Spring 2022

Archival Phonetics & Prosodic Typology in Sixteen Australian Languages

Sarah Babinski
Yale University Graduate School of Arts and Sciences, sarah.babinski@yale.edu

Abstract

Archival Phonetics & Prosodic Typology in Sixteen Australian Languages

Sarah Babinski

2022

In naturalistic speech, the phonetic instantiation of phonological categories is often highly variable (Cohn & Renwick 2021). Speakers have been observed to converge on patterns of phonetic variation that are consistent within languages but variable cross-linguistically for the same phonological phenomenon (Kakadelis 2018). Speakers are evidently sensitive to these sorts of patterns and learn the phonetic variation in a consistent way. Furthermore, the systematicity of this variation suggests that these patterns should change over time systematically as well. Most Australian languages assign lexical stress consistently on the first syllable of the word (Fletcher & Butcher 2014), raising the question of how the phonetics of stress varies across languages with this phonologically stable pattern.

This dissertation presents an investigation into structured variation of the acoustic correlates of stress and prosody in sixteen Indigenous languages of Australia that all have consistent initial stress placement, with a focus on the source(s) of variation in these factors cross-linguistically. Acoustic correlates of stress, despite the phonological uniformity present among these languages, show significant cross-linguistic variation, both in the presence or absence of a particular cue to stress, as well as the size of these effects. The phonological uniformity of stress assignment allows for a more controlled comparison of the acoustic correlates of stress across these languages, since the placement of stress marking remains constant. Acoustic correlates investigated are vowel duration, pre-tonic and post-tonic consonant duration, intensity, f0 (maximum and range), and vowel peripherality. These cues are identified using a series of mixed effects linear regression models.

To identify the source(s) of variation in acoustic correlates to stress, the population genetics tool Analysis of Molecular Variance (AMOVA) is used. This is a statistical tool created for analysis of genetic variance that has been applied to cultural evolution topics such as music (Rzeszutek, Savage & Brown 2012) and folktales (Ross, Greenhill & Atkinson 2013). This model finds significant variation across languages, as well as substantial intra-speaker variation, similarly to the findings for both biological and cultural evolution, but no significant intra-language variation across speakers. These results are also supported by the investigation of inter- and intra-language variation using regresssion modeling.

Another population genetics measure, fixation index or F_{ST}, is used to create a network model of language relationships based on the phonetic correlates of lexical stress. This network shows clear relationships between the Pama Nyungan languages in this sample, as well as some Gunwinyguan languages, supporting the claim that the phonetic cues to stress are stable within language families and change according to the principles of diachronic language change. Smaller groupings in this network also indicate some contact-induced change or areal effects in these phonetic markers.

Phrasal prosody is also investigated in this dissertation, using a toolkit for automated phrasal contour clustering from Kaland (2021). For each language, f0 is measured at regular intervals across the word, which is used as input to a complete-linkage clustering algorithm to identify major categories of phrasal contours. Results of this sort of automatic clustering provide testable hypotheses about phrasal types in each language, while avoiding some common pitfalls of impressionistic analyses of prosodic phrases. As with the investigation into lexical stress, this sort of automated typological work serves as a crucial complement to more detailed language-specific studies for the creation of well-rounded and well-supported theories.

The data used in this dissertation are narrative speech recordings sourced from language archives, collected in varying field settings. In processing these data I have created a large

corpus of these recordings force aligned at the segment level and have worked out *post-hoc* methods for controlling noise and variation in field-collected audio to create a comparable set of language data. I include in the dissertation a lengthy discussion of these methods, with the aim of providing a practical toolkit for the use of archival materials to address novel phonetic questions, as well as to aid in the creation of language revitalization resources.

Archival Phonetics & Prosodic Typology in Sixteen Australian Languages

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Sarah Babinski

Dissertation Director: Dr. Claire Bowern

May 2022

Contents

List of Figures

Acknowledgements

As I have begun to approach this milestone, I have been reflecting on all the events, and especially the people, that have helped me get to this point. I have been very fortunate to have had the support of countless family members, friends, and colleagues over the years. I can only thank some of them here, but know that you are all in my heart.

First and foremost, I would like to thank Claire Bowern for truly being the best advisor any graduate student could hope for. Claire has been a fount of knowledge, a voice of reason, and a source of comfort throughout my graduate program, and I know that I would have been much worse off were it not for her guidance. Meetings with Claire always leave me feeling more clear-headed and confident in what I am doing than when I entered, and she has been a constant presence providing me with the advice, feedback, and structure that have helped me reach this point. I cannot thank her enough for advising me.

Jason Shaw and I both began our time at Yale in Fall 2016, which means I was not expecting to have a phonetician on the faculty when applying to the graduate program. I am very lucky that this was indeed the case, and especially that this phonetician happened to be Jason. He has helped to strengthen my work in this dissertation in many ways, as well as my qualifying papers, term papers, and knowledge of phonetics more generally. I am also very grateful to Jason for providing the cookies that have gotten me through many a Friday afternoon in Phonologroup.

I first met Natalie Weber when we were both graduate students, at a small conference. I was a second-year student and knew no one there, but Natalie was very welcoming and made sure to invite me out with the other students. I was always grateful for their kindness and was very happy when they joined the faculty at Yale a few years later. They have continued to be a kind presence in our department and have provided me with great advice and comments as a member of my committee.

Many thanks also to Doug Whalen, who joined my committee later in the process but has a been a wonderful source of practical wisdom. I have greatly appreciated his comments on my dissertation as well as other work I have presented in Phonologroup over the years, and especially his perspective on the field and on endangered language work.

During my time at Yale, I have been fortunate to have made many great friends and colleagues in the Linguistics Department. I would like to thank Yale graduate students and affiliates past a present, especially Sammy Andersson, Ushasi Banerjee, Rikker Dockum, Marisha Evans, Martin Fuchs, Chris Geissler, Luke Lindemann, Randi Martinez, Josh Phillips, Sigríður Saeunn Sigurðardóttir, Catarina Soares, Mike Stern, Matt Tyler, Irene Yi, and Andy Zhang. Special thanks to all those in the department who have been cheering me on in the final days before submitting this dissertation.

My journey, in graduate school and in life, has been infinitely richer because of my friends near and far who have been there with me. Adriana, Bill, Olivia, and Veda, thank you for your friendship over the past decade and for filling my life with joy, laughs, comfort, and daily memes. Thanks especially to Lewis, who has been friend, classmate, and colleague for many years, and who has been there to commiserate with me ever since high school Latin class. And thank you to Cristian, for providing so much encouragement and soup.

I would never have gotten to Yale if it were not for the love and support of my family. My parents, MaryAnn and James Babinski, have supported me unconditionally in everything I do, and have truly gone above and beyond to help me achieve my goals. My sister, Caroline, has been by my side unwaveringly since the beginning. And I am fortunate to have many extended family members to thank as well, especially my grandparents, Barbara and Paul Babinski, and the late Irene and Edmund Mackarevich; my aunts Debbie, Denise, and Dorette; my cousin Katie; and my goddaughter Madelyn, who has brought us all so much joy.

To my parents

Chapter 1

Introduction

This dissertation's analytical focus is on the phonetics of prosodic phenomena, primarily lexical stress, in sixteen Indigenous languages of Australia. The contribution of this aspect of the dissertation is threefold: (1) increasing accessibility of prosodic research, (2) providing the first study of this kind for many of the languages under investigation, and (3) advancing the study of the typology and diachrony of prosodic phenomena generally, as well as that of Australian languages specifically. It contributes to the further documentation of many Australian languages' prosody, an area of study that is often overlooked in under-resourced languages (Macaulay 2021). I hope that this project will help fellow researchers, as well as community language members and teachers, to better understand the lexical stress system, phonetic variation, and phonology more broadly, of their languages of interest.

Studies on the prosody of words and phrases has often required the creation of targeted experiments and the collection of novel data, making such studies less accessible for many languages with few or no speakers. Likewise, the nature of the methodologies used and language knowledge needed for these types of experiments reduces the accessibility of such work for researchers who do not specialize in phonetics or phonology, or those who are not

highly knowledgeable in the language of interest. This dissertation addresses this issue by demonstrating methods that are largely automated using open source software, removing the barrier of advanced language knowledge, and making use of existing spontaneous speech audio to eliminate the need (at least at first) for novel data collection.

This project makes the case for using natural speech audio from archival sources in doing work on under-researched languages. Establishing best practices for this sort of research, aside from being useful for the community of academic linguists, also contributes to development of automatic speech recognition including speech-to-text technologies, for under-resourced languages with only small data sets. Automatic transcriptions of natural speech data can contribute to community-oriented efforts for captioned videos, spoken-word phrasebooks, and other pedagogical materials in a quick and cost-effective way. This capability could be a vital resource in endangered language communities' revitalization efforts.

To analyze the acoustics of stress and prosody in these languages, I create time-aligned transcripts of natural speech archival audio materials using automatic forced alignment techniques, offering a practical example of the utility of such tools when used on field-based audio that is noisy, i.e. includes background noise, variable sound quality, and utterances in several languages. As a result of this work, I have made the archival materials from AIATSIS (`https://aiatsis.gov.au`), ELAR (`https://elar.soas.ac.uk`), and PARADISEC (`https://paradisec.org.au`) more accessible by providing these entries with addenda containing time-aligned segment-by-segment TextGrids. This will minimize processing time for future researchers and will allow for these archival materials to be investigated for years to come.

In order to look at lexical stress comparatively across languages, individual analyses are conducted for each of the sixteen languages investigated to maintain consistent methods for comparison. Most of these languages have had little, if any, prosodic description in

previous published work. This dissertation therefore presents the first substantial study of prosody in these languages, opening up many avenues of further research which are discussed along with the results. Comparative studies of stress, especially the phonetics of stress, are uncommon, as are studies of phonetic-prosodic change (cf. Whalen, DiCanio & Dockum 2020, Macaulay 2021). This project provides such a study by investigating the acoustic correlates of lexical stress, as well as the major phrasal pitch contours. The results are analyzed for their cross-linguistic variation as well as potential pathways of diachronic change of stress marking.

This chapter is structured as follows. First, the definition of *stress* and other prosodic phenomena as they are conceptualized in this dissertation is discussed in §1.1. The main claims of the dissertation are outlined in §1.3, followed by some background on phonetic precursors to sound change in §1.2. Then, a brief overview of the languages of Australia generally is presented in §1.4. A discussion of archival phonetic methods is presented in §1.5. The chapter ends with a summary of the dissertation structure (§1.6).

1.1 Conceptualization of stress and prosody

The term *stress* is used in this dissertation to mean the particular subset of prominence phenomena that apply at the level of the word. Thus, stress is a type of prominence, one variety of prosody along with higher-level phenomena that apply at the level of the phrase or utterance. Both stress and phrasal prosody are investigated in this dissertation, although the depth of study that falls within the scope of this project differ between them, as is discussed later in this section. Greater focus is placed here, as overall, on stress.

1.1.1 Stress is relative

Linguistic prominence as a general category is defined purely in relative terms — prominent syllables may be longer, louder, more peripheral, etc. all in relation to their non-prominent

counterparts (Gordon 2011). Prominence is a unique phenomenon in that the phonetic factors used in prominence marking differ across languages to the point where two languages may have entirely non-overlapping phonetic definitions of it; these factors are usually some combination of a limited set — duration, intensity, pitch, peripherality, spectral tilt, and few others (Gordon & Roettger 2017, Van Heuven 2018). This differs from other linguistic contrasts which are cued by a more constrained set of phonetic factors, such as Voice Onset Time for laryngeal features (Abramson & Whalen 2017), formants for vowel quality (Garellek 2019), and so on.

Despite this heterogeneity in the phonetic markers to the phenomenon we call *stress,* each language treats its phonetic correlates of stress just like the phonetic markers of any other phenomenon. For example, the English stop voicing contrast consists of two sets of overlapping allophones, such as voiceless /p^h p/ and voiced /p b/. Aspiration is lost on the voiceless stop in certain phonological contexts (e.g. in CC clusters like *spark*), as is true voicing in another context (e.g. word-initially, as in *bark*). This can lead to phonetically very similar stop segments, but there is no true stop neutralization because of the differing phonological contexts in which these allophones occur. Likewise, stress can be considered part of the set of suprasegmental 'phonemes' along with phrasal prosody, phonemic vowel length, lexical tone, boundary lengthening phenomena, and more. While phonetically speaking these elements may be quite similar, the context in which they occur can help us to disambiguate. A stressed syllable may not be the longest syllable in the word, because of the presence of boundary lengthening, or because the stressed syllable happens to fall on a phonemically short vowel. But duration may still be a correlate of stress because of its length relative to unstressed short vowels of the same type, and not necessarily because stressed syllables are globally the longest elements in a given language. Thus the phonetic cues to stress are also defined in relative terms, as it is only the contrast between stressed and unstressed phonetic markers that matters, instead of some quantitative threshold.

1.1.2 Phonological factors

The phonological conceptualization of stress is largely focused on the patterns of stress placement a language employs. The phonological stress literature in linguistics began with Metrical Stress Theory in the 1970s (Hayes 1995, Prince 1990), and has been brought into Optimality Theoretic frameworks from the beginning (Kager 1999, McCarthy & Prince 1993). The main focus of the theoretical work on stress phonology is on rhythmic stress assignment rules, regardless of the phonetic realization of stress in the language.

A phonological typology of stress systems can be encapsulated via the interaction of a few dimensions. First, rhythmic feet are binary (two-syllable) and either headed on the left (trochees) or the right (iambs) (Hayes 1985, Gordon 2011). These feet are then usually attached to either the left or right edges of words, which generates initial, peninitial, penultimate, and final stress assignment patterns. Constraints preferring stress to fall on heavy syllables, as well as constraints against edge-aligned feet, are usually employed to generate stress assignment patterns further from the word edge (McCarthy & Prince 1993).

In most Australian languages, stress has been described as trochaic, left-aligned, and quantity insensitive (Baker 2014, Fletcher & Butcher 2014, Goedemans 2010). Stress falls on the initial syllable of the word, or perhaps the stem, although these two are equivalent most of the time. There are some cases, especially in prefixing non-Pama Nyungan languages, of stress being attached to the root leaving prefixes unstressed (cf. Baker 2014: 156). A small number of Australian languages spoken in Arnhem Land have been reported to have right-aligned trochees (penultimate stress), though this pattern is rare (Goedemans 2010: 72).

1.1.3 Phonetic factors

Duration, intensity, and fundamental frequency (f0) have been identified as potential correlates of stress since at least Fry (1958), whose experiments on English isolated each of these as salient stress cues. Since then, a number of other phonetic factors have been found to mark stress in some languages, including: onset consonant duration; post-tonic consonant duration; peripherality of vowels; spectral tilt; and others (Lehiste 1970, Van Heuven 2018).

Less typological work has been done looking at cross-linguistic variation in the phonetic correlates of stress than has been done for the phonological assignment of it. However, Gordon & Roettger (2017) is a notable example of this type of study, looking at acoustic stress correlates identified in 110 published studies of 75 languages. The authors find in their survey that measurements of duration, f0, intensity, vowel formants, and spectral tilt correlate with stress in between 70 to 85% of the studies that measure them. However, there is a marked difference in the number of studies that look at each of these measures, which could be a confound of this finding; for example, duration is measured in 100 of the studies Gordon and Roettger consider, while spectral tilt is only measured in 22. In addition to this, the ways that each of these factors were measured in the different studies varied, making it less clear how to compare these results directly.

The acoustic measurement that is most often found to correlate with stress in Australian languages is f0 (Goedemans 2010, Fletcher & Butcher 2014). However, many Australian languages have no phonetic description of stress, so whether this observation generalizes across Australia is certainly an open question. In addition to f0, acoustic dimensions that have been found to correlate with stress in at least some Australian languages include duration, intensity, onset duration, post-tonic consonant duration, vowel quality, and spectral tilt (cf. Fletcher et al. 2015, Jepson, Fletcher & Stoakes 2019, Simard 2010, Bishop 2003,

Fletcher & Evans 2002).

1.1.4 Phrase-level prosody

Layered over word-level stress marking is phrase-level prosody, which is marked by intonational patterns and most often correlated with f0. It is used to indicate "paralinguistic meanings" such as the type of speech act, e.g. declarative versus interrogative, discourse functions such as introduction of a new topic, or holding the floor in a conversation (Xu 2019). Prosodic typology has been the subject of significant work, but the exact patterns and contour types can vary widely, made up of different combinations of high and low tones, falls and rises.

Prosodic phenomena above the word have been studied systematically in linguistics since at least Pierrehumbert (1980), which looked at phrase structure in English. A seminal work connecting intonation patterns with syntactic phrase rules came from Nespor & Vogel (1986), which established the commonly-used terms of the phonological word, phonological phrase, and intonational phrase. The analysis of these phenomena have been instantiated in a number of ways, including Autosegmental Metrical Theory (Ladd 1996). A standard transcription system for prosodic units, ToBI, exists for English and a handful of other commonly-researched languages (Beckman, Hirschberg & Shattuck-Hufnagel 2005).

Phrase-level prosodic description in Australian languages is even more rare than word-level stress descriptions, reflective of the broader trend of this type across language documentation work in all areas of the world (Fletcher & Butcher 2014, Whalen, DiCanio & Dockum 2020, Macaulay 2021). However, some work exists for languages such as Kayardild (Fletcher, Evans & Round 2002), Arrernte (Tabain 2016), Jaminjung (Simard 2010), Bininj Gun-wok (Bishop 2003), Djambarrpuyngu (Jepson 2019), and Dalabon (Ross, Fletcher & Nordlinger 2016), among others. Fletcher, Evans & Round (2002) note that while phrasal prosody in Bininj Gun-wok has been found to function exclusively as a way to demarcate

the edges of phrasal units (Bishop 2003), Kayardild seems to use phrasal prosody for both this demarcative function as well as phrasal prominence functions such as focus marking. While subsequent work on phrasal prosody in individual Australian languages has since been published, to my knowledge any comprehensive typology of Australian prosody has yet to be written.

