to have surface level [e o]. On the more successful side, Hypotheses 3–5 (ΔB1, A1*–A2*, CoG) gave significant and easily interpretable results for the ATR value of the Level 2 vowels. I hope that future research can test these hypotheses on other speakers and languages to see if their success holds up cross-linguistically.

Chapter 6

Conclusions

In this thesis, I aimed to contribute to the academic discussion around how the Level 2 vowels [e o] and [ɪ ʊ] can be distinguished and/or identified purely through acoustic means. The results of my acoustic study do, indeed, seem to suggest that this is possible, even if not all of the tests devised gave meaningful or readily interpretable results.

I started in Chapter 2 by summarising the previous literature discussing the problem of distinguishing and identifying Level 2 vowels in Niger-Congo and Nilo-Saharan languages, with a particular focus on Starwalt (2008), the most wide-ranging such acoustic study. Much work had been done on distinguishing these vowels in languages which have both, but I could not find any study that had concretely tried to identify the Level 2 vowel in a given 7-vowel system. On the basis of these works, particularly Starwalt (2008), I identified a number of predictive differences between the [+ATR] vowels and [–ATR] vowels for several acoustic measurements (given in §4.3.2), particularly in how their mean values related to the means of neighbouring Level 1 or Level 3 vowels. These predictive differences went on to form the basis of Hypotheses 3–6 of my statistical analysis, which generally indicated that the Level 2 vowels in Boa-Leboale are [–ATR] [ɪ ʊ]. Hypothesis 6, based on averaged CPP, found no significant results, so this might not be useful in further research, but Hypotheses 3–5, based on ΔB1, A1*–A2* and centre of gravity, produce significant results that mostly agreed with one another (though see below for the issues regarding the automatic measurement of B1 impacting the testing of Hypothesis 3).

In Chapter 3, I outlined the phonology of Boa-Leboale, demonstrating that the vowel system is not easily identifiable through phonology alone, since the static and dynamic vowel patterns do not lend themselves towards any specific interpretation. From comparison with other related languages, particularly Liko, I raised the possibility that Boa-Leboale in fact has 9 vowels (with both sets of Level 2 vowels) underlyingly or on the surface, since this might simplify some aspects of the phonological analysis. This hypothesis, that Boa-Leboale might have more than one set of Level 2 vowels underlyingly or on the surface, was the basis for research question (I). For this, I devised two sets of tests: one to test if there was an underlying difference (Hypothesis 1) and one to test if there was a surface difference between derived and underived Level 2 vowels (Hypothesis 2). In the end, both set of tests ended up difficult to interpret; Hypothesis 1 found evidence of multimodality, but only when looking at all measurements at once rather than any one variable in particular, while Hypothesis 2 found that all measurements but two were statistically equivalent for derived and underived vowels. In the end, this research question was mostly answered by looking at the results of the tests for vowel identity, which were performed separately on underived and derived vowels. These tests found that both the derived and underived vowels were [−ATR] [ɪ ʊ], so Boa-Leboale seems only to have 7 vowels, on the surface and underlyingly. With appropriate adjustment of my statistical analysis – a better choice of statistical tests or a follow-up cluster analysis – I hope that the methodology I employed for Hypotheses 1 and 2 could still have some use. Perhaps with further studies on other languages, particularly those whose vowel systems are already known with confidence, the interpretability of the tests could be improved.

Chapter 4 explained the methodology of the study: how the data was collected by Gerrit de Wit, how the acoustic analysis was performed in Praat and how the statistical analysis was performed in R. I also defended the use of automated, rather than human-mediated, acoustic measurement in terms of its benefits for speed and replicability, though acknowledging its limitations in giving more errors, particularly regarding the tracking and measurement of formants. Though I was unable to include it in this study, I believe the use of supplementary formant tracking packages such as Fast Track (Barreda 2021) in future research could help to minimise these errors.[1] In the study as run, however, there were a number of systematic errors in formant measurement, particularly for the Level 1 vowels [i u]. The most notable was Praat misidentifying F1 and F2 for [u] as a single broad formant. This led to values of F1, F2 and B1 to be higher than their 'true' values (as would be measured in a human-mediated setup). B1 was the most affected measurement in this regard; the resulting measurements were so positively skewed that I chose to perform a logarithmic transformation on the data before any of the statistical analysis. Indeed, I believe that the mismeasurement of B1 for [u] (the vowel with by far the highest variance) had a significant impact on the results of the statistical tests for Hypothesis 3, which found that the mean ΔB1 for [u] was greater than for [ʊ], indicating that the Level 2 vowel was [+ATR] [o] in underived contexts. This was the only statistical test which significantly indicated that the Level 2 vowels were [+ATR], so given the clear probable cause of the error, I believe that on the whole the evidence still points in favour of an analysis with [−ATR] [ɪ ʊ]. I hope that in the future, with the implementation of more accurate formant tracking, this hypothesis can be revisited to determine if it is indeed an appropriate diagnostic for ATR value.

Finally, I presented the results (both descriptive and inferential) in Chapter 5 and discussed them in the light of the research questions they were intended to answer. I believe that Hypotheses 3–5 present a perfect opportunity for future studies testing their applicability to languages other than Boa-Leboale. Another exciting avenue for future research would be to look for other acoustic correlates which can also be tested in a similar fashion. I hope that, with enough reliable tools at our disposal, the problem of ATR identification in Level 2 vowels will be entirely solvable through acoustic means.

---

[1] Another useful intervention in future studies could be using variable maximum formant frequencies: by setting the formant ceiling lower (e.g. around 3,500Hz) for rounded back vowels like [u], the measurement of F1, F2 and B1 could be improved. This can be done by taking multiple sets of measurements with different format ceilings and selecting the appropriate values *post hoc* depending on the vowel being measured.

Bibliography

Allen, Blake, Douglas Pulleyblank & Ọládiípọ̀ Ajíbóyè. 2013. Articulatory mapping of Yoruba vowels: An ultrasound study. *Phonology* 30(2). 183–210. doi:10.1017/S0952675713000110.

Anderson, Colleen G. 2007. F1 and center of gravity interplay in the maintenance of phonological height within a statistical model of a communal grammar: The case of Foodo [ATR] acoustics. Working paper, The University of Texas at Arlington, Arlington, TX.

Baguley, Thom. 2009. Standardized or simple effect size: What should be reported? *British Journal of Psychology (London, England: 1953)* 100(Pt 3). 603–617. doi:10.1348/000712608X377117.

Baković, Eric. 2000. *Harmony, Dominance and Control.* New Brunswick: Rutgers, The State University of New Jersey. PhD dissertation.

Barreda, Santiago. 2021. Fast Track: Fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard* 7(1). doi:10.1515/lingvan-2020-0051.

Boersma, Paul & David Weenink. 2021. Praat: Doing phonetics by computer, version 6.2.02.

Bretz, Frank, Torsten Hothorn & Peter Westfall. 2016. *Multiple comparisons using R.* CRC Press.

Casali, Roderic F. 2002. Nawuri ATR harmony in typological perspective. *Journal of West African Languages* 29(1). 3–43.

Casali, Roderic F. 2003. [ATR] value asymmetries and underlying vowel inventory structure in Niger-Congo and Nilo-Saharan. *Linguistic Typology* 7(3). doi:10.1515/lity.2003.018.

Casali, Roderic F. 2008. ATR harmony in African languages. *Language and Linguistics Compass* 2(3). 496–549. doi:10.1111/j.1749-818X.2008.00064.x.

Casali, Roderic F. 2016. Some inventory-related asymetries in the patterning of tongue root harmony systems. *Studies in African Linguistics* 96–99. doi:10.32473/sal.v45i1.107254.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the world.* Dallas, TX: SIL International 26th edn. Online version: http://www.ethnologue.com.

Edmondson, Jerold A. & John H. Esling. 2006. The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. *Phonology* 23(02). 157–191. doi:10.1017/S095267570600087X.

Elesin, Abisoye. 2018. Prefix Productivity by Tongue Root in Standard Yorùbá. *Journal of The Linguistic Association of Nigeria Supplement* 3. 248–257.

Fant, G. 1972. Vocal tract wall effects, losses, and resonance bandwidths. *STL-QPSR* 13(2–3). 28–52.

Fulop, S.A., E. Kari & P. Ladefoged. 1998. An acoustic study of the tongue root contrast in Degema vowels. *Phonetica* 55(1-2). 80–98. doi:10.1159/000028425.

Ghitza, O. & J. Goldstein. 1986. Scalar LPC quantization based on format JND's. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34(4). 697–708. doi:10.1109/TASSP.1986.1164907.

Gick, Bryan, Douglas Pulleyblank, Fiona Campbell & Ngessimo Mutaka. 2006. Low vowels and transparency in Kinande vowel harmony. *Phonology* 23(1). 1–20.

Gordon, Matthew & Peter Ladefoged. 2001. Phonation types: A cross-linguistic overview. *Journal of Phonetics* 29(4). 383–406. doi:10.1006/jpho.2001.0147.

Guion, Susan G., Mark W. Post & Doris L. Payne. 2004. Phonetic correlates of tongue root vowel contrasts in Maa. *Journal of Phonetics* 32(4). 517–542. doi:10.1016/j.wocn.2004.04.002.

Halkin, Joseph. 1910. *Les Ababua* (Le Mouvement Sociologique International 11). Brussels: Organe de la société belge de sociologie.

Hess, Susan. 1992. Assimilatory effects in a vowel harmony system: An acoustic analysis of advanced tongue root in Akan. *Journal of Phonetics* 20(4). 475–492.

Hillenbrand, J., R. A. Cleveland & R. L. Erickson. 1994. Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* 37(4). 769–778. doi:10.1044/jshr.3704.769.

Hudu, Fusheini Angulu. 2010. *Dagbani tongue-root harmony : A formal account with ultrasound investigation*. Vancouver: University of British Columbia. PhD dissertation.

Jacobson, Leon C. 1980. Voice-quality harmony in Western Nilotic languages. In Robert M. Vago (ed.), *Issues in vowel harmony proceedings of the CUNY Linguistics Conference on Vowel Harmony, 14th May 1977* (Studies in Language Companion Series), 183. John Benjamins Publishing Company. doi:10.1075/slcs.6.10jac.

Jacobson, Leon Carl. 1978. DhoLuo vowel harmony: A phonetic investigation. Working Paper 43 UCLA Los Angeles, CA.

Kingston, John, Neil A. Macmillan, Laura Walsh Dickey, Rachel Thorburn & Christine Bartels. 1997. Integrality in the perception of tongue root position and voice quality in vowels. *The Journal of the Acoustical Society of America* 101(3). 1696–1709. doi:10.1121/1.418179.

Ladefoged, Peter. 1964. *A phonetic study of West African languages* (West African Language Monographs 1). Cambridge: Cambridge University Press.

Lakens, Daniël, Anne M. Scheel & Peder M. Isager. 2018. Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science* 1(2). 259–269. doi:10.1177/2515245918770963.

Lindau, Mona. 1978. Vowel Features. *Language* 54(3). 541–563. doi:10.2307/412786.