1.2 Phonetic precursors to sound change

Ohala (1993) characterizes phonetic influence on sound change as instances of misperception (hypercorrection and hypocorrection). These usually result in the phonologization of chance phonetic distinctions that were previously by-products of some other phonological contrast, e.g. confusion of coarticulatory effects resulting in a change from /anpa/ to /ampa/. Further work has since been published on the role of the speaker in these sorts of changes, finding that some speakers are more likely to be innovators and early adopters of change due to a high social awareness, while other speakers' cognitive styles make them more closely attuned to coarticulatory effects in the speech signal and thus less likely to make perceptual mistakes (Garrett & Johnson 2013, Yu et al. 2011, A. C. L. Yu 2010). This work has led others to propose overarching models of sound change that take these individual differences into account. Blevins (2004), Wedel (2006), and others have laid out models of sound change with a perspective borrowed and modified from research on biological change. This work conceptualizes sound change as an evolutionary system, in which the inherent noise and variation in language serve as the seeds of linguistic change. As language is transmitted across generations as well as in everyday communicative interactions, the innovators and early adopters drive a slow shift in a particular linguistic phenomenon that is eventually phonologized and recognized as a change.

This dissertation explores the influence of phonetic variation, specifically acoustic variation correlated with stress and other types of prominence, on sound change. The crucial

distinction between this work and the cases Ohala (1993) focuses on is that this type of phonetic variation is already phonologized in one domain, namely the domain of stress. The basic theory of change used will be the evolutionary models of Blevins (2004) and Wedel (2006), by assuming that change essentially draws from a pool of synchronic phonetic variation. Certain phonological and phonetic factors surrounding lexical stress in a language may 'tip the scales,' making it more likely for some changes to happen over others. For this reason, a typological study of phonetic variation is crucial to this project. Variation across languages is presented in Chapter 4, while within-language speaker variation is discussed in Chapter 5.

This project has much to contribute to current research on language change and language contact, particularly to the study of sound change. The relative uniformity of stress assignment systems in Australian languages make them an ideal group of languages to study for insights into the role(s) the phonetics of stress play in language change. Through this large-scale comparison of many Australian languages, variation is explored both within and across languages, and this variation is quantified in Chapter 6. This sort of study has not been performed in any previous research, and there are no existing theories of which I am aware that propose any principles by which lexical stress or the acoustic cues that mark it undergo diachronic changes. This is a clear gap in linguistic research that this dissertation will begin to fill, drawing on literature concerning the phonetics of lexical stress as well as existing research concerning the phonetic precursors of sound change.

1.3 The current study: Claims

(1.1) The phonetic factors that cue linguistic prominence are linguistically heritable, meaning that they vary and change in similar ways to phonemes. They vary in a structured way within a language and will remain relatively stable until a change occurs, in a way that is analogous to phonological change. For this claim to be true,

the following must hold:

a. Prominence cues are consistent across speakers of the same language. Speakers converge on the same cues to prominence and use these to the exclusion of other potential cues. (This point already has lots of evidence.)

b. Closely related languages will be more likely to share cues to prominence than languages that are more distantly related. This should be the most clearly observed when languages have the same pattern of stress assignment, as changes in stress position will increase the likelihood of cue changes.

c. Study of population structure will show that significant variation exists across languages, separately from any within-language or within-speaker variation.

(1.2) Just like all other linguistically variable phenomena, the phonetic cues to stress can also vary along sociolinguistic lines within a language. In order for this to hold, the following must be true:

a. Some cues to stress within a language may only be cues for some speakers, and not for others. Similarly, speakers may vary in their use of these cues based on the social situation.

b. This variation falls along defineable social lines, such as gender identity, dialect, social status, register, etc. (This cannot be studied well with these data)

c. Study of population structure will show that significant variation exists across speakers within a language, separately from any cross-linguistic or within-speaker variation.

(1.3) Different prominence cues may co-occur with one another, marking the same type of prominence with multiple acoustic factors. However, the presence of multiple cues may make each individual factor more unstable in the system, as the crucial contrast would not be lost with the loss of one cue. Some cues may hold for all

speakers in the language, while others may be sociolinguistically variable.

1.4 Australian languages

1.4.1 Overview

Bowern (2011) estimates 363 languages were spoken in Australia at the time of European contact. Out of these languages, 275 (about 75%) are members of the Pama Nyungan language family (see Figure 1.1). Pama Nyungan is the major family of the continent, while the remaining languages are either part of very small families or are language isolates. However, even most languages that are outside Pama Nyungan in Australia share several linguistic traits that make these languages look similar. For example, phonemic inventories across Australian languages are largely uniform (Fletcher & Butcher 2014, Gasser & Bowern 2014). Vowel inventories are usually sparse, stops do not have voicing distinctions, and retroflex consonants are common across these languages, regardless of whether they are historically related. These facts, among others, have led researchers to claim that traits only directly observed in a handful of Australian languages apply to all languages on the continent, leading to overgeneralizations that ignore clear variation that would be found if these languages were all studied on their own terms.

Despite surface-level similarity in the phoneme inventories of Australian languages, previous work suggests that this does not preclude the existence of rich phonetic variation. Gasser & Bowern (2014) found considerable variation in the phonotactics of Australian languages, e.g. differences in phoneme frequencies and minimal word length requirements. Similarly, Kakadelis (2018) conducted a phonetic study of three unrelated languages without a stop voicing distinction (Bardi, Arapaho, and Nahuatl). She found that while these languages have the superficial phonemic similarity of 'no stop voicing distinction,' they actually vary substantially in average oral stop segment duration, lenition, and voice on-

set time (VOT). These languages differed in many non-phonologized patterns of phonetic variation in statistically significant ways. For example, Bardi showed a tendency toward phonetically voiced stops in intervocalic and intersonorant contexts, while Nahuatl showed the opposite effect, with more phonetically devoiced stops in these contexts.

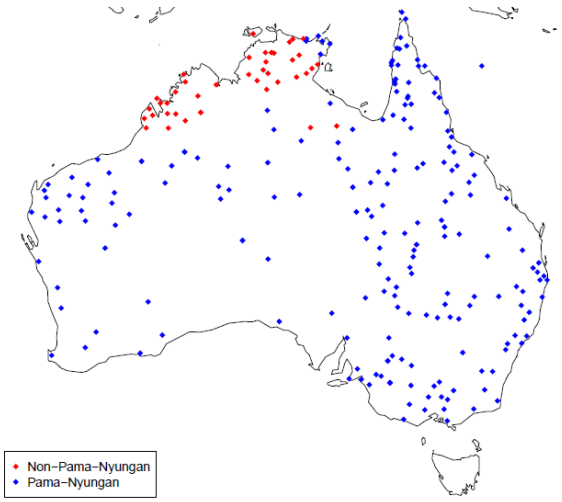

Figure 1.1: Map of Australian languages, from Chirila database (Bowern 2016).

Most Australian languages (about 80%) have consistent initial lexical stress (Fletcher & Butcher 2014, Goedemans 2010). However, as has been found in other cases of superficial uniformity, the phonetic variation in the correlates of lexical stress in these languages is substantial. The claim that is often made about Australian languages is that the primary acoustic correlate of initial stress is f0 (pitch), based on a handful of studies of individual languages: Simard (2010) for Jaminjung; Bishop (2003) for Bininj Gun-wok; Fletcher & Evans (2002) for Dalabon and Bininj Gun-wok; among others. Not only is this unsubstantiated in the many Australian languages that lack rigorous acoustic analyses, but work on the likelihood of stress correlates to be the same across related languages is sparse to non-existent: there is no empirical basis for such a claim. One of the outcomes of this project is that it adds to existing work to more accurately construct a typology of stress correlates in Australian languages.

One motivation for this project is to use suprasegmental variation in our reconstructions of proto-Pama Nyungan. This language family is known for its surprising dearth of phonological variation among the modern languages, but recent work has found that sub-

stantial variation exists in phonotactics and other phonetic factors (Gasser & Bowern 2014, Bowern 2018b). By defining some principles of suprasegmental change, its influence on segments, and describing suprasegmental variation across Australia, perhaps sound change can become a more informative player in subgrouping and reconstruction of Pama Nyungan.

1.4.2 The Pama Nyungan family

The existence of a Pama Nyungan language family was first proposed in O'Grady, Voegelin & Voegelin (1966). While the proposed makeup of this family in terms of subgrouping and membership has changed over the years, most Australianists accept this language family as valid (Dixon (2001, 2002) does not). The general consensus is that Pama Nyungan is a large family of languages spoken in Australia, comprising roughly 75% of all languages spoken in Australia before European contact (Bowern 2011).

A landmark study in the reconstruction and subgrouping of Pama Nyungan is Bowern & Atkinson (2012) (revised in Bouckaert, Bowern & Atkinson (2018)). This study utilizes Bayesian phylogenetic methods to determine the statistical probability of genetic relationships between all languages proposed to be part of Pama Nyungan. The authors compiled data from 194 Australian languages, coding for 189 cognate sets across these, and calculated the most probable subgroups and split relations in the family. Their results found four major splits in the family which roughly correspond to geographical regions— Southeastern, Northern, Central, and Western. Similarly, 25 of the 28 subgroups that had been proposed in previous literature were reconstructed in the Bayesian tree, with the missing subgroups likely being confounded by missing data.

The phonological similarity of Australian languages is what has led some scholars to doubt the historical reconstructability of these languages. One way of identifying loanwords in a language is to determine whether observed historical sound changes have operated on

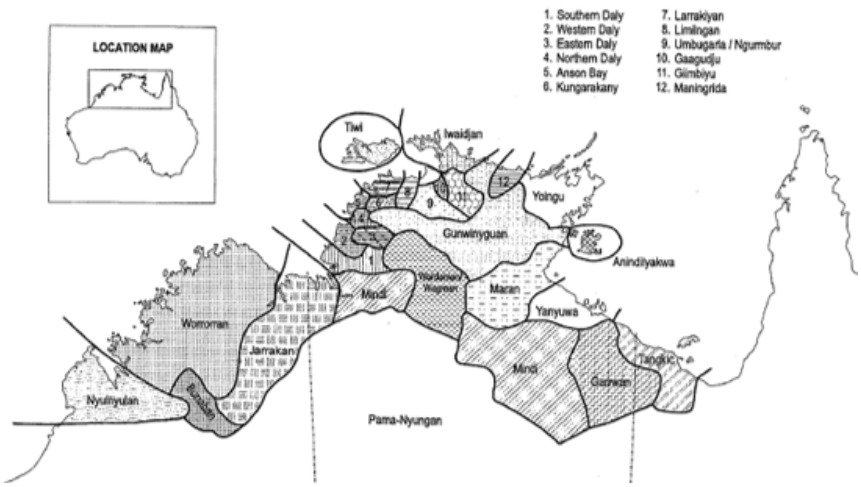

Figure 1.2: From Evans (2003b). Map of non-Pama Nyungan language subgroups.

them or not. When phonological variation, and therefore described sound changes, are scarce, it becomes more difficult to identify loans. However, as Bowern & Atkinson (2012) argue, the presence of loanwords in a phylogenetic data set does not affect results unless loaning is extremely frequent, which only applies to a handful of Australian languages.

1.4.3 Non-Pama Nyungan languages

While the majority of historical linguistics work and reconstruction has been done on the large Pama Nyungan family, there are many languages, concentrated in the north-central region of the Australian continent, which are non-Pama Nyungan. These are mapped out in Evans (2003b), reproduced as Figure 1.2 here. These languages are all Australian, i.e. proposed to have descended from proto-Australian, but they belong to branches of the Australian language family that were not as fruitful as Pama Nyungan. Twenty-eight language groups have been proposed for these languages, all of which are laid out in Figure 1.2. These families often consist of only a handful of languages, and in most cases the phylogenetic relations between these families in the greater Australian family structure is unclear. Some groups, such as Tangkic, show some evidence of belonging to the Pama Nyungan

family but this evidence is not entirely clear (cf. Bowern 2020). Some languages are more clearly distinct from this major family, but their relationship to other non-Pama Nyungan languages is tenuous at best; some of the groups in Fig. 1.2 may only be groups of geographically close languages that are not in fact linguistically related. Other languages are likely isolates, such as Tiwi, Gaagudju, and Giimbiyu (Evans 2003b).

1.4.4 Language endangerment

Fewer than half of the languages spoken in Australia before European contact continue to be spoken. Among these, only around 20 languages are still actively spoken and passed down to children, and the other hundred or so are primarily spoken by older people and no longer being passed down (Commonwealth of Australia 2005). This is a dire situation for indigenous Australian languages, and many communities have revitalization programs in place to help combat this; there are around fifty active language reclamation programs in Australia at time of writing (p.c. Claire Bowern). For the purposes of linguistic research, the fact of the matter is that most Australian languages today only survive in archival materials. This dissertation makes the case for using such materials, which were generally not collected with modern phonetic analysis in mind, as a fruitful source of acoustic data.

1.5 Archival phonetics

For fieldworkers, archiving collected data to preserve it for future use is a critical part of research. There are now decades of archived linguistic data from many, often under-researched, endangered, and sometimes extinct, languages of the world. And while these archival deposits often require some additional processing before they are ready for phonetic analysis (Babinski & Bowern 2021, Babinski 2021a) the data available here is invaluable, especially for comparative work and when a language is no longer spoken.

Linguists and other field researchers often spend many years, sometimes their entire

careers, building relationships with the communities in which they work. This relationship-building is crucial for ethical fieldwork, and allows for both the researcher(s) and the community to achieve their goals in a mutually beneficial manner. It therefore becomes untenable for one researcher to collect their own linguistic data for a large comparative study without perpetuating harmful practices such as 'helicopter research.' Only in rare scenarios would one researcher be able to personally gather data from over a dozen languages without perpetuating such harmful and unethical research practices.

Archival resources, then, become vitally important for conducting ethical typological and otherwise multi-lingual linguistic work. However, the field of phonetics is still largely based on the assumption that the researcher can and should collect their own data that is carefully controlled in an experimental setup. Less focus has been placed on how to glean meaningful insights from natural language data that was not collected for the purpose of one's specific research question. However, a body of recent work has made use of archival materials for phonetic study; these studies and other practical considerations of working with this type of data are discussed in Chapter 8.

1.6 Chapter summary

There are nine chapters in this dissertation. This is the end of the introductory chapter. Chapters 2 and 3 provide background information for the results presented in Chapters 4-7, Chapter 8 provides discussion of the practical use of archival materials in linguistics research, and Chapter 9 presents an overall discussion and conclusion to the dissertation.

Information on the data sources used in this project, including a review of the relevant background on each of the 16 languages in this study, can be found in Chapter 2. Background on data processing and analytical methods is outlined in Chapter 3. This chapter provides discussion of the process of force aligning the archival materials using the Montreal Forced Aligner (McAuliffe et al. 2017), extracting acoustic information from Praat

(Boersma & Weenink 2018), and analyzing measurements in R (R Core Development Team 2020).

There are four chapters presenting results of different types. Chapter 4 presents the results of an acoustic study of stress for each language; the focus is on cross-linguistic results and implications for theories of suprasegmental change. Detail on inter-speaker (intra-language) results are discussed in Chapter 5 for each language with more than one speaker's audio available. Chapter 6 synthesizes these two sets of results, presenting a quantitative analysis of the variation in stress cues both within and across languages using phylogenetic methods. Chapter 7 presents the results of an automated method to hypothesize major categories of phrasal contours in each language, using the methods first put forth by Kaland (2021).

Chapter 8 provides an in-depth discussion of using existing archival materials for linguistics research, specifically phonetics work. This chapter is meant to serve as a practical discussion of the methods used in this dissertation and elsewhere, to benefit those wishing to conduct similar work in the future. Finally, an overall discussion of the results and insights of this project can be found in Chapter 9, which also concludes the dissertation. [See the Table of Contents for information on appendices.]

Chapter 2

Data

As mentioned in Chapter 1, the type of audio data used for the analysis of stress and prosody in this dissertation was natural, spontaneous speech in the forms of narratives or conversations. These were identified within archival collections made by researchers in the past, resulting in a corpus consisting of spontaneous speech recordings that vary in size and in the date of recording. In this chapter, the sources of data used in this dissertation are described in detail.

Section 2.1 describes the methodology used in identifying and downloading audio data from online language archives. Sixteen languages are included in this project, and each of these is described in the remainder of this chapter, grouped by historical relatedness. Languages of the Pama-Nyungan family are discussed first, in §2.2.1, followed by non-Pama Nyungan langauges in §2.2.2. For each language, there is description of both the existing documentation of the language, with a focus on prosodic documentation of any sort, as well as a description of the particular archival collection that is used in the analysis here. An overview of each archival collection used in this project can be found in Appendix A.

2.1 Sources of Data

This study makes use of archived audio narratives from three archives: the Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS, https://aiatsis.gov.au), the Endangered Languages Archive (ELAR, https://elar.soas.ac.uk), and the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC, https://paradisec.org.au). Access to these archival deposits, in most cases, required a request to be sent to the archive, and usually approval by the depositors themselves. A small group of deposits on ELAR and PARADISEC were available for direct download without consultation with the archive or with the depositor(s). Beyond the data acquired through archival requests, a small set of language data in the present study was acquired personally from the researchers who had collected the recordings themselves. The sources of each language data set, along with depositors and method of acquisition, are given in Appendix A.

Archival deposits were screened for recordings of narratives, conversations, and other types of natural, running speech. Recordings of word or sentence elicitations were excluded from the present study, in order to avoid potential confounds such as list intonation and potentially unnatural prosody in ungrammatical or otherwise non-naturalistic utterances. Archival deposits with utterance-level transcriptions as *ELAN* (2018) files were preferred over fully unaligned transcripts.