Lindau, Mona. 1979. The feature expanded. *Journal of Phonetics* 7(2). 163–176. doi:10.1016/S0095-4470(19)31047-2.

Lindau, Mona & Peter Ladefoged. 1986. Variability of feature specifications. In Joseph S. Perkell & Dennis H. Klatt (eds.), *Invariance and variability in speech processes*, 464–478. Hillsdale, NJ: Lawrence Erlbaum Associates.

Motingea, Mangulu, André. 2005. *Leboale et lebaate: Langues bantoues du plateau des Uélé, Afrique Centrale* (ILCAA Language Monograph Series 3). Tokyo: Research Institute for Languages and Cultures of Asia and Africa (ILCAA).

Olejarczuk, Paul, Manuel A. Otero & Melissa M. Baese-Berk. 2019. Acoustic correlates of anticipatory and progressive [ATR] harmony processes in Ethiopian Komo. *Journal of Phonetics* 74. 18–41. doi:10.1016/j.wocn.2019.01.004.

Orie, Ọlanikẹ Ọla. 2003. Two harmony theories and high vowel patterns in Ebira and Yoruba. *The Linguistic Review* 20(1). 1–35. doi:10.1515/tlir.2003.001.

Painter, Colin. 1973. Cineradiographic data on the feature ‘covered’ in Twi vowel harmony. *Phonetica* 28(2). 97–120. doi:10.1159/000259449.

Pike, Kenneth L. 1967. Tongue-root position in practical phonetics. *Phonetica* 17(3). 129–140. doi:10.1159/000258583.

Posit team. 2023. RStudio: Integrated development environment for R, version 2023.3.0.386. `http://www.posit.co/`.

Przezdziecki, Marek A. 2005. *Vowel harmony and coarticulation in three dialects of Yoruba: Phonetics determining phonology*. Ithaca, NY: Cornell University. PhD dissertation.

R Core Team. 2023. R: A language and environment for statistical computing, version 4.2.3. `https://www.R-project.org/`.

Reeder, JeDene. 1998. *Pagibete, a northern Bantu borderlands language: A grammatical sketch*. Arlington, TX: The University of Texas at Arlington. MA thesis.

Remijsen, Bert, Otto G. Ayoker & Timothy Mills. 2011. Shilluk. *Journal of the International Phonetic Association* 41(1). 111–125. doi:10.1017/S0025100310000289.

Rose, Sharon. 2018. ATR vowel harmony: New patterns and diagnostics. *Proceedings of the Annual Meetings on Phonology* 5. doi:10.3765/amp.v5i0.4254.

Siffer, Alban, Pierre-Alain Fouque, Alexandre Termier & Christine Largouët. 2018. Are your data gathered?: The folding test of unimodality. In *KDD 2018 - 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Minin*, 2210. doi:10.1145/3219819.3219994.

Starwalt, Coleen Grace Anderson. 2008. *The acoustic correlates of ATR harmony in seven-and nine-vowel African languages: A phonetic inquiry into phonological structure*. Arlington, TX: The University of Texas at Arlington. PhD dissertation.

Stewart, John M. 1967a. A theory of the origin of Akan vowel harmony. In *Proceedings of the Sixth International Congress on Phonetic Sciences*, 863–865. Prague: Publishing House of the Czechoslovak Academy of Sciences.

Stewart, John M. 1967b. Tongue root position in Akan vowel harmony. *Phonetica* 16(4). 185–204. doi:10.1159/000258568.

Thalheimer, W. & Samantha R. Cook. 2002. How to calculate effect sizes from published research: A simplified methodology.

Védy, Louis. 1904. Les Ababua. *Bulletin de la Société royale belge de géographie* 28(3–4). 191–205, 265–294.

de Wit, Gerrit. 2015. *Liko Phonology and Grammar*. Utrecht: LOT.

de Wit, Gerrit. 2020. Boa-Yewu, a Bantu language with a seven-vowel system and ATR vowel harmony. In Jenneke van der Wal, Heleen Smits, Sara Petrollino, Victoria Nyst & Maarten G. Kossmann (eds.), *Essays on African languages and linguistics: In honour of Maarten Mous*, 417–448. Leiden: African Studies Centre Leiden (ASCL).

de Wit, Gerrit. 2022a. The behaviour of high and mid vowels in Boa-Leboale. Paper presented at the 52nd Colloquium on African Languages and Linguistics (CALL 2022), Leiden, 29–31 Aug 2022.

de Wit, Gerrit. 2022b. Boa-Leboale phonology. Unpublished book chapter.

Appendix A

SpectralMeasures.Praat

```
#  Read multiple acoustic measurements (duration, F1, F2, B1, *A1-A2, CoG, CPP) into arrays
#  S K Ahmed, created 06-04-2022
#  S K Ahmed, revised 07-04-2022, fixing A1-A2 measurement using an intermediary 1-to-1 LTAS (on the basis of code by Chad Vicenik, Bert Remijsen, Christian DiCanio) and adding A1-A2 modelling for normalised *A1-A2 output
#  W F L Heeren, revised 25-04-2022, adding form for formant settings, and replacing Strings object by 2nd tier in textgrid
#  G de Wit, revised 12-05-2022, adding separators between fields to facilitate import in Excel
#  S K Ahmed, revised 07-10-2022, adding options for CSV and TSV export, and adding B1 modelling for DeltaB1 output
#  S K Ahmed, revised 20-01-2023, adding CPP measurement
#  S K Ahmed, revised 29-03-2023, adding CPP average across different filters (following Hillenbrand et al. (1994)), changing Spectrogram time window (to Praat standard 30ms), adding heuristics (from Barreda 2021) for F1 and B1 errors
#  S K Ahmed, revised 28-04-2023, parabolic interpolation for A1, A2 measurement and cepstrum slice taken at t2 instead of 0.1s

# Check if a sound and textgrid are selected at the start
if numberOfSelected ("Sound") <> 1 or numberOfSelected ("TextGrid") <> 1
    exit Please select a Sound and a TextGrid first.
endif

# Formant and output settings: How many formants in what frequency range? What should the output look like?

form Settings
    comment Formant settings
    positive nFormants 5
    positive freq_range 5000
```

```
    comment Output settings
    choice Column_separator: 1
        button comma (.csv)
        button tab (.tsv)
        button pipe (.txt)
    text outputfile data
endform

if column_separator$ = "comma (.csv)"
    sep$ = ","
    ext$ = ".csv"

    elsif column_separator$ = "tab (.tsv)"
    sep$ = tab$
    ext$ = ".tsv"

    elsif column_separator$ = "pipe (.txt)"
    sep$ = "|"
    ext$ = ".txt"

    else
    exit Error with column separator.
endif

# Assigning variables to 2 selected objects:
# Sound object = relevant file
# TextGrid object = annotated for the Sound object with intervals labelled
   * marking vowels where measurements should be taken

idSnd = selected ("Sound")
idTG = selected ("TextGrid")
filename$ = selected$ ("TextGrid")

# Preparing output file with headers (use if not accompanied by auxiliary
   script)
# writeFileLine: "'outputfile$''ext$'", "Filename", sep$, "Word", sep$,
   "Time", sep$, "Vowel", sep$, "URvowel" sep$, "ATR", sep$, "MorphGroup",
   sep$, "Duration", sep$, "F1", sep$, "F2", sep$, "B1", sep$, "DeltaB1",
   sep$, "NormA1-A2", sep$, "CoG", sep$, "CPPNoFilt", sep$, "CPPAverage"

# Counting number of intervals marked * in TextGrid object

selectObject: idTG
nVowels = 0
nIntervals = Get number of intervals: 1
for i from 1 to nIntervals
    label$ = Get label of interval: 1, i
    if label$ = "*"
```

```
        nVowels = nVowels + 1
        vowels[nVowels] = i
    endif
endfor

# Creating a spectrogram from Sound object

selectObject: idSnd
idVowelSpecgram = To Spectrogram: 0.03, 11025, 0.005, 5, "Gaussian"

# Cycling through each marked vowel and saving acoustic measurements into
    arrays (and printing in the output file)

for i from 1 to nVowels
    interval = vowels[i]

    # Time measurements

    selectObject: idTG
    starttime[i] = Get start time of interval: 1, interval
    endtime[i] = Get end time of interval: 1, interval
    duration[i] = endtime[i] - starttime[i]
    t1[i] = 0.3 * duration[i] + starttime[i]
    t2[i] = 0.5 * duration[i] + starttime[i]
    t3[i] = 0.7 * duration[i] + starttime[i]

    # Taking a spectral slice at vowel midpoint (t2) and converting to
        LTAS (1-to-1) for A1/A2 readings

    selectObject: idVowelSpecgram
    idVowelSlice = To Spectrum (slice): t2[i]
    idVoweltas = To Ltas (1-to-1)

    # Extracting vowel window without transition effects (between t1 and
        t3)
    # Making Spectrum (for CoG), Formant and Cepstrum (for CPP) objects of
        this range

    selectObject: idSnd
    idVowelSnd = Extract part: t1[i], t3[i], "rectangular", 1, "yes"
    selectObject: idVowelSnd
    idVowelSpecRange = To Spectrum: "no"
    selectObject: idVowelSnd
    idVowelForm = To Formant (burg): 0, nFormants, freq_range, 0.025, 50
    selectObject: idVowelSnd
    idVowelSndBandPass = Filter (pass Hann band): 2500, 3500, 100
    selectObject: idVowelSnd
    idVowelSndHighPass = Filter (pass Hann band): 2500, 0, 100
```

```
#NoFilter
selectObject: idVowelSnd
idCepsgramNF = To PowerCepstrogram: 60, 0.002, 5000, 50
idCepsRangeNF = To PowerCepstrum (slice): t2[i]
cpp_nofilt[i] = Get peak prominence: 60, 333.3, "parabolic", 0.001,
   0.05, "Straight", "Robust slow"
removeObject: idCepsgramNF, idCepsRangeNF

#BandPass
selectObject: idVowelSndBandPass
idCepsgramBP = To PowerCepstrogram: 60, 0.002, 5000, 50
idCepsRangeBP = To PowerCepstrum (slice): t2[i]
cpp_band[i] = Get peak prominence: 60, 333.3, "parabolic", 0.001,
   0.05, "Straight", "Robust slow"
removeObject: idCepsgramBP, idCepsRangeBP

#HighPass
selectObject: idVowelSndHighPass