Because of the nature of the present data set, the conditions of recording and recording devices used for each language in this study were unable to be controlled for. This fact distinguishes the present study from most other acoustic studies of stress, which tend to collect new, highly controlled experimental data to analyze (cf. Fry 1958, Gordon & Roettger 2017). This approach is infeasible for a study of this sort for a number of reasons. First, collecting substantial amounts of high-quality experimental data for each language in this

study would be time-consuming beyond the limits of a dissertation. The recordings that are used in the present study were made by researchers who spent many months or even years working on these languages, and have close relationships with the communities of speakers who participated in them. Fostering such relationships is extremely valuable and crucial to ethical data collection in linguistics. Second, conducting the types of phonetic experiments that are often used in studies of lexical stress is a much more difficult task when working in a remote community, often with elderly individuals and no access to an environment as truly silent as a phonetics booth or lab setting. And finally, using archival data allows the researcher to access languages that are no longer spoken, a fact that is true of a large proportion of the languages in the present study. Including these languages adds rich information about the variation that has existed in Australia, even if there are not current speakers to participate in a new experiment.

2.2 Languages

This section provides some relevant background information about the languages in the present study, and details about the archival deposit when possible. Information about the identity of speakers, recording conditions, equipment used, and other relevant information about each deposit were not always available directly from the archival entry, so some of this information may be missing. A map of all the languages in this sample is given in Figure 2.1.

2.2.1 Pama-Nyungan Languages

This dissertation project includes data from five languages in the Pama Nyungan family. Four of these are part of the major Western subgroup, while one (Yidiny) is part of the Northern subgroup. Two of the Western languages (Warlpiri and Wanyjirra) are members of the Ngumpin-Yapa subgroup within the Western designation. This section provides a

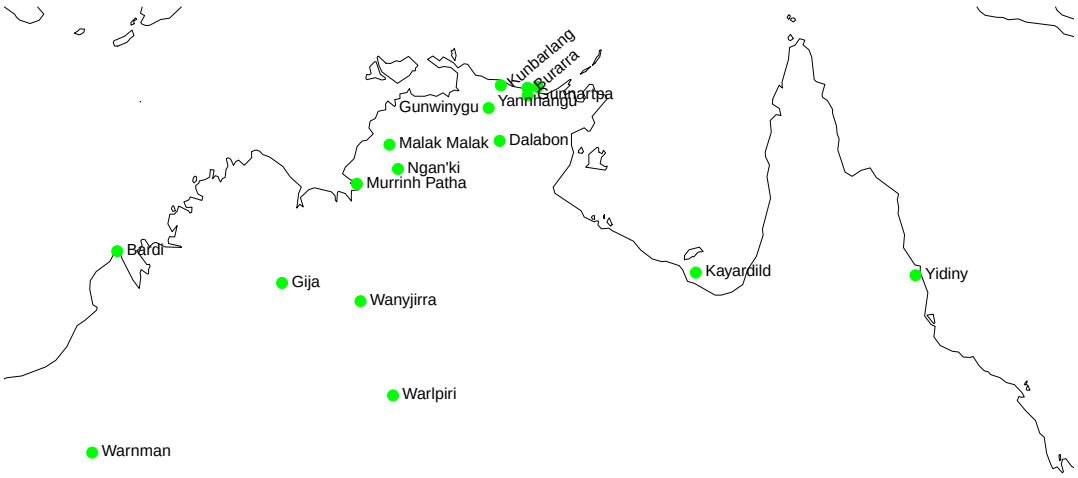

Figure 2.1: Map of all languages in sample.

brief discussion of each of these languages.

Northern: Yidiny

One language, Yidiny, represents the Northern subgroup of Pama-Nyungan in this project. Yidiny [yii] is a Pama-Nyungan language in the Paman subgroup, spoken in the Cape York Peninsula in Queensland, Australia. It is closely related to Djabugay [dyy], though there is some disagreement as to whether this grouping constitutes an independent subgroup of Pama-Nyungan (Dixon 2002) or a clade within the larger Paman subgroup (Bouckaert, Bowern & Atkinson 2018). It was first described in Dixon (1977a,b), who suggested that its system of stress assignment was variable according to a complex set of rules, including stress-to-weight. This idea was challenged by Bowern, Alpher & Round (2013) using the same audio data used by Dixon, concluding that Yidiny in fact has consistent initial stress.

Recordings used for this project were made and transcribed at the utterance level by R.M.W. Dixon sometime between 1963 and 1975 with contributions from two Yidiny speakers, Tilly Fuller and Dick Moses (Dixon 1977a: xv). At the time of writing the

grammar, Dixon (1977a) estimated there were around 2,000 Yidiny speakers in northern Queensland; according to a 2016 census, there are only 19 speakers of the language today (Eberhard, Simons & Fenning 2021).

The audio used in this dissertation project were digitized from the original tape recordings and are deposited with AIATSIS. Materials were acquired by Claire Bowern and have been used in a previous study of Yidiny stress placement (Bowern, Alpher & Round 2013) and a study comparing the performance of various forced alignment programs on low-resource language data (Babinski et al. 2019). There are around 45 minutes of audio consisting of around 25,000 word tokens combined from speakers Tilly Fuller (F) and Dick Moses (M). This audio was force aligned with the Montreal Forced Aligner (McAuliffe et al. 2017) and manually corrected for the (Babinski et al. 2019) project; these files are the ones used in the analysis of Yidiny in this dissertation.

Western: Wanyjirra, Warlpiri, Warnman, Yan-nhangu

The western grouping of Pama Nyungan languages consists largely of those languages spoken in the large state of Western Australia. The outgroups of this branch are the Yolŋgu and Ngarna subgroups, with the main group of Western languages being part of the Nyungic branch. Four of the languages described below are members of this Nyungic branch, specifically Ngumpin-Yapa and Wati subgroups, and Yan-nhangu is a Yolŋu language.

Wanyjirra

Wanyjirra is a Ngumpin-Yapa language and a member of the Ngumpin dialect continuum (Senge 2016: 6). This language has been classified by many as a dialect of Gurindji [gue], including in the Ethnologue (Eberhard, Simons & Fenning 2021). However, Senge (2016) presents arguments that Wanyjirra is a language in its own right that is very closely related to Gurindji, Jaru, Malngin and other languages in the Ngumpin dialect chain. It was once

spoken in the Victoria River District of Northern Territory in Australia, but no Aboriginal people live in this region today (Senge 2016: 5).

The language has been described in Senge (2016) based in part on the field recordings in the ELAR deposit sourced for this project. Stress is described briefly in the grammar as being found consistently on the first syllable of the word and marked by "increased loudness, and sometimes, length although it is not phonologically contrastive" (Senge 2016: 101).

Data for this language come from Chikako Senge, who collected the materials in 2011 and deposited the materials on ELAR. Access to the deposit was requested by me and granted via ELAR in 2019. The recordings were made at the Kimberley Language Resource Centre in Halls Creek, Western Australia. All of the audio used for analysis in this project come from one speaker, Tiny McCale (F), who was around 83 years old at the time of recording. All recordings were made using a Zoom H4n recorder, sometimes with an external condenser microphone, and utterance-level transcriptions were made in ELAN by Senge.

Warlpiri

Warlpiri was first described comprehensively in Nash (1980), which includes substantial remarks on stress in the language. The basic description given is that "primary stress is uniformly on the initial syllable of the word" (Nash 1980: 100). Further discussion of the stress patterns of Warlpiri concerns the placement of secondary stress, which is not included in the analyses in this dissertation and so will not be discussed here. The impressionistic description of the correlates of stress made by Nash are: increased intensity, increased duration, and higher pitch.

This is the only language where the data used did not already have time-aligned utterance alignments. Instead, I had the audio and a text transcription, for which I aligned utterances based on my judgment and the presence of pauses in speech. Recordings were

made by David Nash and are not part of an archival deposit, instead being shared personally with Claire Bowern and given to me for this project.

Warnman

Warnman [wbt] is a language in the Wati subgroup and originates near Jigalong in Western Australia. According to a 2016 census, there may be no remaining Warnman speakers (Eberhard, Simons & Fenning 2021).

The data for this project come from a PARADISEC deposit made by Nick Thieberger, of recordings made in 1984. Among other documentary audio materials, this deposit includes traditional stories that can be used for this sort of project. Data come from about 30 minutes of audio from only one speaker.

Yan-nhangu

Yan-nhangu is a Yolŋgu language spoken in Arnhem Land in north central Australia. This is a largely multilingual region, with Yan-nhangu specifically having notable contact with Burarra, a non-Pama Nyungan language also included in this dissertation project (Bowern & James 2005). While documentation materials were being collected around 2007, there were only six remaining Yan-nhangu speakers. Documentation of this language is sparse, with the notable exception of a learner's guide (Bowern 2006) published as part of a general documentation project that also included the recording of the materials used here. Audio was recorded by Bowern between 2004 and 2006 and deposited at ELAR (Bowern 2007).

2.2.2 Non-Pama-Nyungan Languages

The remaining eleven languages in this survey have not been found to be a member of the Pama Nyungan family. These are described here, grouped together into their smaller families when relevant.

Nyulnyulan: Bardi

Bardi [bcj] is a language of the Nyulnyulan family in the Kimberley region. The audio data used in this project was recorded on tape between 1990 and 2008 and then digitized by Claire Bowern. There are five fluent Bardi speakers who contributed to the recordings: Bessie Ejai (F), Tudor Ejai (M), Nancy Isaac (F), Jessie Sampi (F), and David Wiggan (M). All were over 70 years of age at time of recording.

Primary stress in Bardi has been described in both Bowern (2012) and Bowern, McDonough & Kelliher (2012) as demarcative, occurring consistently on the first syllable of the word. Every word has this initial stress pattern, including pronouns, adverbs, preverbs, and loanwords (Bowern 2012: 112). The acoustic correlates of stress in Bardi have been observed to be increased vowel duration, higher intensity, and potentially vowel peripheralization and pitch, though this latter correlate may occur as a function of higher-level prosody rather than word-level stress *per se*.

Bardi has a complex system for assigning secondary stress, although this is not at issue in this dissertation. Secondary stress assignment is sensitive to morphological structure, prosodic vowel deletion, and syllable weight, all of which add surface complexity to what seems to be basically a rightward aligning trochaic system of secondary stress (Bowern 2012: 113).

A preliminary analysis of lexical stress in Bardi has been presented in Bowern, McDonough & Kelliher (2012) and Bowern (2012). The correlates of stress in Bardi were identified as increased vowel duration and intensity. Increases in pitch are hypothesized to be correlates of phrase-level prosody instead of word-level stress, and stressed vowels are found to be slightly more peripheral than unstressed ones (Bowern, McDonough & Kelliher 2012: 344).

Maningrida: Burarra/Gunnartpa

Burarra [bvr] is generally considered part of the small Maningrida family, spoken in Arnhem Land in northern Australia (Evans 2003b). It has some grammatical description (R. Green 1987) and is proposed to have consistent initial stress (Glasgow 1981). Gunnartpa is one of the three dialects of Burarra, spoken in the area of the Cadell River. Two Burarra sources were available for the present study, both of which focused on speakers of the Gunnartpa dialect. One of these data sources is a collection of digitized tape recordings in the PARADISEC archive (Carew 1993). This audio was collected between 1993-1996 in Arnhem Land by Margaret Carew. Recordings were originally made on casette tape and digitized to .wav files in 2010. The other source is from ELAR, which also contains narrative recordings collected by Margaret Carew, all of which are transcribed and aligned at the utterance level. These materials were collected between 1995 and 2012.

The Burarra sources contain audio from 13 speakers: Rose Darcy, Betty Warnduk, Margie West, Jeannie Brown, Laurie Miyaga, Crusoe Baterra England, Margaret Nulla, Katy Fry, England Bangala, Harry Gamarrang Litchfield, Michael Borrorrbuma, Rosie Jinmujinggul, and Terry Gela Ngamandara. Speakers' ages were not listed in the metadata for the archival deposit.

The acoustic correlates of stress in Burarra/Gunnartpa have been analyzed for the first time in this dissertation and will be submitted as part of a more general phonetic sketch to *Illustrations of the IPA* (Babinski et al. in prep.).

Jarrakan: Gija

Gija [gia] is a member of the Jarrakan language group along with two other languages, Miriwoong [mep] and Gajirrabeng [gdh]. It is spoken in the eastern Kimberley region but is no longer learned by children. All current Gija speakers are over eighty years of age

in a community of about 800 (Kofod 2013, de Dear, Possemato & Blythe 2020). Data for this project comes from an ELAR deposit made by Frances Kofod in 2013. The deposit is meant to document cultural and historical Gija knowledge, as well as the language used to talk about forms of artistic expression such as painting and dance. The recordings used for this project feature five speakers: Mabel Jawalji, Madigan Thomas, Peggy Patrick, Rusty Peters, and Paddy Springvale. Most recordings were made between 2008-2011, with a smaller portion of the materials collected in 1988.

A short description of Gija stress was presented in Taylor & Taylor (1971: 108), who note that stress in "normal intonation" falls on the initial syllable of the word.

Gunwinyguan: Dalabon, Gunwinggu, Kunbarlang

The Gunwinyguan family is a small group of related languages spoken in Arnhem Land in north central Australia. In addition to the languages included here, the family consists of Jawoyn, Mangarrayi, Ngalakgan, Ngandi, Nunggubuyu, and Warray. Reconstruction of Proto-Gunwinyguan has been presented in a few published works including Harvey et al. (2003), Alpher, Evans, Harvey, et al. (2003), and Evans (2003a). The three languages from the Gunwinyguan family used in this dissertation are all part of the *marne* group, in the center of the tree in Figure 2.2.

Dalabon

Dalabon is a Gunwinyguan language spoken in southwestern Arnhem Land in Northern Territory, Australia. At the time the Dalabon collection was deposited to ELAR (2011), it had fewer than ten speakers. It is classified as a Central Gunwinyguan language along with varieties of Bininj Gun-wok (cf. Fig. 2.2). The prosody of Dalabon has been investigated in Ross (2011) and Ross, Fletcher & Nordlinger (2016).

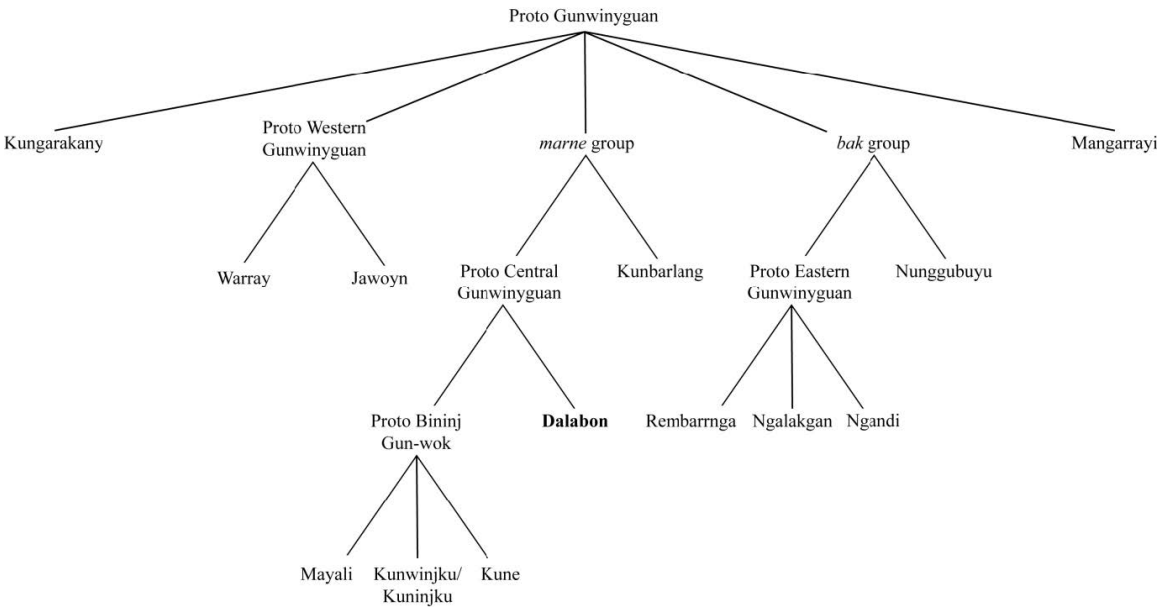

Figure 2.2: Proto-Gunwinyguan language family (Evans 2003a, Ross 2011).

Data for Dalabon were sourced from Maia Ponsonnet's ELAR deposit, which contains recordings from 2010-2012. The narrative recordings selected for this project contain speech from five speakers: June Jolly Ashley (F), Lily Bennett (F), Maggie Tukumba (F), Nikibini Darluk (F), and Philip Ashley (M). Speaker ages ranged from around 40 to around 70 years old at time of recording.

Gunwinggu

Gunwinggu [gup], a variety of Bininj Gun-Wok, and is spoken in Arnhem Land and is described in Evans (2003a). It is more closely related to Dalabon as another Central Gunwinyguan language, than to Kunbarlang (cf. Fig. 2.2).

Data come from a PARADISEC audio collection (Si 2014). Recordings were taken as a linguistic fieldwork project (some elicitation, some narrative) by Aung Si near Maningrida in the Northern Territory. Three speakers are represented in the data used for this dissertation: Jill Yirrindili, Charlie Brian, and another speaker whose name was not recoverable in

the deposit, referred to by the initial "A" in transcription documents.

Kunbarlang

Kunbarlang is the out-group language in the *marne* group of the Gunwinyguan family (cf. Fig.2.2) and is the only coastal Gunwinyguan language (see the map in Fig. 2.1). O'Keeffe et al. (2017) mentions that Kunbarlang's place in the family is uncertain, however, likely due to contact influences with Iwaidjan and Maningridan languages that are spoken in the area.

Materials for Kunbarlang come from the ELAR deposit O'Keeffe et al. (2017). At the time of deposit, there were fewer than 50 Kunbarlang speakers. Two speakers are represented in the data used for this project, Frank Ambidjambidj and Sandra Makurlngu.

Tangkic: Kayardild

Kayardild [gyd] is a Tangkic language spoken on Bentinck and Mornington Islands, part of the South Wellesley Islands in Queensland, Australia. It is listed here as a non-Pama Nyungan language, although recent work has considered whether Tangkic should be considered a subgroup of Pama Nyungan (Bowern 2020).

At the time of data collection in 2004-2005, depositor Erich Round documented the speech of the eight remaining fluent speakers of Kayardild, all of whom were over the age of 60 (Round 2015). The narrative audio data used for this project includes speech from five of these speakers, all of whom are female. They are Amy Loogatha, May Moodoonuthi, Dawn Naranatjil, Paula Paul, and Ethel Thomas.

Northern Daly: Malak Malak

MalakMalak [mpb] has been classified by Nordlinger (2017) as a Northern Daly language, a small grouping that consists of MalakMalak and one other language, Kuwema [woa].

As other languages in the Daly subgroups, it is spoken in the Daly river region of Arnhem Land. Data for this language come from an ELAR deposit made by Dorothea Hoffmann. At the time of deposit, 2012-2013, there were eleven fluent native speakers of MalakMalak remaining. All of these speakers primarily communicate with others in either Kriol or English, and almost all of them also know Matngala [zml] (Hoffmann 2015). The two speakers represented in the Malak Malak data used here are Thommy Wurra and Karrangan.

Southern Daly Languages: Murrinhpatha, Ngan'gi

Murrinhpatha [mwf] and Ngan'gityemerri [nam], of which Ngan'gi is a dialect, were once considered language isolates; however, it has been proposed by I. Green (1995) that these two languages are related to one another in the Southern Daly subgroup. The evidence Green presents in favor of this subgrouping are largely based on similarities in verbal morphology, although the two languages do not have a high degree of lexical similarity.

Murrinhpatha [mwf] is spoken in the Daly River region in the Northern Territory of Australia. It was described by Walsh (1976) as a PhD thesis, and in a learner's grammar (Street 1987); there is also a Murrinhpatha-English dictionary (Street 2012). The language has since been the subject of a number of studies, including a comprehensive morphophonological sketch (Mansfield 2019) from which my assumptions about Murrinhpatha prominence come. Audio recordings were made by John Mansfield and come from a PARADISEC collection archived by Danielle Barth (Barth 2009). Recordings were made as part of a structured social cognition task, where speakers were asked to create or retell a narrative. Only one speaker, Gerald Mardinga, is included in this project.

The presence and placement of stress in Murrinhpatha has been a point of contention in the literature (cf. especially Mansfield (2019)). Most accounts have been based on the assumption that Murrinhpatha is typical among Australian languages in this respect, but the phonetic facts of the language clearly deviate from this assumption, thus confounding these

explanations.

Ngan'ki [nam] is also spoken in the Daly River region in the Northern Territory of Australia. It is one of two major dialects of the language Ngan'gityemerri, which is the proposed sister to MurrinhPatha in the Southern Daly subgroup. A grammar of this language was written by Reid (1990). Audio was extracted from video recordings available from PARADISEC (Reid 2013), collected by Nicholas Reid in the community Nauiyu in the Northern Territory. There are nine Ngan'gi speakers who are represented in this collection, but only first names were recoverable in deposit metadata. The speakers' names are Benigna, Christina, Kinbi, Louise, Anne, Molly, Monica, John, and Agnes.

2.3 Summary: Relationships between languages

There are both historically related and areally related languages in the data sets described in this chapter. They have been introduced here grouped by historical affiliation; these groups, especially the Pama Nyungan languages, should be noted for potential signs of historical stability and change. Some groups, especially the Daly langauges, are more regional groupings than definitive phylogenetic groups, and should be noted for potential areal patterns. Finally, the languages of Arnhem Land— Yan-nhangu, Burarra/Gunnartpa, Kunbarlang, Gunwinggu, Malak Malak, Dalabon, Ngan'gi, and Murrinh Patha— are spoken in a dense linguistic area in the Northern Territory, and thus are in contact with languages from many different historically related groups. These languages should be noted for potential effects of language contact on the phonetics of prosodic phenomena going forward.

Chapter 3

Methodology

Before any sort of linguistic analysis, one's data set must be cleaned to check for errors; this is especially true for projects like this one, which use spontaneous speech recordings made in variable field conditions. Beginning with the archival data described in Chapter 2, audio files were associated with utterance-level transcripts from the depositors themselves. However, in order to conduct this study of stress, word- and segment-level alignments were also needed. These were achieved using forced alignment, as discussed in §3.1.

After alignment, further processing was required to extract acoustic measurements, normalize those measurements, and conduct statistical modeling to analyze the cues to stress in all of these languages; all of these methods are described in §3.2. These methods are especially relevant for the results presented in Chapters 4 and 5. Methods for the analyses in Chapters 6 and 7 are discussed separately within these chapters. Finally, §3.3 presents a brief discussion of the quality of the data being processed and analyzed in this dissertation.

3.1 Forced Alignment

Forced alignment algorithms offer an automatic way to align segment labels to spoken language audio. The Montreal Forced Aligner (McAuliffe et al. 2017) was used in this dis-

sertation project to add segment-level alignments given manually created utterance level transcripts. The automatic alignments allowed for much more data to be processed than if it were to be done manually, and required relatively minimal manual correction post-alignment. This section introduces the technical details of forced alignment, and the specific process used to prepare and align the languages in this sample.

3.1.1 What is forced alignment?

There are two major Automatic Speech Recognition (ASR) toolkits on which a handful of forced alignment programs are based. One is the HTK algorithm, or Hidden Markov Model (HMM) Toolkit, on which many popular aligners are based, such as the University of Pennsylvania Phonetics Lab Aligner (P2FA) (Evanini, Isard & Liberman 2009), the Prosodylab Aligner (Gorman, Howell & Wagner 2011), FAVE align (Rosenfelder et al. 2011), and Web-MAUS (Kisler, Schiel & Sloetjes 2012), among others. HTK uses a Hidden Markov Model to identify speech segments. HMMs are probabilistic models that, in the context of ASR, can identify the boundaries of a speech segment based on its acoustic features. The model is first trained on some amount of segmented audio, from which the HMM can construct probabilistic acoustic models of the features of each phone included in the data. The HMM can then be used to predict where these boundaries are in new audio data.

The other major ASR toolkit used for forced alignment is the Kaldi Automatic Speech Recognition toolkit, which the Montreal Forced Aligner (MFA) (McAuliffe et al. 2017), DARLA (Reddy & Stanford 2015), Gentle (Ochshorn & Hawkins 2017), and others are based on. The Kaldi ASR Toolkit is based on a Hidden Markov Model-Gaussian Mixture Model (HMM-GMM). A GMM is a way of identifying whether a particular data point is part of some category, where each category is defined as a Gaussian distribution. Speech segments in this sort of model are represented as a Gaussian distribution of relevant acoustic features.

Regardless of the alignment algorithm used, forced alignment works in the same basic way. The algorithm must first be trained on some amount of manually segmented material. This creates a language-specific model of segments that allows the algorithm to match transcribed segments to acoustic features. Words need to be transcribed into sequences of phones in some standardized transcription alphabet; here, the ARPABET phones are used, as described in more detail in §3.1.4. The training model consists of acoustic models for each phone in the data, which are then used as standards for identifying the boundaries of those phones in the new data being aligned. Ideally, the training model would consist of segmented material from the language being studied; however, as discussed in §3.1.2, the realities of endangered language research can make this infeasible and can necessitate the use of a model trained on a different language with appropriate manual checking of outputs.

Forced alignment algorithms may be constrained or unconstrained in nature. Constrained forced alignment requires that transcripts be aligned at the level of the utterance or breath group for the aligner to work, while unconstrained forced alignment does not require this. While unconstrained forced alignment is saves time in pre-processing, the risks of compounding errors in alignment tilt the scales in favor of constrained forced alignment. Compounding errors can occur when a mistake is made early on in a sound file, which results in a ripple effect for all following alignments in that file. With existing utterance-level alignments, a mistake in alignment can only have compounding effects on the remaining segments in that utterance, instead of most or all of a long sound recording. This limits the ripple effects of such errors and more easily deals with anomalies such as long pauses, stretches of speech in another language, or non-speech background noise that are especially common when working with natural language recordings made in the field.

3.1.2 Forced alignment for under-resourced languages

Many forced alignment algorithms require large amounts of data to create a language-specific training model. When conducting research on endangered or under-researched languages, this amount of aligned audio is often non-existent and would require a massive time investment on the part of the researcher(s) in order to do by hand. For this reason, a growing body of work has investigated the usability and accuracy of forced aligners in endangered and under-researched language work. Using a forced alignment algorithm with a model trained on a language that differs from the target language results in higher rates of error in the resulting alignment, but with manual correction it can offer a way of greatly reducing segmentation time and facilitating not just linguistic research but also language reclamation projects (cf. Coto-Solano et al. (2022)).

DiCanio et al. (2015) compares the performance of the P2FA and HM-Align alignment models on Yoloxchitl Mixtec language data. The data consisted of elicited word lists that were constructed to collect words of varying lengths. They found that HM-Align made fewer errors in alignment than P2FA, and that certain types of segments had higher error rates than others. The authors attribute the differences in performance primarily to the fact that HM-Align uses an allophonic English phone set, allowing for a greater phonetic specificity than P2FA, which uses a context-free phonemic English set of phones. When using English-trained models on non-English language data, the availability of a wider set of phones for transcription can only increase the accuracy of forced alignment results.

Another study investigating the feasibility of forced alignment for aligning endangered language recordings is Johnson, Di Paolo & Bell (2018), who look at the performance of the Prosodylab Aligner on Tongan field data. The recordings used in this test were word lists in Tongan, although the authors note that they planned to run forced alignment on connected speech in the future. Recordings were made in a field setting with all the requisite back-

ground noise that that environment entails. The authors considered both the raw recorded audio ('dirty' files) as well as audio that had been cleaned of ambient noise ('clean' files). Results were fairly accurate for both types of files, as long as the aligner was trained only on cleaned data. Furthermore, when compared to two different humans' manual alignments of a subset of the data, the Prosodylab Aligner's results did not differ from manual alignment any more than one human's alignment might differ from another. The authors conclude that using forced alignment in this way, even with manual corrections post-alignment, is a viable time-saving option for those looking to align their field recordings.

Babinski et al. (2019) considered the performance of forced alignment specifically for Australian languages, using around 45 minutes of running Yidiny speech as the test data. The data were collected in a field setting and transcribed at the utterance level. The aligners compared in this study were P2FA, DARLA, and MFA. Of particular interest here were the various potential approximations in transcription using the ARPABET transcription system. Because ARPABET was created to transcribe English phones, some Yidiny sounds were not available and needed to be approximated in some way. Specifically, alternatives were considered for the transcription of stops, which need to be specified for voicing in ARPABET despite having no voicing contrast in Yidiny; the palatal nasal, which was transcribed alternately as N, Y, and N+Y; and the trill, tap, and retroflex rhotic, which are separate phonemes in Yidiny but have very different distributions in English. The optimal transcriptions for these sets of phones were: stops transcribed as voiceless P T K; the palatal nasal transcribed as N; and the rhotics as R (trill and retroflex, as they are allophonic), D (tap). These are the standards used in transcribing language data in this dissertation; more details on the ARPABET transcription are presented in §3.1.4.

The P2FA and MFA aligners performed similarly to one another in terms of prosodic alignment, vowel measurements, and consonant durations, and were fairly accurate to the gold standard manual alignments. The unconstrained version of DARLA was used, and this

model performed significantly worse than the other aligners tested. Both of these models can save the researcher time in segmentation of audio even with manual correction. This dissertation uses the Montreal Forced Aligner, which is described in the following section.

3.1.3 The Montreal Forced Aligner

The Montreal Forced Aligner (MFA) is an open source forced aligner first released with McAuliffe et al. (2017) as an update to the Prosodylab aligner. In contrast to the Prosodylab aligner, however, MFA is based on the Kaldi ASR toolkit. For alignment, MFA uses a Hidden Markov Model-Gaussian Mixture Model (HMM-GMM) and passes over the data three times to create the phones in the trained language model. First, each phone is modelled independent of phonological context (the monophone models stage). Then, phones are modelled in their various contexts, considering one phone on either side of the target; this stage creates the triphone models. And finally, speaker differences are taken into account to create speaker-adapted triphone models. All monophones in the input to the aligner must be present in the training model, but this does not hold for all triphones.

MFA uses constrained alignment, requiring utterance-level transcriptions as input. It takes as input the following files: WAV format sound files sampled at least at 16kHz; Praat TextGrids (*.TextGrid), or Prosodylab (*.lab) files, with utterance-level transcription; and a dictionary file, which lists each word in the transcript and associates it with a string of phones. It is optional to specify a language-specific training model, either from the existing pretrained models (of which there are 26 at the time of writing), or from one's own training model created from manually segmented audio. While the inclusion of the training model is optional, including it greatly increases the accuracy of the resulting alignments, even if the training model was trained on a language that does not match the input language. This is why, for the data in this dissertation, the pretrained English model was used in the alignment of the Australian language data.

Phone transcriptions for non-English MFA training models use the GlobalPhone (GP) corpora and phone sets for model training. GP standards are established in a language specific way for each language, and in some cases come with a language-specific grapheme-to-phoneme transliteration algorithm. The English-trained model, however, uses the ARPABET transcription system, a phone set that was created specifically to transliterate English phones. As none of the pretrained MFA models are specific to any Australian languages, it is reasonable to suspect that using any of these on Australian language data comes with some amount of error. The decision to use the English pretrained model instead of another option was made for convenience and in light of the knowledge that any MFA output would be checked manually for errors in alignment.

3.1.4 Data preparation

Audio files

The Montreal Forced Aligner requires that audio files be in WAV format and have a sampling rate of at least 16kHz. Archival deposits were identified as having WAV format audio files before requesting access, and none of them had a sampling rate below this threshold. In order to run MFA on a group of audio files, they must all have the same sampling rate. In some cases, the audio files as obtained from the archive had different sampling rates from one another, e.g. 16kHz and 44.1kHz. In these cases, those files with the higher sampling rate were downsampled to match the lower rate.

Utterance-level transcriptions

Utterance-level transcriptions were created by the archival depositors for each language in the sample. In most cases, these transcriptions came in the form of ELAN files. These files varied in their structure and in the amount of detail included. Only the highest level of

transcription was extracted from ELAN to the TextGrids that were used as input to MFA, i.e. morpheme-level transcriptions and syntactic information were not included. Usually, audio including speech from multiple speakers had the transcriptions for each speaker in a separate tier, and these tiers were preserved in the TextGrids and resulting alignments. Some language deposits included information about the speakers such as name, age, and other demographic and personal history information, but some only included the initials of the speakers that served as the titles of the ELAN tiers.

Usually, ELAN transcriptions were made in the orthography for the language that is considered standard either by the language communities or by researchers. Transcriptions into ARPABET were modulated to account for any cross-linguistic variation in orthographic conventions. These transcriptions were extracted from ELAN in two forms: as Praat TextGrids, to be used as direct input to MFA, and as a list of words, which was used to construct the dictionary file for MFA input.

Some language materials did not come as ELAN transcription. The Bardi transcriptions were done manually by Claire Bowern before ELAN software was widely used. The Warlpiri transcriptions were present in a plain text (.txt) file instead of an ELAN (.eaf) file. In this case, the author of this dissertation created the utterance-level boundaries manually and pasted the transcript at the appropriate place in the TextGrid. Yidiny utterance-level transcriptions were aligned using P2FA from text transcripts made by R.M.W. Dixon, and manually corrected.

Dictionary files

A .txt dictionary file is required input to MFA. This file was created by extracting ELAN transcriptions as a list of words in the ELAN export menu. The lists of words for each file in a language were combined in Notepad++, where the words were alphabetized and duplicates removed. Each line in the dictionary file was structured as the example in (3.1), with the

word as it appears in the transcript followed by the transcription, with each ARPABET character separated with a space. Because the orthographies of the languages in the sample are fairly transparent, the transcription into ARPABET characters was easily achieved using the Find & Replace function in the text editor.

(3.1) slipped S L IH1 P T

ARPABET Transcriptions

The ARPABET was created by the Advanced Research Projects Agency (ARPA) in the 1970s to transcribe English phonemes. Now known as the Defense Advanced Research Projects Agency, DARPA is an agency within the US Department of Defense dedicated to research and innovation that has existed since 1958. Among many other things, ARPA created ARPANET, the earliest predecessor to the Internet, and had a strong focus on computer technologies and language processing systems during the Cold War. The ARPABET is capable of encoding the phonemes and allophones of English, as well as stress marking and morpheme, word, and utterance boundaries.

Because ARPABET characters were created to transcribe English phones only, some approximations were necessary to fit the available characters to the phonemes in Australian languages. Most of these decisions were necessary for the Yidiny forced alignment in Babinski et al. (2019), where multiple transcription options were considered in some cases and evaluated for accuracy in automatic alignment. This section provides descriptions of the transcription conventions settled upon, partly as a result of these comparisons.

An inclusive phoneme inventories of the Australian languages in the present sample are given in Tables 3.1 and 3.2. The major ARPABET approximations needed to accommodate this sort of phoneme inventory include: vowel qualities, phonemic vowel length distinctions, (lack of) stop voicing, retroflex and palatal stops, retroflex and palatal laterals, and

the alveolar trill. All transcription conventions, including the more transparent decisions, are described here. An overview of the transcription conventions, with IPA, orthographic, and ARPABET equivalents, is given in Appendix B.

	front	central	back
close	i iː		u uː
	e eː		o oː
open		a aː	

Table 3.1: Inclusive vowel inventory of sample languages.