idCepsgramHP = To PowerCepstrogram: 60, 0.002, 5000, 50
idCepsRangeHP = To PowerCepstrum (slice): t2[i]
cpp_high[i] = Get peak prominence: 60, 333.3, "parabolic", 0.001,
   0.05, "Straight", "Robust slow"
removeObject: idCepsgramHP, idCepsRangeHP

cpp_avg[i] = (cpp_nofilt[i] + cpp_band[i] + cpp_high[i] ) / 3


# Adding word from textgrid tier 2 into the words$[] array

selectObject: idTG
wordIntervalNumber = Get interval at time: 2, t2[i]
words$[i] = Get label of interval: 2, wordIntervalNumber

# Formant measurements (F1, F2 with the range t1-t3, then
   B1/F1mid/F2mid specifically at the midpoint t2 for A1/A2 analysis)

selectObject: idVowelForm
f1[i] = Get mean: 1, t1[i], t3[i], "hertz"
f2[i] = Get mean: 2, t1[i], t3[i], "hertz"
b1[i] = Get bandwidth at time: 1, t2[i], "hertz", "linear"

# Formant measurements at midpoint (for A1, A2 analysis and modelling)

f1mid = Get value at time: 1, t2[i], "hertz", "linear"
f2mid = Get value at time: 2, t2[i], "hertz", "linear"
f3mid = Get value at time: 3, t2[i], "hertz", "linear"
```

```
# Error-reducing heuristics in the vein of Barreda (2021) - to avoid common formant tracking errors

if f1[i] > 1000 or b1[i] > 500 or f1mid > 1000
    # appendInfoLine: "Barreda heuristic condition met: ", words$[i], " f1: ", f1[i], " b1: ", b1[i]

    selectObject: idVowelSnd
    idVowelFormAlt = To Formant (burg): 0, nFormants, 4900, 0.025, 50

    selectObject: idVowelFormAlt
    f1[i] = Get mean: 1, t1[i], t3[i], "hertz"
    f2[i] = Get mean: 2, t1[i], t3[i], "hertz"
    b1[i] = Get bandwidth at time: 1, t2[i], "hertz", "linear"
    f1mid = Get value at time: 1, t2[i], "hertz", "linear"
    f2mid = Get value at time: 2, t2[i], "hertz", "linear"
    f3mid = Get value at time: 3, t2[i], "hertz", "linear"

    if f1[i] > 1000 or b1[i] > 500 or f1mid > 1000
        # appendInfoLine: "Barreda heuristic still met: ", words$[i], " f1: ", f1[i], " b1: ", b1[i]

        selectObject: idVowelSnd
        idVowelFormAlt2 = To Formant (burg): 0, nFormants, 4800, 0.025, 50

        selectObject: idVowelFormAlt2
        f1[i] = Get mean: 1, t1[i], t3[i], "hertz"
        f2[i] = Get mean: 2, t1[i], t3[i], "hertz"
        b1[i] = Get bandwidth at time: 1, t2[i], "hertz", "linear"
        f1mid = Get value at time: 1, t2[i], "hertz", "linear"
        f2mid = Get value at time: 2, t2[i], "hertz", "linear"
        f3mid = Get value at time: 3, t2[i], "hertz", "linear"

        removeObject: idVowelFormAlt2
    endif
    # appendInfoLine: "New values: ", words$[i], " f1: ", f1[i], " b1: ", b1[i]
    removeObject: idVowelFormAlt
endif

# Modelling DeltaB1 following Fant (1972) and Starwalt (2008)

deltab1[i] = b1[i] - (15 * (500 / f1mid) ^ 2 + 20 * (f1mid / 500) ^ 0.5 + 5 * (f1mid / 500) ^ 2)

# Modelling A1-A2 following Fulop, Kari & Ladefoged (1998)
```

```
modelb1 = 30
modelb2 = 80
modelb3 = 150
f = f1mid
modela1_f1 = 20 * log10 ((f1mid ^ 2 + (modelb1 / 2) ^ 2)/(sqrt((f -
    f1mid) ^ 2 + (modelb1 / 2) ^ 2) * sqrt((f + f1mid) ^ 2 + (modelb1 /
    2) ^ 2)))
modela1_f2 = 20 * log10 ((f2mid ^ 2 + (modelb2 / 2) ^ 2)/(sqrt((f -
    f2mid) ^ 2 + (modelb2 / 2) ^ 2) * sqrt((f + f2mid) ^ 2 + (modelb2 /
    2) ^ 2)))
modela1_f3 = 20 * log10 ((f3mid ^ 2 + (modelb3 / 2) ^ 2)/(sqrt((f -
    f3mid) ^ 2 + (modelb3 / 2) ^ 2) * sqrt((f + f3mid) ^ 2 + (modelb3 /
    2) ^ 2)))
modela1_fN = 0.72 * (f / 492) ^ 2 + 0.0033 * (f / 492) ^ 4
modela1_glot = 20 * log10 (2 * ((f / 100) / (1 + (f / 100) ^ 2)))
modela1 = modela1_f1 + modela1_f2 + modela1_f3 + modela1_fN +
    modela1_glot
f = f2mid
modela2_f1 = 20 * log10 ((f1mid ^ 2 + (modelb1 / 2) ^ 2)/(sqrt((f -
    f1mid) ^ 2 + (modelb1 / 2) ^ 2) * sqrt((f + f1mid) ^ 2 + (modelb1 /
    2) ^ 2)))
modela2_f2 = 20 * log10 ((f2mid ^ 2 + (modelb2 / 2) ^ 2)/(sqrt((f -
    f2mid) ^ 2 + (modelb2 / 2) ^ 2) * sqrt((f + f2mid) ^ 2 + (modelb2 /
    2) ^ 2)))
modela2_f3 = 20 * log10 ((f3mid ^ 2 + (modelb3 / 2) ^ 2)/(sqrt((f -
    f3mid) ^ 2 + (modelb3 / 2) ^ 2) * sqrt((f + f3mid) ^ 2 + (modelb3 /
    2) ^ 2)))
modela2_fN = 0.72 * (f / 492) ^ 2 + 0.0033 * (f / 492) ^ 4
modela2_glot = 20 * log10 (2 * ((f / 100) / (1 + (f / 100) ^ 2)))
modela2 = modela2_f1 + modela2_f2 + modela2_f3 + modela2_fN +
    modela2_glot

modela1_a2 = modela1 - modela2

# Formant amplitude measurements (A1, A2, A1-A2, normalised A1-A2)

selectObject: idVoweltas
a1[i] = Get maximum: f1mid - 150, f1mid + 150, "parabolic"
a2[i] = Get maximum: f2mid - 150, f2mid + 150, "parabolic"
a1_a2[i] = a1[i] - a2[i]

norma1_a2[i] = a1_a2[i] - modela1_a2

# Centre of gravity measurement

selectObject: idVowelSpecRange
cog[i] = Get centre of gravity: 2.0
```

```
    removeObject: idVowelSnd, idVowelSpecRange, idVowelSlice, idVoweltas,
        idVowelForm, idVowelSndBandPass, idVowelSndHighPass

    # Recording all measurements into the output file
    appendFileLine: "'outputfile$''ext$'", filename$, sep$, words$[i],
        sep$, t2[i], sep$, sep$, sep$, sep$, sep$, duration[i], sep$,
        f1[i], sep$, f2[i], sep$, b1[i], sep$, deltab1[i], sep$,
        norma1_a2[i], sep$, cog[i], sep$, cpp_nofilt[i], sep$, cpp_avg[i]

endfor

removeObject: idVowelSpecgram
selectObject: idSnd, idTG
```

Appendix B

BoaStats.R

```
#Installing and loading required packages

packages <- c("car", "contrast", "dplyr", "lsr", "moments", "multcomp",
   "phonR", "readxl", "Rfolding", "rstatix", "TOSTER", "ufs", "vioplot")
new.packages <- packages[!(packages %in% installed.packages()[,"Package"])]
if(length(new.packages)) install.packages(new.packages)
lapply(packages, FUN = require, character.only = TRUE)

#Setting up recurring vectors
vowels <- c("i", " ", " ", "a", " ", " ", "u")
measures <- c("f1", "f2", "b1", "deltab1", "norma1a2", "cog", "cppnofilt",
   "cppavg")
measures.dur <- c("duration", "f1", "f2", "b1", "deltab1", "norma1a2",
   "cog", "cppnofilt", "cppavg")
measures.red <- c("f1", "f2", "deltab1", "norma1a2", "cog", "cppavg")

#Importing acoustic measurements from Excel file

BoaData <- read_excel("C:/Users/samue/OneDrive - Universiteit
   Leiden/Documents/@Leiden/5 Boa VH thesis/BoaData-v8.xlsx")

BoaData$vowel <- factor(BoaData$vowel, levels = vowels)

#Descriptive stats

descstats <- function(data, scale.ext) {
  # Plot F1 and F2 for all vowel tokens
  setwd("C:/Users/samue/OneDrive - Universiteit Leiden/Documents/@Leiden/5
     Boa VH thesis")
  png(paste("images/BoaVowels",scale.ext,".png", sep=""), width = 1200,
     height = 1200, res = 150)
  with(data, plotVowels(f1, f2, vowel,
                        var.col.by = vowel,
                        pretty = TRUE,
```

```
                        pch.tokens = vowel,
                        cex.tokens = 0.7,
                        alpha.tokens = 0.2,
                        plot.means = TRUE,
                        pch.means = vowel,
                        cex.means = 1.4,
                        family = "Charis SIL",
                        ellipse.line = TRUE,
                        ellipse.conf = 0.95))
dev.off()

#Save all diamond plots and statistics
sink(file = paste("desc-statistics", scale.ext, ".csv", sep = ""))
paste("Vowel,Measure,Min,LQ,Median,Mean,UQ,Max,Skew,StDev", "\n", sep =
   "") %>% cat()
for (f in 1:length(measures.dur)) {
  msr <- measures.dur[f]
  intervalsM <- data.frame(matrix(nrow = 600, ncol = 7))
  colnames(intervalsM) <- rev(vowels)
  for (v in vowels) {
    data.subset <- subset(data, vowel == v)
    paste(paste(v, msr,
          summary(data.subset[[msr]])[[1]],
          summary(data.subset[[msr]])[[2]],
          summary(data.subset[[msr]])[[3]],
          summary(data.subset[[msr]])[[4]],
          summary(data.subset[[msr]])[[5]],
          summary(data.subset[[msr]])[[6]],
          skewness(data.subset[[msr]]),
          sd(data.subset[[msr]]),
          sep = ",", collapse = NULL), "\n", sep = "") %>% cat();
    data.subset[(nrow(data.subset)+1):600,] <- NA
    intervalsM[[v]] = data.subset[[msr]]
  }
  meansDiamondPlot(intervalsM,
                  jitterHeight = 0.05,
                  jitterWidth = 0,
                  showData = TRUE,
                  conf.level = 0.95,
                  color = "blue",
                  alpha = 0,
                  dataSize = 1,
                  dataAlpha = 0.1,
                  dataColor = "black",
                  xlab = c("Duration (s)", "F1 (Hz)", "F2 (Hz)", "B1
                     (Hz)", "ΔB1 (Hz)", "A1*-A2* (dB)", "Centre of
                     Gravity (Hz)", "Cepstral Peak Prominence - no
                     filter (dB)", "Cepstral Peak Prominence - averaged
```

```
                     across filters (dB)")[f],
                  outputFile = paste("C:/Users/samue/OneDrive -
                     Universiteit Leiden/Documents/@Leiden/5 Boa VH
                     thesis/images/diamond/diamond-", msr, scale.ext,
                     ".png", sep = ""))
}
sink(file = NULL)

# Derived vs underived diamond plots

for (f in 1:length(measures.dur)) {
  msr <- measures.dur[f]
  uvs <- c(" ", " ", " ", " ")
  svs <- c(" ", " ", " ", " ")
  intervalsD <- data.frame(matrix(nrow = 600, ncol = length(uvs)))
  for (n in 1:length(uvs)) {
    data.subset <- subset(data, vowel == svs[n] & urvowel == uvs[n])