ARPABET: Vowels

Many Australian languages have a 3-vowel inventory described phonemically as /i a u/. A smaller number of languages have a five-vowel system /i e a o u/, and many languages additionally have phonemic vowel length distinctions.

In contrast to these relatively sparse inventories, ARPABET offers characters for 19 distinct vowel qualities, as well as primary, secondary, or tertiary stress marking. Stress in English is marked by increased duration and intensity, as well as vowel decentralization (Fry 1958). In order to avoid any errors caused by differences in the English-trained model of vowel stress, every vowel in the Australian dictionary files was marked as having primary stress, indicated by a numeral "1" after the vowel phone.

Phonemic vowel length distinctions are not present in most varieties of English and are not represented in ARPABET characters, nor are monophthongal /e a o/ vowels. In the case of short and long /i u/ phonemes, the length distinction was approximated as a tense-lax distinction. Long vowels were transcribed as tense /i/ and /u/, ARPABET "IY1" and "UW1," and short vowels were transcribed as lax /ɪ/ "IH1" and /ʊ/ "UH1." Long mid vowels were transcribed as diphthongal /eɪ/ "EY1" and /oʊ/ "OW1," while mid short vowels were written as lax mid vowels /ɛ/ "EH1" and /ɔ/ "AO1." Finally, the long and short low

central vowels were transcribed as /ɑ/ “AA1” and /ʌ/ “AH1,” respectively.

	bilabial	dental	alveolar	retroflex	palatal	velar
plosive	p	t̪	t	ʈ	c	k
nasal	m	n̪	n	ɳ	ɲ	ŋ
trill/tap			r, ɾ			
lateral			l	ɭ	ʎ	
approximant	w			ɻ	j	

Table 3.2: Inclusive consonant inventory of sample languages.

ARPABET: Stop voicing

None of the languages in the present sample have a phonemic stop voicing distinction, in keeping with the vast majority of Australian languages. While allophonic variation in this respect abounds (cf. Kakadelis (2018)), for the purposes of forced alignment transcription in ARPABET either the voiced or voiceless English stops must be decided upon. Babinski et al. (2019) found that using voiceless ARPABET stops in transcription yielded slightly more accurate segmentation results than using voiced ARPABET stops. For this reason, all stops in the sample languages were transcribed as voiceless.

ARPABET: Bilabial consonants

The bilabial consonants in the sample languages all have equivalents in English and are transcribed as such, “P” “M” “W.”

ARPABET: Dental consonants

Some Australian languages have dental stops /t̪/ and /n̪/. These are transcribed in ARPABET as “T” and “N” respectively. These stops were important to identify in the process of creating the pronunciation dictionary because they are usually written in orthography as ‘th’ and ‘nh’ but have no relation to the glottal fricative.

ARPABET: Alveolar consonants

The alveolar consonants /t/, /n/, and /l/ are all present in English and were straightforwardly transcribed in ARPABET characters. Many Australian languages additionally have an alveolar trill /r/, which is not present in English. This phone was transcribed as "R" in ARPABET, which corresponds to the retroflex approximant but indicates the rhotic quality of this phone and encodes the allophonic alternation that sometimes exists between /r/ and /ɻ/ in these languages. The trill also often exists in contrast to a tap /ɾ/, which in English is an allophonic realization of alveolar oral stops. As the voiced stops are not used to transcribe Australian stops here, the tap was transcribed as "D" in ARPABET. These were usually written orthographically as 'rr' /r/ and 'r' /ɾ/ and were distinguished in transcription in this way.

ARPABET: Retroflex consonants

Most Australian languages have retroflex consonants, as shown in Table 3.2. As English has the retroflex approximant /ɻ/, this phone is used to color the aligner's expectations of a stop or lateral. This is achieved by collocating ARPABET characters "R" and the relevant alveolar stop or lateral. This results in transcription of the retroflex stop /ʈ/ as "R T" in ARPABET. In cases such as these, measures of segment duration are calculated by recombining these two phones in post-processing.

ARPABET: Palatal consonants

Many Australian languages have a set of palatal consonants. The glide /j/ is also present in English and is transcribed in ARPABET as "Y". The palatal stop /c/ often exhibits some amount of frication on release and is transcribed as ARPABET "CH," the English /tʃ/. The palatal nasal is present in Yidiny and was tested with three different transcription conven-

tions in Babinski et al. (2019): as plain "N," plain "Y," and combined "N Y." It was found that plain "N" resulted in sufficiently accurate alignments and had the advantage over "N Y" of being one character to notate one phone. The aligner did not recognize /ɲ/ well when transcribed as "Y" because of the lack of nasality in this notation. Thus, the "N" convention is used for the palatal nasal. Additionally, while Yidiny does not have a palatal lateral, the same convention was extended to those languages with /ʎ/ by transcribing these as "L."

ARPABET: Velar consonants

Both velar stops are present in English and are available characters in ARPABET, "K" and "NG."

3.1.5 Post-alignment

The TextGrids resulting from MFA alignment were checked manually by either the author or one of the following Yale undergraduate research assistants: Jeremiah Jewell, Shayley Martin, and Ronnie Rodriguez. These students all had some coursework in phonetics and were familiar with Praat and spectrogram reading before working on this project. In the case of Yidiny alignments, the TextGrids were checked by one of the authors of Babinski et al. (2019).

Conceding that even human-made alignments are likely to vary across individuals, RAs were asked to look specifically for errors that looked particularly unlike a decision a human would make. For example, automatic alignments sometimes produce very long initial stops because the beginning of the closure cannot be located, while a human would have an idea of a reasonable boundary to place here. Another common issue comes about in vowel-glide sequences where the boundary is difficult to determine. An automatic system will often make the first segment in this sequence quite long, as it does not find a reason to place an end boundary, leaving the second segment to be very short. In contrast, a human aligner

would be able to better approximate these boundaries using audio cues and common sense.

MFA is particularly good at locating clear consonant-vowel transitions, and the RAs were able to double-check these to ensure that there were not systematic mistakes in the alignments. As this dissertation places a particular focus on vowel acoustics, RAs were asked to look closely at vowel segmentation and ensure that the TextGrid boundaries accurately captured vocalic periodicity in the waveform. The corrected files were reviewed by the author, who also addressed any issues that came up in the process of TextGrid review.

3.2 Statistical Methods

This section outlines the way that acoustic measurements are taken and how they are extracted (§3.2.1) and the statistical methods by which stress correlates are identified (§3.2.2). Five acoustic measurements with potential correlation to stress are the focus of this dissertation: vowel duration, consonant duration, intensity, f0, and vowel space. In what follows, each of these measurements is considered in turn.

3.2.1 Acoustic measurements

Potential acoustic correlates of stress were extracted from audio data using Praat scripts, and in some cases were normalized in R (Boersma & Weenink 2018, R Core Development Team 2020). Praat scripts were modified from original scripts by Dicanio (2017), Lennes (2018), and McCloy (2012). This section describes how each measurement was taken and how the measurements were manipulated in the analysis.

Duration

Duration measurements were extracted in milliseconds (ms) from Praat using the Dicanio script. They were determined by the alignments at word and segment level as produced by the MFA and checked manually. In Chapter 4, all duration measurements are reported

as log-normalized from the original measurements. For measurements of vowel duration, outliers shorter than 50 ms and longer than 200 ms were excluded. This amounted to the exclusion of 5,835 vowel tokens out of 87,222 total in the corpus (about 6%). Other measurements of these outlier vowels are also excluded from analysis.

Intensity

Maximum and minimum intensity measurements were taken, in decibels (dB), using the Dicanio script. Maxima and minima were extracted for each segment as well as each word. These measurements of intensity were highly variable as a direct result of the variability of the recording conditions in each set of language data. Amount of background noise, location of recording (e.g. indoors or outdoors), and distance between the speaker and the microphone all affect intensity measurements. For this reason, intensity measurements were normalized relative to the intensity of the following vowel, thus creating values that were comparable across recordings and languages.

$$I_{rel} = I_a - I_b$$

The equation for this relative intensity measure (I_{rel}) is the difference between the maximum intensity of the target vowel (I_a) and the maximum intensity of the vowel following the target (I_b). As a result, the relative intensity measure is a positive number when the target vowel has a relatively higher intensity than the vowel than follows, and it is a negative number when the target vowel has relatively lower intensity. Chapter 4 reports distributions and statistical analyses using this relative measure.

Fundamental frequency (f0)

Fundamental frequency (f0) is used as the acoustic correlate of pitch in this dissertation. Both average f0 measures for each segment and pitch tracks over each word are used in the

analysis, as described here.

Maximum and minimum f0 measurements were taken for each segment in the data using the Dicanio script. These measurements were normalized in semitones computed using each speaker's average pitch reading at the reference. Measures used in the modeling for stress correlates include the average of the minimum and maximum, as well as the pitch range (difference between maximum and minimum f0). Zhang (2018) found that this method was optimal in a comparison of sixteen f0 normalization methods, finding that it performed the best in preserving talker differences in a case study of tone in Wu dialects while normalizing over the effects of physiological differences on f0. Descriptive statistics and stress correlation results are reported using this normalized measure in Chapter 4.

F0 is also used in the investigation into prosodic phrase categories, presented in Chapter 7. For this portion of the dissertation, scripts from Kaland (2021) are used; see Chapter 7 for detailed discussion of these methods.

Vowel formants

First and second formant measurements were taken in Praat using the Dicanio script at the midpoint of the vowel. These measurements were then normalized using the average-spacing ΔF normalization method (Johnson 2020). This is a type of vowel normalization that is vowel extrinsic, speaker intrinsic, and formant extrinsic. This method normalizes vowel formant measures to an estimated vocal tract for each speaker, using the following formula (from Johnson (2020: 5-7)):

$$\Delta F = \frac{1}{mn} \sum_{j}^{m} \sum_{i}^{n} \left[\frac{F_{ij}}{i - 0.5} \right]$$

This value is used to normalize each individual formant measurement, as below:

$$F1' = \frac{F1}{\Delta F}$$

The ΔF method was chosen for this project because of its cross-linguistic consistency. This method provides vowel extrinsic normalization that, unlike other methods of this sort, does not rely on 'point' vowels that are not uniform across languages and make it difficult to compare normalized values in a typological study such as this one (cf. Fabricius, Watt & Johnson (2009), Lobanov (2005), Nearey (1978)). The values produced using this method are on the same measurement scale no matter the language, which makes differences across languages more directly comparable than they would be using other methods.

3.2.2 Determining stress correlates

Stress correlates were determined for each language and each measurement individually, and then compared to one another. Mixed effects linear regression models were run in R using the `lmerTest` package (Kuznetsova, Brockhoff & Christensen 2017) with the dependent variable being the acoustic measurement in question. This results in five separate models for each language. The maximum model for each measurement is given in (3.2). Depending on the nature of the data, some of these variables were excluded from individual language models. For example, some of the data used for this project only includes one speaker of the language, so the random effect of speaker was excluded from those models. Likewise, languages that do not have a phonemic vowel length contrast did not have vowel length included as a fixed factor in these models.

For any individual model, one or more of the factors in the maximum model may not contribute any explanatory power to the model. For each model in each language, the maximum model was cut down when needed, using the `step()` function in the `lmerTest` package (Kuznetsova, Brockhoff & Christensen 2017). In some cases, the 'stress' factor was eliminated using this function, but the non-significant results are reported in the following chapter for thoroughness. Full model results are reported in Appendix C.

(3.2) a. **Regression Model A:** vowel duration

```
lmer(vowel.duration ~ (1|word) + (1|seg.identity)
+ (1|speaker) + phonemic.length + word.finality + stress
```

b. **Regression Model B:** consonant duration

```
lmer(consonant.duration ~ (1|word) + (1|seg.identity)
+ (1|speaker) + stress
```

c. **Regression Model C:** intensity

```
lmer(rel.intensity ~ (1|word) + (1|seg.identity)
+ (1|speaker) + phonemic.length + word.finality + stress
```

d. **Regression Model D:** f0

```
lmer(avg.f0 ~ (1|word) + (1|seg.identity) + (1|speaker)
+ word.finality + stress
```

e. **Regression Model E:** vowel peripherality

```
lmer(Euclidean.dist ~ (1|word) + (1|seg.identity)
+ (1|speaker) + word.finality + stress
```

Using this method, as opposed to a logistic regression model with 'stress' as the dependent variable and continuous measurements as fixed factors, allows for more control over the other (non-stress) factors that may influence each measurement. For example, in a duration model, phonemic vowel length and word finality can be included in the regression, in order to tease apart the effect of stress from these other factors. A logistic model investigating the same question (i.e. the relationship between stress and vowel duration) cannot easily handle many control variables, which are especially crucial with naturalistic data that was not collected in highly controlled carrier sentences.

3.3 Data Quality

As discussed in the previous chapter, the data for this project were collected in a variety of circumstances that cannot be controlled for when using archival materials. Many recordings were made outside or in otherwise noisy conditions; factors such as speakers' distance from the microphone will likely vary across language samples; and different audio recorders and audio processing tools were used across samples, among any number of other factors that make these language samples variable in ways that highly controlled experimental data would not be. Because of the nature of the data being used, I performed some quality checks on each data set to determine the non-linguistic variability present in the data. These checks are summarized here.

Data quality was measured by looking at signal-to-noise ratio across all the audio for each language. This measurement is a way of measuring how loud periods of speech are compared to periods of non-speech, thus estimating the amount of background noise in a recording. The formula for the signal-to-noise ratio (SNR) is given in Equation (3.1), where P = power.

$$SNR = 10log_{10}\left(\frac{P_{signal}}{P_{noise}}\right) = 10log_{10}(P_{signal}) - 10log_{10}(P_{noise}) \tag{3.1}$$

It is important to note here that the standard way of measuring intensity in acoustic analysis, decibels (dB), is already a logarithmic transformation of the power of a signal. Because of this, the way of calculating SNR for the language data in this project only required subtracting the average intensity in dB of untranscribed segments of the recording from the average intensity in dB of segments of transcribed speaking. Calculations were done using averages across each file in each archival collection, and then averaging those ratios to get the final value as shown in Figure 3.3.

A positive SNR value indicates that the signal (in this case, transcribed speech) is louder

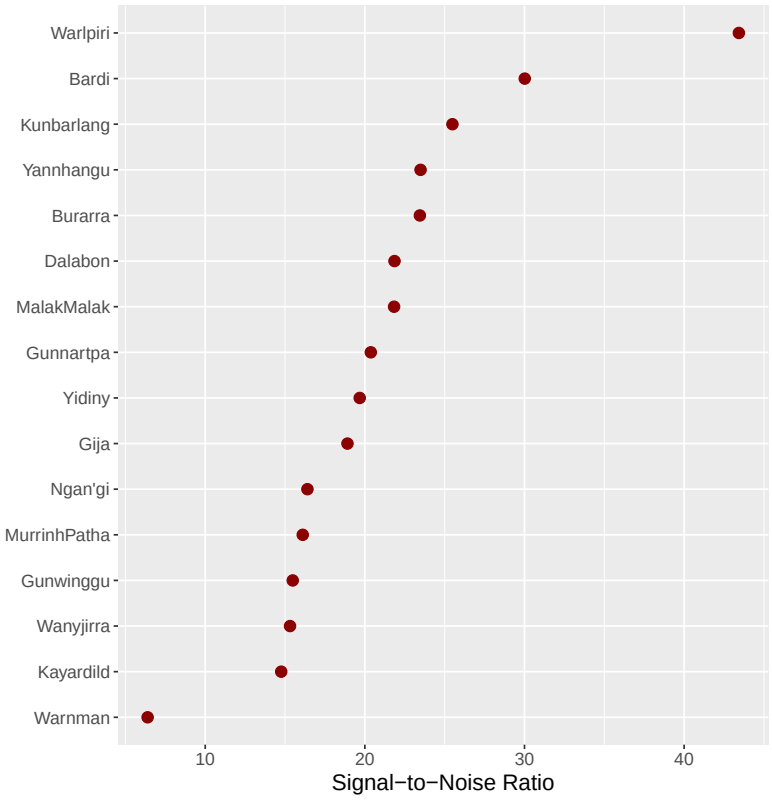

Figure 3.1: Average signal-to-noise ratio for each language.

and therefore more distinct that the background noise (any periods of silence, background noise, and untranscribed speech). All of the language collections in Figure 3.3 have positive SNR values. However, the standard SNR for high quality audio is generally considered to be 50 dB (Sanker et al. 2021), which only one collection, Warlpiri, approaches. These collections are consistently noisier than audio collected in laboratory conditions, as is expected given the uncontrolled nature of the recording conditions. However, I also suggest that these values may be artificially deflated because of the nature of transcription, rather than recording conditions and background noise. That is, the transcribed speech in these recordings are not necessarily the only speech sounds present in the audio. The files will often include speech from the researcher (usually in English), and may additionally include speech from other target-language speakers that happens to remain untranscribed. So, what is measured here as a signal-to-noise ratio may include signal in the noise portion and artificially deflate the measure. With such a general measurement as SNR, it is difficult to determine to what extent this confounding factor applies for each collection and each file. But it is very likely that there is a lot of speech in the recordings that is classified as "noise" due to its being untranscribed, that inflates the average intensity of noise segments without actually indicating noisy audio. Some of this speech would be the speech of the researcher collecting the audio; their questions, explanations, and conversational turns are not necessarily transcribed in these materials, especially when they are not in the target language. The other major source of untranscribed speech could be contributions from other speakers in a group who did not agree to contribute officially, or who happen to be visiting the recording location while the audio is being collected. Untranscribed speech may also come from children playing in the background and other sources, which may more conceivably be considered background noise for the purposes of determining audio quality because this speech may overlap with the intended speech audio and interfere with the integrity of the signal.

The methods used for measurement normalization were outlined in §3.2. It is important to note, given the variability in data quality summarized here, that the normalization methods chosen for these measurements, especially intensity, were chosen with this variability in audio quality in mind, in order to eliminate audio quality as a confounding factor to the greatest extent possible.