    data.subset[(nrow(data.subset)+1):600,] <- NA
    intervalsD[n] = data.subset[[msr]]
    colnames(intervalsD)[n] <- paste("/", uvs[n], "/ [", svs[n], "]",
       sep = "")
  }
  meansDiamondPlot(intervalsD,
                   jitterHeight = 0.05,
                   jitterWidth = 0,
                   showData = TRUE,
                   conf.level = 0.95,
                   color = "blue",
                   alpha = 0,
                   dataSize = 1,
                   dataAlpha = 0.1,
                   dataColor = "black",
                   xlab = c("Duration (s)", "F1 (Hz)", "F2 (Hz)", "B1
                      (Hz)", "ΔB1 (Hz)", "A1*-A2* (dB)", "Centre of
                      Gravity (Hz)", "Cepstral Peak Prominence - no
                      filter (dB)", "Cepstral Peak Prominence - averaged
                      across filters (dB)")[f],
                   outputFile = paste("C:/Users/samue/OneDrive -
                      Universiteit Leiden/Documents/@Leiden/5 Boa VH
                      thesis/images/deriv/diamond-deriv-", msr,
                      scale.ext, ".png", sep = "")
  )
}

#Save all boxplots and violin plots

for (f in measures) {
  png(paste("images/boxplots/boxplot-", f, scale.ext, ".png", sep = ""),
```

```
    width = 1200, height = 1200, res = 150)
  boxplot(data[[f]] ~ data$vowel, data = data, notch = TRUE, xlab =
    "Vowel", ylab = f)
  dev.off()
  data$vowel <- factor(data$vowel, levels = rev(vowels))
  png(paste("images/vioplots/vioplot-", f, scale.ext, ".png", sep = ""),
    width = 1200, height = 1200, res = 150)
  with(data, vioplot(formula = formula(paste(f, "~ vowel")),
    col=rgb(0.1,0.4,0.7,0.7) , names=rev(vowels), horizontal = TRUE,
    side = "right"))
  dev.off()
  data$vowel <- factor(data$vowel, levels = vowels)
 }
}

#Unadjusted values/plots

descstats(data = BoaData, scale.ext = "")

#Adjustment

BoaData <- BoaData %>%
 mutate(
  derived = if_else(
   vowel == urvowel, "underived", "derived"
  )
 )

BoaScaled <- BoaData
BoaScaled$b1 <- log10(BoaData$b1)
BoaScaled$deltab1 <- log10(BoaData$deltab1 - min(BoaData$deltab1) + 5)
BoaScaled$cog <- log10(BoaData$cog)

#Adjusted values/plots

descstats(data = BoaScaled, scale.ext = "-trsfm")

#Hypothesis 1

sink(file = "h1-unimodality.csv")
 "Measure ,,\n" %>% cat();
 h1.F <- folding.statistics(data.matrix(subset(BoaScaled, vowel ==
   " ")[,measures.red]))
 h1.B <- folding.statistics(data.matrix(subset(BoaScaled, vowel ==
   " ")[,measures.red]))
 paste("all,", h1.F, ",", h1.B, "\n", sep = "")  %>% cat();
 for (m in measures.red) {
  phi.F <- folding.statistics(data.matrix(subset(BoaScaled, vowel ==
```

```
       " ")[[m]]))
    phi.B <- folding.statistics(data.matrix(subset(BoaScaled, vowel ==
       " ")[[m]]))
    paste(m, ",", phi.F, ",", phi.B, "\n", sep = "") %>% cat();
  }
  paste("n,", nrow(subset(BoaScaled, vowel == " ")), ",",
     nrow(subset(BoaScaled, vowel == " ")), "\n", sep = "") %>% cat();
sink(file = NULL)

#Hypothesis 2

sink(file = "h2-equivalence.csv")
  "Vowel,Measure,meanderiv,meanundrv,sdderiv,sdundrv,nderiv,nundrv,df,TOSTLower-t,TOSTUp
     %>% cat();
  tost <- list()
  for (v in c(" ", " ")) {
    for (m in measures) {
      dd <- subset(BoaScaled, vowel == v & derived == "derived")[[m]]
      ud <- subset(BoaScaled, vowel == v & derived == "underived")[[m]]
      tost[v] <- tsum_TOST(m1 = mean(dd), m2 = mean(ud), sd1 = sd(dd), sd2
         = sd(ud), n1 = length(dd), n2 = length(ud), eqb = 0.880, alpha =
         0.05, var.equal = FALSE, eqbound_type = "SMD")
      paste(v, m, mean(dd), mean(ud), sd(dd), sd(ud), length(dd),
         length(ud), tost[v][[1]]["t-test","df"], tost[v][[1]]["TOST
         Lower","t"], tost[v][[1]]["TOST Upper","t"],
         tost[v][[1]]["t-test","t"], tost[v][[1]]["TOST Lower","p.value"],
         tost[v][[1]]["TOST Upper","p.value"],
         tost[v][[1]]["t-test","p.value"], sep = ",")  %>% cat();
      "\n" %>% cat();
    }
  }
sink(file = NULL)

#Hypotheses 3-7