Chapter 4

Results: Cross-linguistic variation

The best evidence in support of Claim (9.1), that cues to stress are linguistically heritable, is to establish that (a) each language uses a certain set of prominence cues in a consistent way, and (b) that these cues vary across languages, with more closely related languages being more similar in their use of stress cues than more distantly related ones. This requires first establishing what the cues to stress are in each of the languages in this study, and then comparing these cross-linguistically. This chapter presents the results[1] of language-by-language investigations into the acoustic correlates of stress, with a focus on cross-linguistic variation.

In what follows, I delve into the results of the acoustic study outlined in Chapter 3 with a focus on cross-linguistic variation in the acoustic correlates of lexical stress. Variation in these correlates across speakers of the same language is discussed in Chapter 5. Here, I present the descriptive facts about each acoustic measurement investigated in the dissertation, along with each of these factors' contribution to stress marking in all languages. These measurements are: vowel duration (§4.1), onset and coda consonant durations (§4.2), in-

1. Full R markdown output with all model results is available online at https://doi.org/10.5281/zenodo.6354645.

tensity (§4.3), f0 (§4.4), and vowel space (§4.5). A summary of acoustic stress correlates across each language in this study is given in §4.6; overall, much variation is observed across languages. While f0 and vowel duration are the most common correlates of stress in these languages, all of the tested acoustic factors correlate with stress in at least one of the languages in this study.

4.1 Vowel Duration

Measurements of vowel duration have been binned into four categories based on stress status and phonemic vowel length. All vowels shorter than 50 ms and longer than 200 ms were excluded, and these measurements were then log transformed. Vowel duration is a significant correlate of stress in six of the languages in this study.

The distribution of durations in the short vowels for each language are shown in Figure 4.1, binned into 'stressed' and 'unstressed' categories. These distributions have long right tails, with some languages showing a more defined peak at the leftward edge, roughly between -3.0 (around 50 ms) and -2.3 (around 100 ms).

Not all Australian languages have phonemically long vowels, and they are often infrequent when present. As can be seen in Fig. 4.2, only 3 of 16 study languages have at least 5% long vowel tokens in their respective corpora: Bardi, Kayardild, and Yidiny. In these languages, long vowels occur in both stressed and unstressed contexts. The languages with under 5% long vowels fall into one of two categories. In one category, the infrequent long vowels are distributed in this same way, with long vowels in both stressed and unstressed syllables (cf. Dalabon, Burarra). In the other category, the small number of long vowels are exclusively or almost exclusively stressed (cf. Warnman, Warlpiri, Murrinh Patha). In these cases the total number of long vowels in the corpus is under 10.

With the relative proportions of long vowels in mind, the distribution of long vowel durations are given in Fig. 4.3, grouped into stressed and unstressed groups. In many

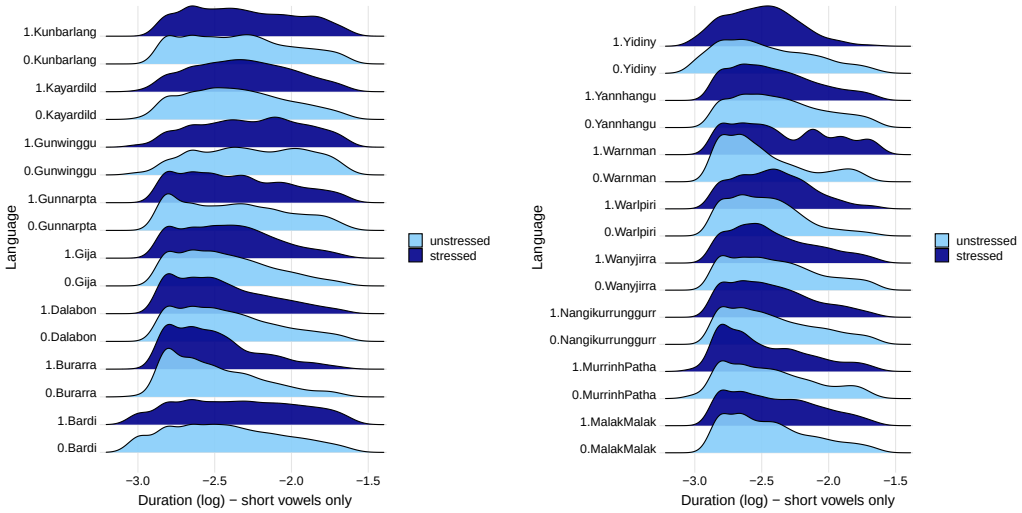

Figure 4.1: Log-transformed durations of short vowels, by language.

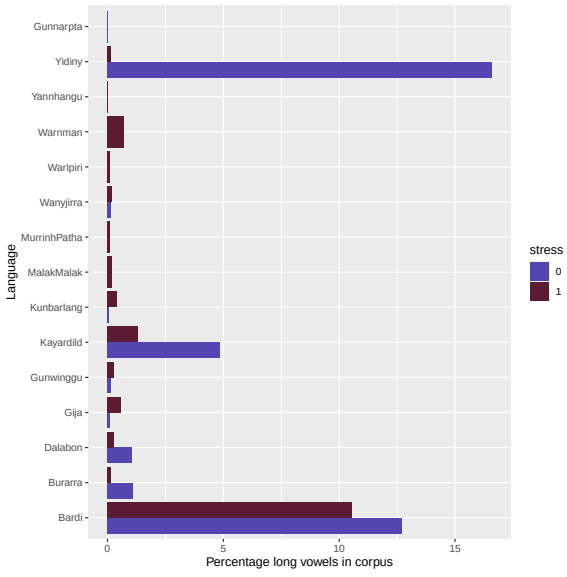

Figure 4.2: Percentage of phonemically long vowels in the corpus, by language.

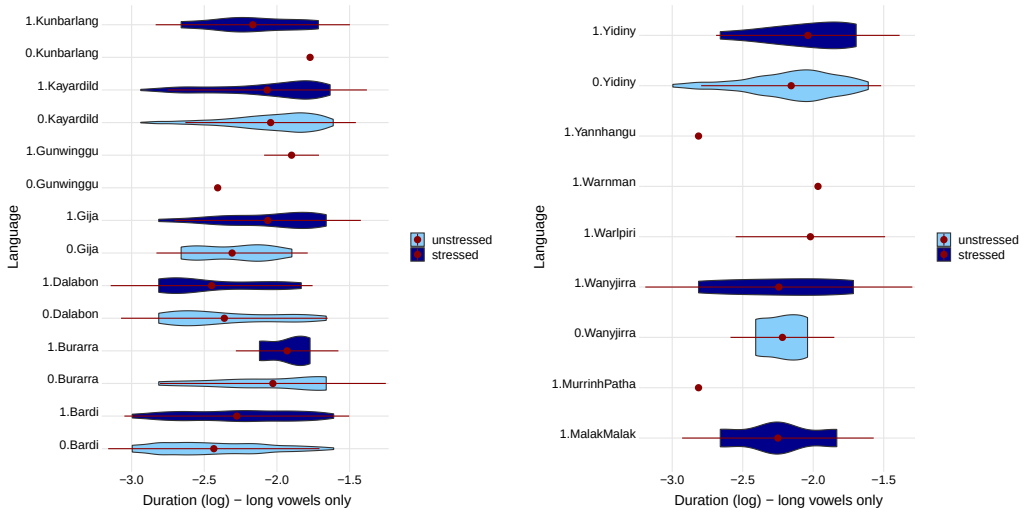

Figure 4.3: Log-transformed durations of long vowels, by language. Red dots mark mean value, red lines indicate length of one standard deviation from the mean.

languages (e.g. Malak Malak, Burarra) the numbers of long vowels are so low that the smoothed distributions have a strange looking shape. The numbers of long vowels are only large enough in three languages to show some reliable distribution that looks somewhat normal: Bardi, Kayardild, and Yidiny. For this reason, these languages will be discussed specifically and the remaining languages' distributions will not be investigated in much detail.

Both Kayardild and Yidiny long vowels have a normal-looking distribution with long left tails, adding a complement to their short vowel distributions which were roughly normal with long right tails. In Bardi, the distribution of long vowels looks much flatter, and this again matches with the language's short vowels which also have a very distributed spread of duration measurements.

When the number of long vowels in a language is very low, this factor was left out of the statistical analysis as the models could not make predictions based on so little data. Figure 4.4 summarizes the effect of stress on vowel duration in each language, as determined using the method for Regression Model A, described in Chapter 3.

Duration is a common correlate of stress, as shown in Figure 4.4. Each dot in this plot shows the regression model estimate value for duration when stressed, relative to duration when unstressed, and each line shows the standard error of this value. When the line does not cross the zero mark (the dark dotted line), the effect is significant at $\alpha = 0.05$, while lines that do cross the zero mark are not. The unit of measure is logarithmic, as is the duration measure input as the dependent variable for Regression Model A. This means that an effect of $+0.1$ represents a roughly 6.7% increase in duration, since the range of duration measurements is about 1.5 logarithmic units. Gunnartpa, Yidiny, Bardi, Wanyjirra, and Malak Malak all have significant effects of about this size. For the purposes of categorizing relative size of effects here, I consider an effect at or above $+0.1$ to be a 'large' effect of stress on duration. The other languages with significant effects, Kayardild, Warlpiri, Yan-nhangu, Ngan'gi, and Gija show what I will consider a 'moderate' effect size. The models for these languages predict stressed syllables to be around 3% longer than unstressed ones. The remaining five languages in Figure 4.4 did not have a significant effect of stress.

Another factor included in Regression Model A that is generally of interest for studies of word-level prosody is word finality. Languages often show increased duration in word-final syllables, often as a marker of word boundaries. As shown in Figure 4.5, fourteen of the languages in this study have some significant effect of word finality on vowel duration. In fact, in the cases of Gunnartpa, Yan-nhangu, and Yidiny, word-final vowels are around 10% longer than non-final vowels. Only Kunbarlang and Gunwinggu have no word-final lengthening effect, with the remaining languages lengthening final vowels $3-7\%$ over their non-final counterparts.

These word-final lengthening effects underscore the many uses of vowel duration in these languages. Along with lengthening final vowels to mark word boundaries, many of these languages also use duration differences in their phonologies, as they have phonemic length distinctions. As shown in Figure 4.4, many of these languages additionally use dura-

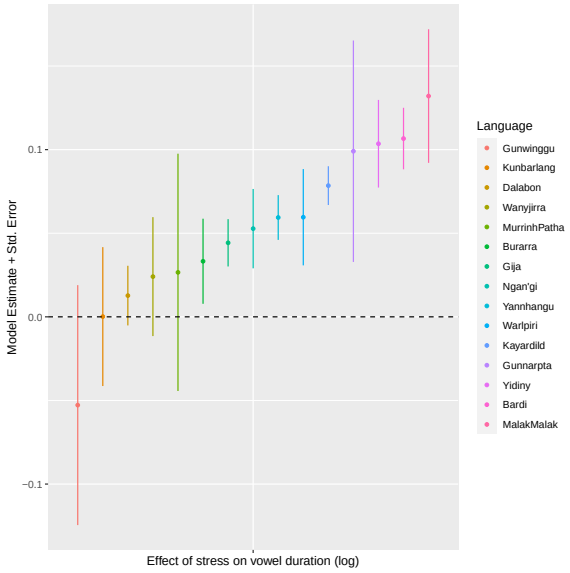

Figure 4.4: Results of regression model A; model estimate and standard error values for binary factor 'stress' shown. In legend, topmost labels correspond to leftmost dot-whiskers. Lines that cross the zero mark (dark dashed line) represent non-significant model results.

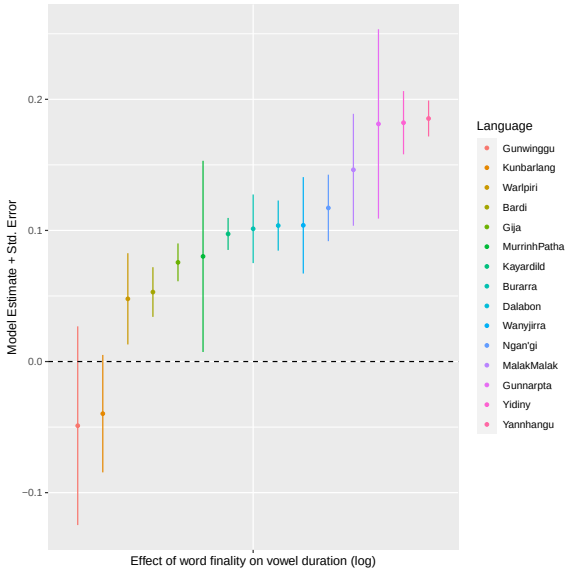

Figure 4.5: Results from regression model A; model estimate and standard error values for binary factor 'finality' shown. In legend, topmost labels correspond to leftmost dot-whiskers. Lines that cross the zero mark (dark dashed line) represent non-significant model results. Lines that cross the zero mark (dark dashed line) represent non-significant model results.

tion to some extent in marking stress. It is possible that the use of vowel duration differences to mark a phonemic contrast in vowel quality in these languages increases the salience of duration differences in all contexts, thus making further meaningful uses of duration differences more probable than average.

These observations seemingly run counter to the Functional Load Hypothesis of Berinstein (1979). This hypothesis, and this particular use of the term 'functional load,' refers back to an argument in Berinstein's thesis (quoted here in (4.1)) to explain why duration is not a perceptual correlate of stress in K'ekchi, a finding which seemingly ran counter to the prevailing idea of a universal hierarchy of stress correlates as put forth in Hyman (1977).

(4.1) "Change in F0, increased duration, and increased intensity, in that order, constitute the unmarked universal hierarchy for perception of stress in languages with no phonetic contrasts in tone or vowel length; in languages with such contrasts the perceptual cue correlated with that contrast (i.e. F0 with tone and duration with length) will be superseded by the other cues in the hierarchy." (Berinstein 1979: 2)

This hypothesis predicts that the presence of phonemic vowel length in a language inhibits the meaningful use of vowel duration differences elsewhere; similarly, the presence of phonemic tonal distinctions should inhibit the use of f0 to mark stress, and so on.

Recent work testing the predictions of Berinstein's Functional Load Hypothesis (FLH) has produced mixed results. It was more or less upheld by Vogel, Athanasopoulou & Pincus (2016), which considered a survey of prominence cues in Greek, Hungarian, Spanish, and Turkish. In Hungarian, the only language of this group with phonemic vowel length, duration was not used to mark stress. Evidence is also found in this study for an extension of Berinstein's FLH predicting that stress and focus marking should not use the same acoustic cues, though Hungarian serves as a counter-example to this point (Vogel, Athanasopoulou & Pincus 2016: 37). However, Lunden et al. (2017) find no relationship between

the presence of phonemic length contrasts and use of duration as a cue to lexical stress in a larger study of stress correlates that sources information from published theses and other academic works. Out of 82 languages in this study that do have contrastive vowel length, 45 (54.9%) use duration as a stress correlate. This is only slightly less frequent than among the 58 languages that lack contrastive vowel length; 38 of these (65.5%) use duration as a cue to stress. This difference in frequency is not found to be statistically significant and the authors conclude that Berinstein's FLH is not supported (Lunden et al. 2017: 573-574).

While the FLH may hold and even have some cognitive reality in some languages, the results just presented for Australian languages provides more evidence problematizing the universality of this claim and provides an argument for including a more diverse array of languages in studies of this kind.

4.2 Consonant Duration

4.2.1 Post-tonic lengthening

Post-tonic consonant lengthening has been noted as a common correlate of stress in Australian languages (Fletcher & Butcher 2014, Fletcher et al. 2015). This lengthening occurs after a stressed vowel, regardless of the consonant's syllable membership (i.e. both codas and onsets of the following syllable are affected). Fletcher & Butcher (2014) have claimed that post-tonic lengthening is present in most if not all Australian languages. However, evidence from some languages suggests that this is not a universal across the continent. Some of these refuting studies are Jepson, Fletcher & Stoakes (2019) for Djambarrpuyngu, Fletcher et al. (2015) for Mawng, and Pentland (2004) for Warlpiri, among others.

Based on the findings in previous work on the post-tonic consonant lengthening phenomenon, consonants were binned into three categories; stops, nasals, and glides were investigated for this effect separately. As the following results demonstrate, some languages

only show this effect in one of the categories, and only one language (Kayardild) has an effect in all three categories.

Stop consonants are the category most likely to show an effect of this post-tonic lengthening phenomenon. Four languages— Kayardild, Warlpiri, Yidiny, and Murrinh Patha— have significantly longer stops post-tonically than otherwise (see Fig. 4.7). As the range of consonant durations is around 2 logarithmic units (cf. Fig. 4.6), an effect of 0.1 units indicates a difference in duration of about 5%. Yidiny and Murrinh Patha both have significant effects over +0.2, meaning that stressed vowels are associated with post-tonic lengthening of over 10% on average.

Only two languages have an effect of the stress factor on post-tonic nasal duration: Kayardild and Gunwinggu (Fig. 4.8). These effects are both around about +0.1, indicating an increase in duration of about 5% on average. While Kayardild shows this effect for all post-tonic consonant groups, for Gunwinggu this effect is only seen in the post-tonic nasals.

Three languages — Kayardild, Gija, and Warlpiri — have a significant effect of stress status on the duration of post-tonic glide consonants, as shown in Fig. 4.9. These effects are again around +0.1 or about 5% longer on average after stressed vowels compared to following unstressed ones. Kayardild shows this effect in all consonant groups. Warlpiri shows this effect in the stop and glide groupings only; and for Gija the glide consonants are the only group where a post-tonic lengthening effect is seen.

The correlation of stress with post-tonic consonant lengthening is not especially common in the languages surveyed in this dissertation. This is further evidence against the generalization of post-tonic lengthening as extremely common across languages of Australia. However, as has been seen, post-tonic consonant lengthening does occur in six of these 16 languages for at least one grouping of consonant types.