BoaScaled.U <- subset(BoaScaled, (vowel != " " & vowel != " ") | derived ==
   "underived")
BoaScaled.D <- subset(BoaScaled, (vowel != " " & vowel != " ") | derived ==
   "derived")

h3h7.contrasts <-  c(c(1, -1,  0,  0,  0,  0,  0),
                      c(0,   0,   0,   0,   0,  -1,   1),
                      c(1,  -2,   1,   0,   0,   0,   0),
                      c(0,   0,   0,   0,   1,  -2,   1),
                      c(0,   0,   1,   0,   0,  -1,   0))

h3h7.contrast.matrix <- matrix(h3h7.contrasts, nrow = 5, ncol = 7,
   byrow=TRUE)
```

```
colnames(h3h7.contrast.matrix) <- vowels
rownames(h3h7.contrast.matrix) <- c("ivsF", "uvsB", "Fcomp", "Bcomp",
   "EvsB")

models.U <- data.frame(matrix(nrow = 30))
models.D <- data.frame(matrix(nrow = 30))
for (m in measures){
  models.U[[m]] <- glm(formula(paste(m, "~ vowel")), data=BoaScaled.U,
     family='gaussian')
  models.D[[m]] <- glm(formula(paste(m, "~ vowel")), data=BoaScaled.D,
     family='gaussian')
}
h3h7 <- data.frame(matrix(nrow = 10))

sink(file = "h3h7-linear.csv")
  "Test,vowel,deriv,zvalue,pvalue\n"  %>% cat();
  h3h7$h3U <- summary(glht(model = models.U$deltab1, linfct = mcp(vowel =
     h3h7.contrast.matrix[c("ivsF", "uvsB"),])),
     test=adjusted("single-step"))
  h3h7$h4U <- summary(glht(model = models.U$norma1a2, linfct = mcp(vowel =
     h3h7.contrast.matrix[c("ivsF", "uvsB"),])),
     test=adjusted("single-step"))
  h3h7$h5U <- summary(glht(model = models.U$cog, linfct = mcp(vowel =
     h3h7.contrast.matrix["Fcomp",])), test=adjusted("single-step"))
  h3h7$h6U <- summary(glht(model = models.U$cppavg, linfct = mcp(vowel =
     h3h7.contrast.matrix[c("Fcomp", "Bcomp"),])),
     test=adjusted("single-step"))
  h3h7$h7U <- summary(glht(model = models.U$deltab1, linfct = mcp(vowel =
     h3h7.contrast.matrix[c("EvsB"),]), alternative = "greater"),
     test=adjusted("single-step"))
  h3h7$h3D <- summary(glht(model = models.D$deltab1, linfct = mcp(vowel =
     h3h7.contrast.matrix[c("ivsF", "uvsB"),])),
     test=adjusted("single-step"))
  h3h7$h4D <- summary(glht(model = models.D$norma1a2, linfct = mcp(vowel =
     h3h7.contrast.matrix[c("ivsF", "uvsB"),])),
     test=adjusted("single-step"))
  h3h7$h5D <- summary(glht(model = models.D$cog, linfct = mcp(vowel =
     h3h7.contrast.matrix["Fcomp",])), test=adjusted("single-step"))
  h3h7$h6D <- summary(glht(model = models.D$cppavg, linfct = mcp(vowel =
     h3h7.contrast.matrix[c("Fcomp", "Bcomp"),])),
     test=adjusted("single-step"))
  h3h7$h7D <- summary(glht(model = models.D$deltab1, linfct = mcp(vowel =
     h3h7.contrast.matrix[c("EvsB"),]), alternative = "greater"),
     test=adjusted("single-step"))
  for (h in c("h3U", "h4U", "h6U")) {
    mdl <- h3h7[[h]]$test
    for (t in 1:2) {
      paste(h, c("F", "B")[t], "underived", mdl$tstat[[t]],
```

```
         mdl$pvalues[t], sep = ",")  %>% cat();
      "\n" %>% cat();
    }
  }
  for (h in c("h5U", "h7U")) {
    mdl <- h3h7[[h]]$test
    paste(h, "X,underived", mdl$tstat[[1]], mdl$pvalues[1], sep = ",")
       %>% cat();
    "\n" %>% cat();
  }
  for (h in c("h3D", "h4D", "h6D")) {
    mdl <- h3h7[[h]]$test
    for (t in 1:2) {
      paste(h, c("F", "B")[t], "derived", mdl$tstat[[t]], mdl$pvalues[t],
         sep = ",")  %>% cat();
      "\n" %>% cat();
    }
  }
  for (h in c("h5D", "h7D")) {
    mdl <- h3h7[[h]]$test
    paste(h, "X,derived", mdl$tstat[[1]], mdl$pvalues[1], sep = ",")  %>%
       cat();
    "\n" %>% cat();
  }
sink(file = NULL)

sink(file = "hX-summaries.txt")
  #h2
  "Hypothesis 2\n" %>% cat()
  print(tost[" "])
  print(tost[" "])
  #h3h7
  "Hypothesis 3 (DeltaB1)\n" %>% cat()
  print(h3h7$h3U)
  print(h3h7$h3D)
  "Hypothesis 4 (A1*-A2*)\n" %>% cat()
  print(h3h7$h4U)
  print(h3h7$h4D)
  "Hypothesis 5 (CoG)\n" %>% cat()
  print(h3h7$h5U)
  print(h3h7$h5D)
  "Hypothesis 6 (CPP)\n" %>% cat()
  print(h3h7$h6U)
  print(h3h7$h6D)
  "Hypothesis 7 (DeltaB1 again)\n" %>% cat()
  print(h3h7$h7U)
  print(h3h7$h7D)
sink(file = NULL)
```