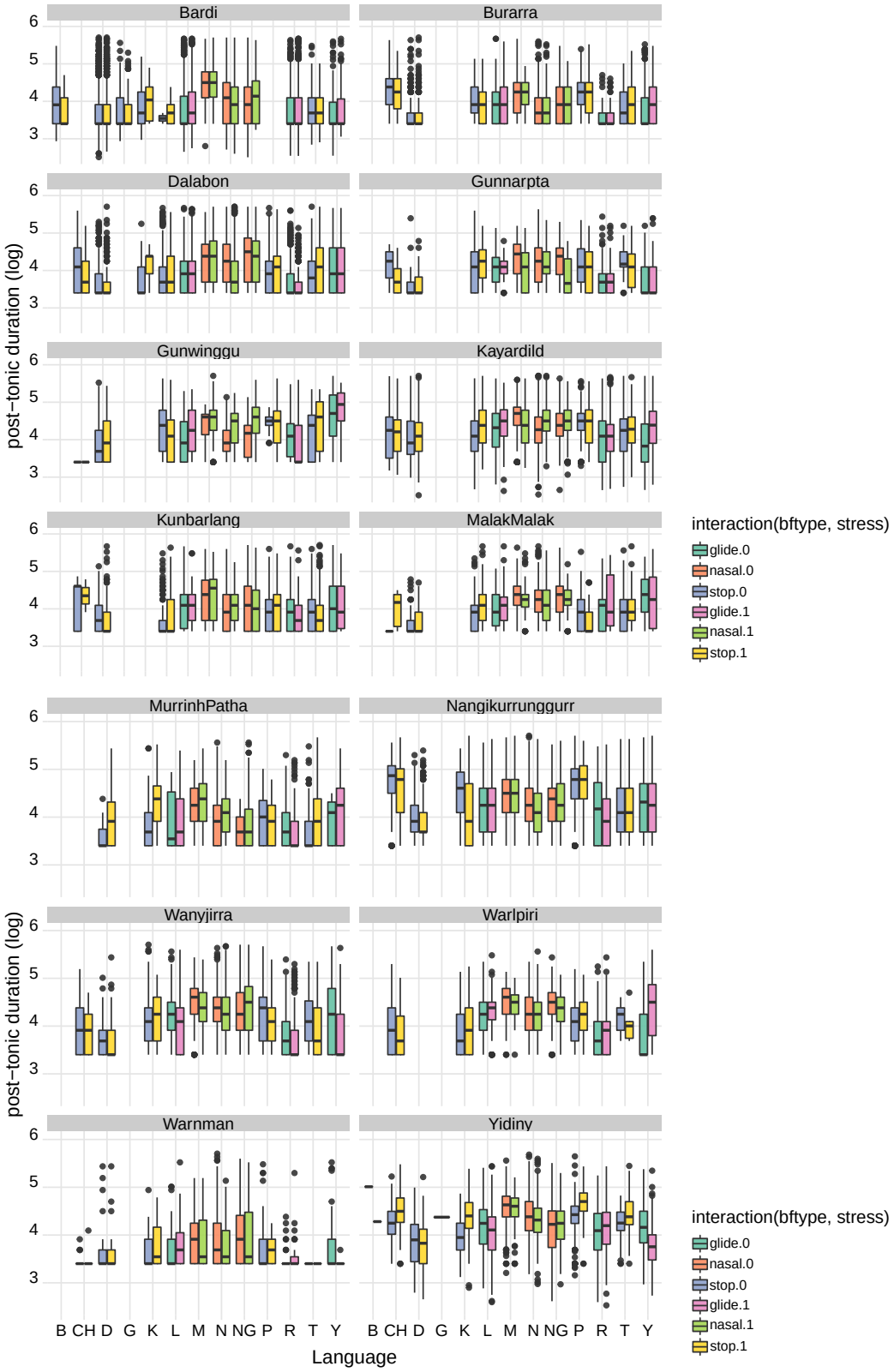

Figure 4.6: Distribution of consonant durations in post-vowel position.

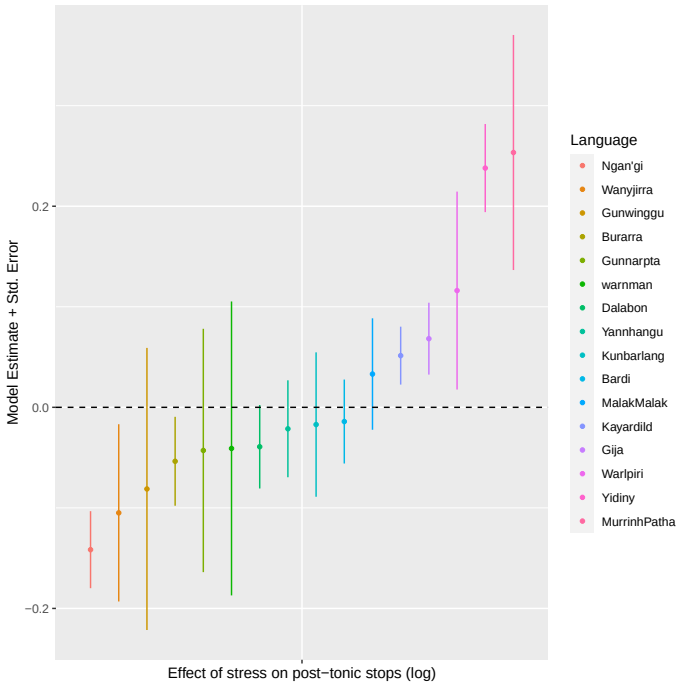

Figure 4.7: Model effect of the fixed binary factor 'stress' on duration of the following stop consonant.

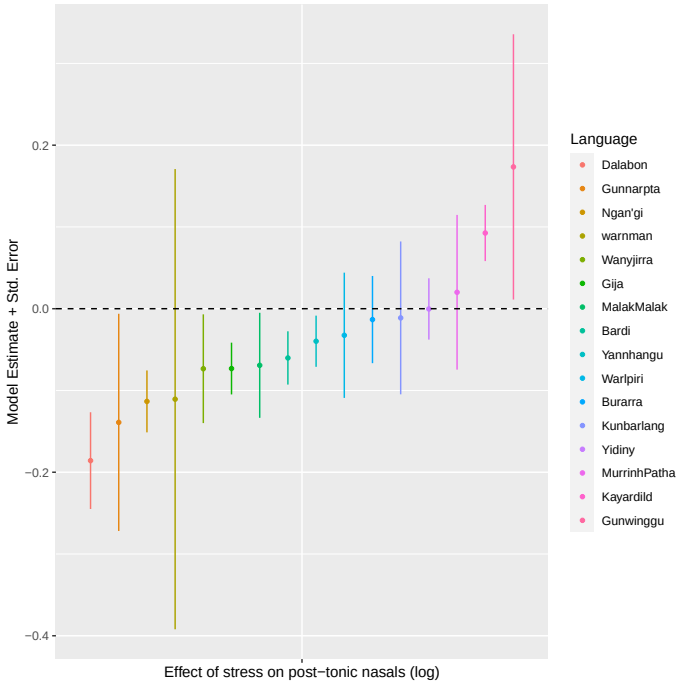

Figure 4.8: Model effect of the fixed binary factor 'stress' on duration of the following nasal consonant.

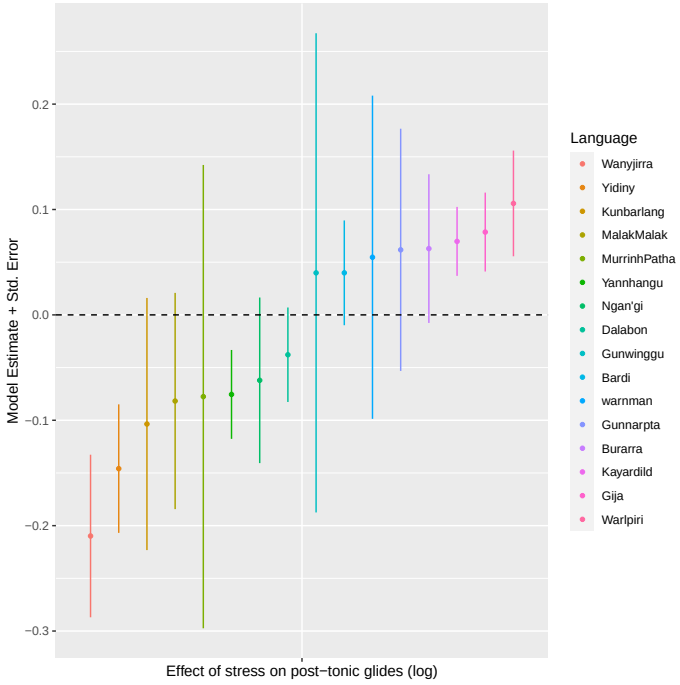

Figure 4.9: Model effect of the fixed binary factor 'stress' on duration of the following glide consonant.

4.2.2 Onset duration

Fletcher & Butcher (2014) and others have noted that duration of the pre-tonic onset consonant is often lengthened in Australian languages. Similarly to post-tonic lengthening, this effect may differ across the consonant categories stop, nasal, and glide, and these groups are considered separately in the results reported here. The distribution of onset consonant durations is presented in Figure 4.10, which color-codes each consonant segment into one of these three categories, and splits duration distributions by stress.

More languages have a correlation of onset stop duration and stress than either onset nasal or onset glide duration and stress. Nine of sixteen languages in Fig. 4.11 show significantly longer stop duration in onset position of stressed syllables when compared to unstressed syllables. These effects are often larger than any of the effects seen for post-tonic lengthening as well. Again assuming a distributional range of roughly 2 logarithmic units,

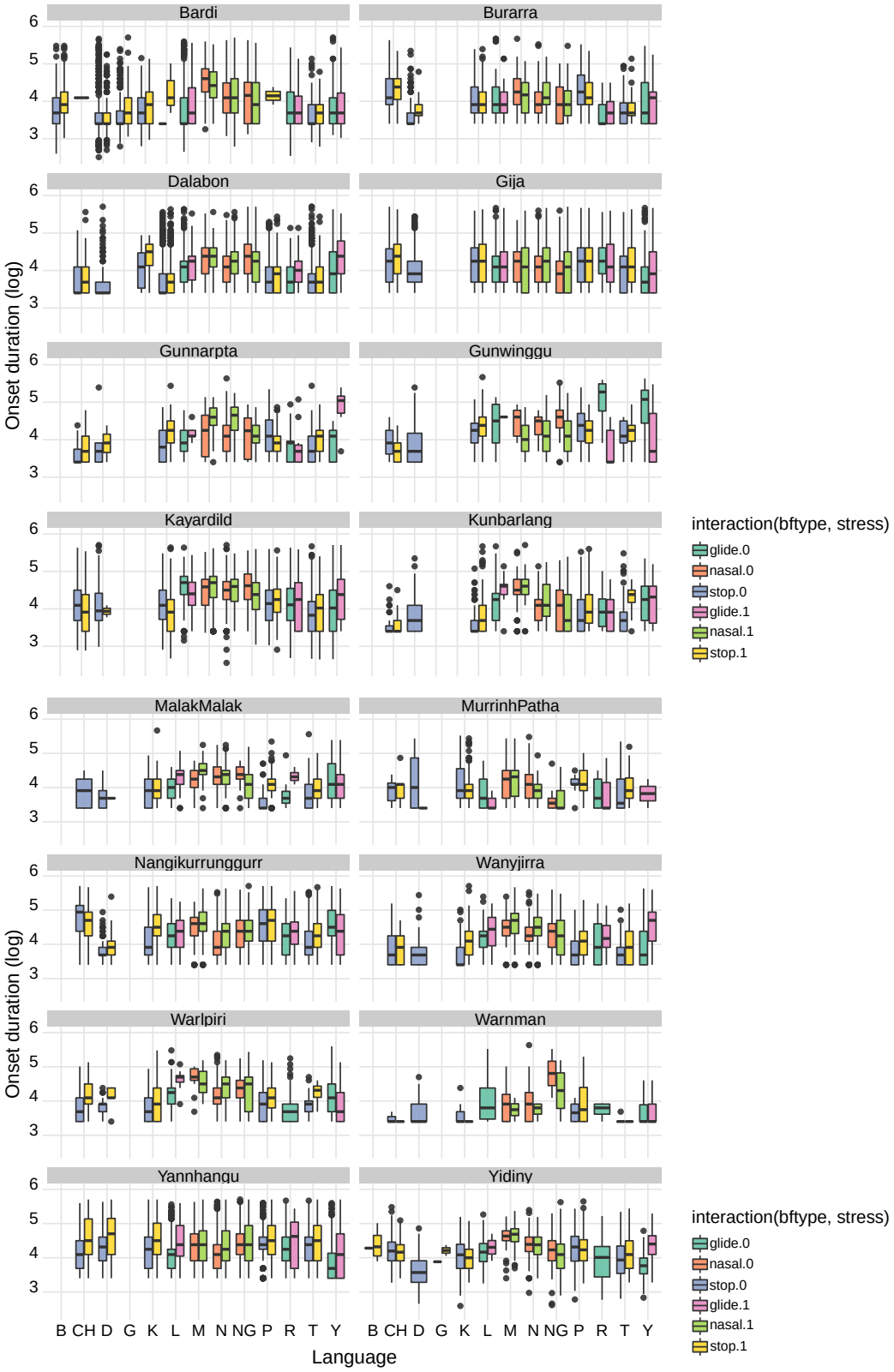

Figure 4.10: Distribution of consonant durations in onset position.

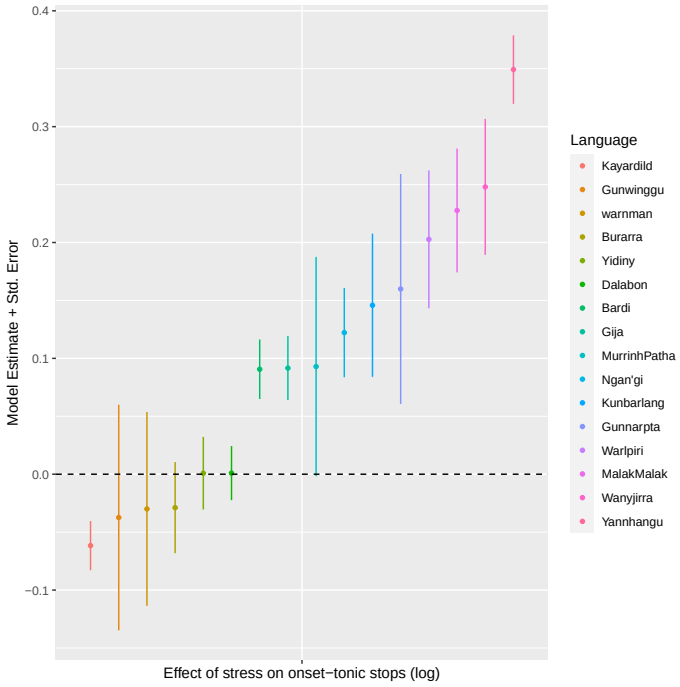

Figure 4.11: Model effect of the fixed binary factor 'stress' on duration of the onset stop consonant.

the range of effect sizes in Fig. 4.11 ranges from 5% (+0.1) to 17.5% (+0.35) longer onset consonants in stressed syllables when compared to unstressed ones. The largest effect is observed for Yan-nhangu with an effect size of +0.35; this language also shows significant and rather large effects for the other consonant categories.

Five languages show an effect of stress on the duration of onset nasal consonants in Fig. 4.12. These languages— Gunnartpa, Yan-nhangu, Ngan'gi, Gija, and Wanyjirra— also have significant effects of stress on the duration of onset stop consonants. The effect sizes here range from around +0.1 to +0.25, or around 5-12% longer nasal consonants in the onsets of stressed versus unstressed syllables.

Finally, seven languages in this data set show significantly longer glides in the onsets of stressed versus unstressed syllables. Four of these languages— Gunnartpa, Gija, Wanyjirra, and Yan-nhangu— have effects here along with the other consonant categories.

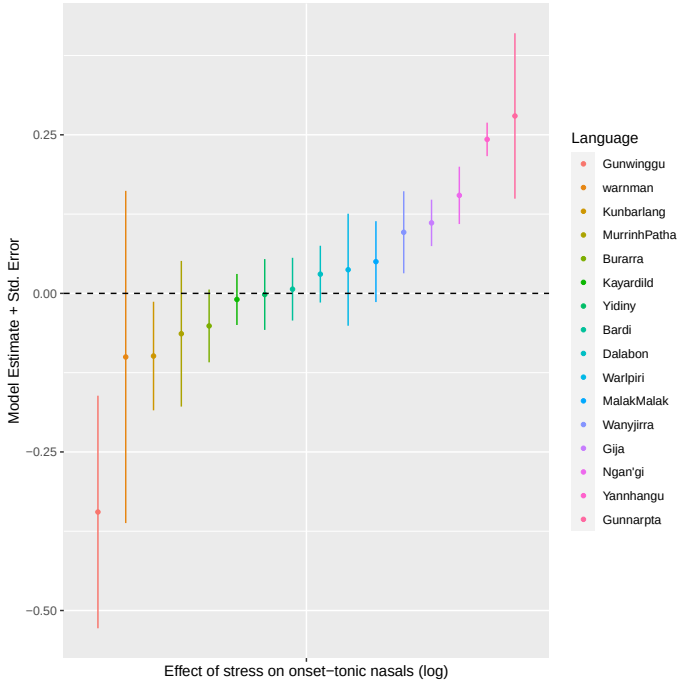

Figure 4.12: Model effect of the fixed binary factor ‘stress’ on duration of the onset nasal consonant.

Malak Malak has an effect for glides and stops but not nasals, and the other languages with significant effects here, Yidiny and Dalabon, only have an effect of onset lengthening in the glide consonants.

Overall, 11 of the 16 languages in this sample have an effect of onset lengthening in stressed syllables for at least one of these consonant groupings. This is the same number of languages that use vowel duration as a correlate of stress, although these are not entirely overlapping groups, and it is almost twice as many languages that show some amount of post-tonic lengthening.

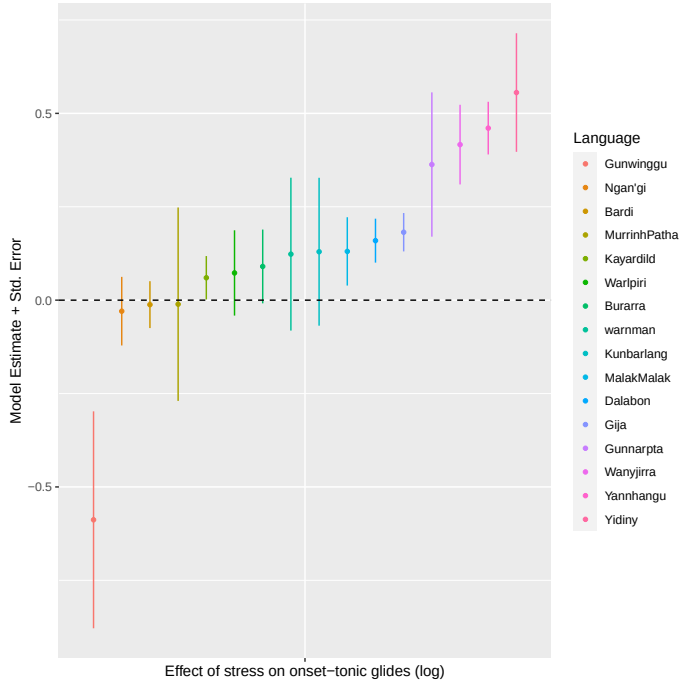

Figure 4.13: Model effect of the fixed binary factor 'stress' on duration of the onset glide consonant.

4.3 Intensity

Measures of intensity are highly variable, especially given that the data sources used in this project were recorded at different times, in different locations, and in different recording situations. For this reason, intensity measurements were normalized relative to the intensity of the following vowel. The equation for this relative intensity measure (I_{rel}) is the difference between the maximum intensity of the target vowel (I_a) and the maximum intensity of the vowel following the target (I_b).

$$I_{rel} = I_a - I_b$$

As a result, the relative intensity measure is a positive number when the target vowel has a relatively higher intensity than the vowel than follows, and it is a negative number when the target vowel has relatively lower intensity.

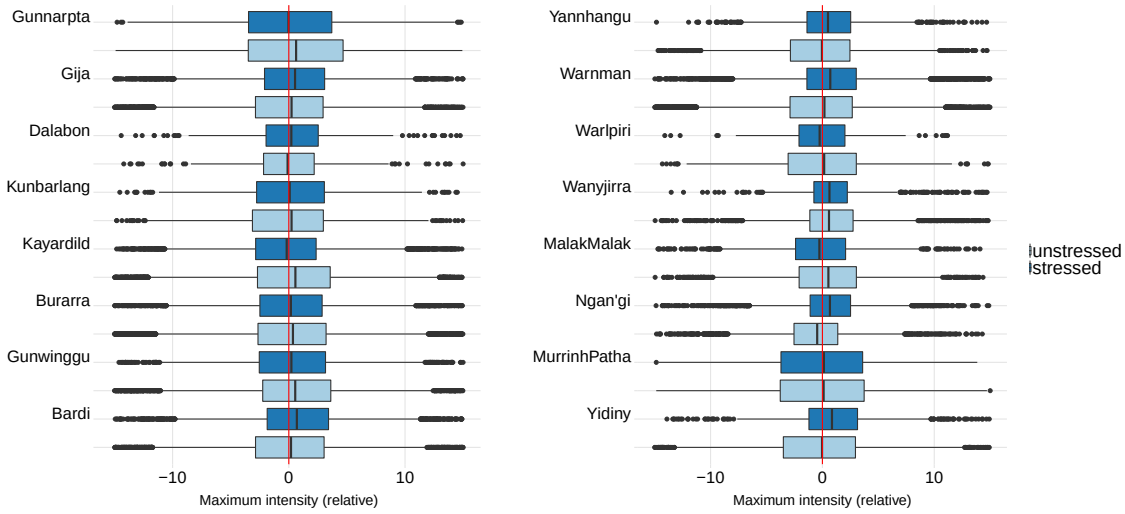

Figure 4.14: Distributions of relative intensity measure, grouped by stress status. Stressed vowels shown in dark blue.

Figure 4.14 shows the distributions of this relative intensity measure for each language, for both stressed and unstressed vowels. These distributions can be interpreted as follows. When intensity tends to lessen gradually over the course of the word, both stressed and unstressed vowels will have a distributional mean above zero, as each syllable will be more intense than the following one. When both stressed and unstressed vowels have means around zero, there is no variation in intensity based on stress or syllable position. When only stressed vowels have a distributional mean above zero, and unstressed vowels have a mean close to zero, we may expect that intensity is a correlate of stress in that language. Such a distribution indicates that stressed syllables are consistently higher than the ones that follow it, while consecutive unstressed syllables tend to hold a steady intensity, i.e. there is not a positional effect. A potential correlation of intensity with stress based on this metric is observed in six languages: Gija, Bardi, Yan-nhangu, Warnman, Ngan'gi, and Yidiny.

Despite these apparent trends in the data distributions, the regression model results in Figure 4.15 do not show intensity and stress to be correlated in all of these languages. Bardi, Yidiny, and Ngan'gi do show significant effects of stress here, along with Malak Malak,

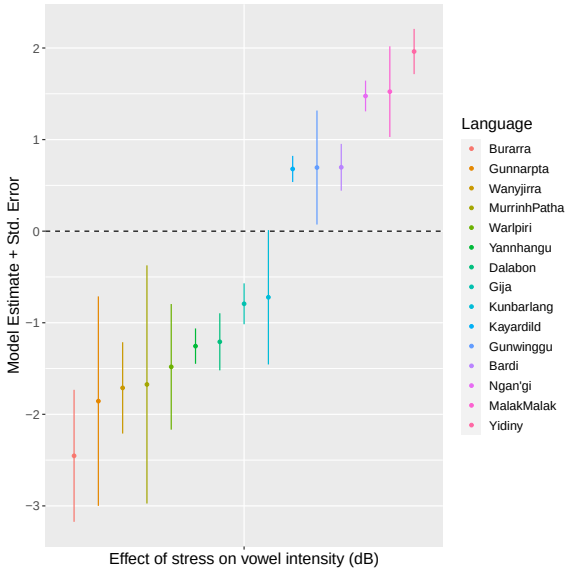

Figure 4.15: Results of regression model C; model estimate and standard error values for binary factor 'stress' shown.

Gunwinggu, and Kayardild. All of these effects are small, with estimated values for stressed vowels being between 0.5 and 2 dB more intense than unstressed vowels. Given that the range of relative intensity values in Figure 4.14 ranges something like 25 dB, this amounts to an increase in intensity in the range of 2 to 8%. This effect may well be perceptually salient for some speakers, especially speakers of Yidiny which is on the high end of this range, but follow-up study would be needed to test this.

All other languages besides Kunbarlang have significant effects in the opposite direction; stressed syllables are significantly less intense than unstressed ones. It is important in these cases to remember that the stressed syllables in this case are all word-initial as well. As there is not an argument for stress or any sort of prominence to be marked by a quieter vowel, I conclude from the results in Figure 4.15 that intensity is not a salient marker of stress in any of these languages with significant negative effects. Perhaps another factor is driving these effects, such as a non-word initial high phrasal tone that increases intensity elsewhere and results in these negative effects initially. A solution to this issue could be

resolved by including information about phrasal prosody in the regression models. This is beyond the scope of the current thesis but is an avenue of future research; for some investigation into phrasal contours, see Chapter 7.

4.4 F0

As discussed in Chapter 3, increased f0 is often cited as the primary correlate of initial stress in Australian languages as a whole. Using f0 to mark consistent initial stress may function not just as a stress correlate, but also as a prosodic marker of phonological word and phrase boundaries. This section considers both the distributions and correlation with stress of f0, as in previous sections, and the findings from an investigation into the anchoring of the pitch gesture to various landmarks at the beginning of the word.

4.4.1 F0 maximum

The distributions of normalized f0 measurements are given in Figure 4.16. Some clear distributional differences in stressed versus unstressed vowels are present in these data. For example, the stressed distributions for Yidiny, Yan-nhangu, and Warlpiri are clearly centered at a higher value than the corresponding unstressed distributions. These suggest a correlation with stress as has been noted to be common among Australian languages.

These distributional trends hold up in the regression model results shown in Figure 4.17. Yidiny, Yan-nhangu, and Warlpiri all have correlations predicting stressed syllables to have f0 values about 1 semitone higher than unstressed ones. As the range of distributions in Figure 4.16 is about 20 semitones, this amounts to stressed syllables being about 5% higher than unstressed. This size of effect is also seen for Malak Malak, and slightly smaller effects of about 2.5% are found for Wanyjirra, Bardi, Gunwinggu, and Kayardild. Gija shows a very small effect of stress, but no significant effect is found for Burarra, Dalabon, Kunbarlang, or Gunnartpa. Finally, Ngan'gi and Murrinh Patha both have small but

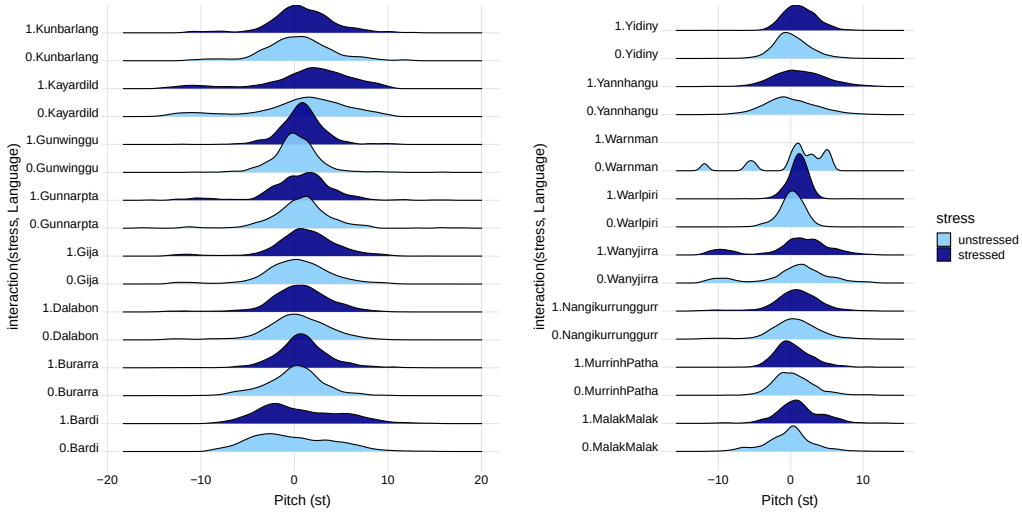

Figure 4.16: Distribution of normalized f0 measurements (in semitones) for each language, grouped by stress status.

significant negative effects, predicting stressed (initial) syllables to be slightly lower in pitch than unstressed ones. In these cases, and similarly to the conclusion drawn in §4.3, I do not propose that this effect has to do with stress. Instead, the effect may be due to some other factor, especially if there are higher-level phrasal contours that move the maximum f0 to another location in the word, and f0 maximum does not correlate with stress at all.

It is particularly notable that all of the Pama Nyungan languages in this study show significant effects of stress on f0 measurements. Generally speaking, and partially as a result of the family's large size, the Pama Nyungan family has been studied more extensively by linguists than non-Pama Nyungan languages have been. As these results suggest that f0 is in fact a very common and strong correlate of stress in Pama Nyungan languages, it stands to reason that a generalization would be made regarding all languages of Australia that f0 is the most common correlate of stress. However, these results suggest that this is not a universally true fact about Australian languages, and that f0 may be a less common correlate of stress particularly outside of the Pama Nyungan family.

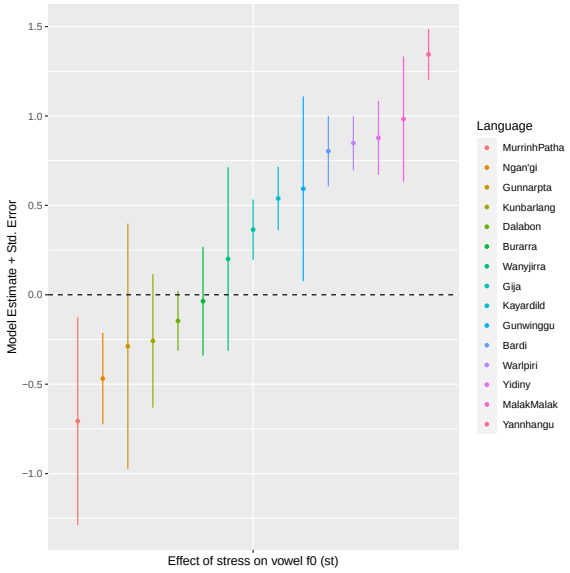

Figure 4.17: Results of regression model D; model estimate and standard error values for binary factor ‘stress’ shown.

4.4.2 F0 range

Some languages may not correlate especially high f0 with stress, instead using steeper f0 contours on stressed syllables when compared to unstressed ones. This phenomenon would be reflected in a greater f0 range across the vowel, which this section investigates. The distributions of f0 range across languages are presented in Fig. 4.18, grouped by stress category. While these range values are often smaller than 2 semitones, they can be quite large, up to around 10 st. Some languages, especially Bardi, Gija, and Kunbarlang, visibly have higher mean f0 range values in stressed vowels compared to unstressed.

The results of regression modeling with f0 range as the dependent variable are shown in Fig. 4.19. As observed in the distributional data, Bardi, Gija, and Kunbarlang all have significant positive effects of the binary ‘stress’ factor on f0 range, along with Kayardild as well. The sizes of these effects are around +0.25 semitones (around +0.35 for Kunbarlang), which is an increase in f0 range of around 2.5% given the distributional range.

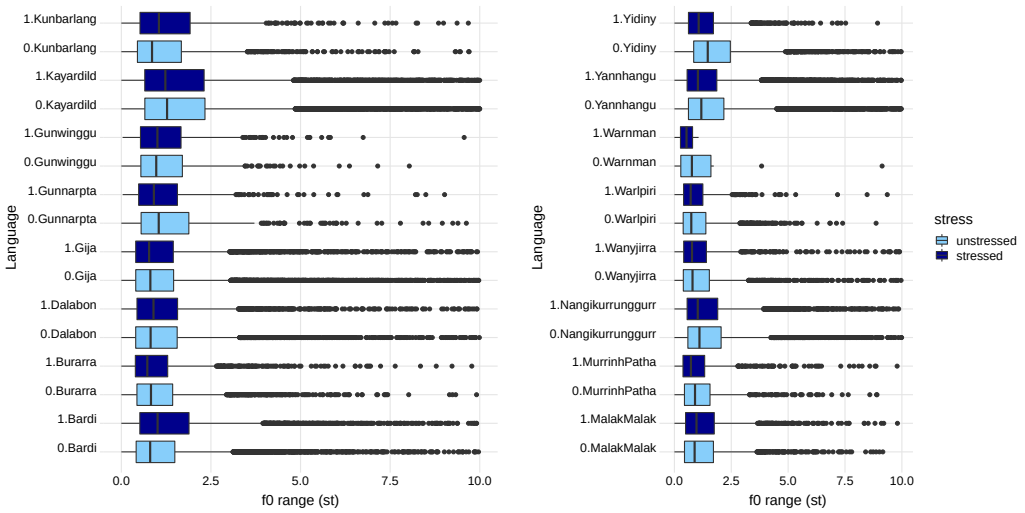

Figure 4.18: Distribution of normalized f0 range (in semitones), grouped by stress status.

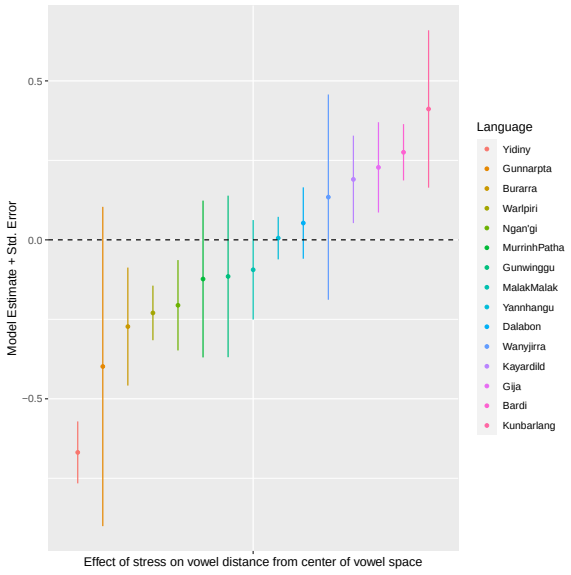

Figure 4.19: Model effect of the fixed binary factor ‘stress’ on f0 range.

It is important to note that Bardi, Gija, and Kayardild also have significant positive effects of stress on f0 maximum, as just discussed in the previous section. It stands to reason that this relatively small effect of f0 range may serve as a secondary correlate of the high f0 stress cue, reflecting a steep f0 fall as a result of the high f0 maximum, so a large f0 range may not be a major cue to stress here but a secondary correlate of what we have already seen. Kunbarlang, on the other hand, does not have an effect of stress on f0 maximum, and has the largest effect in Fig. 4.19, suggesting that f0 range may be a more salient cue to stress in this language. Again, this is likely indicating a steep f0 fall, but in this case the fall does not necessarily coincide with an especially high f0 maximum.

Four languages— Yidiny, Burarra, Warlpiri, and Ngan'gi— have significant negative effects of stress on f0 range. This may mean one of two things. It is possible that stress is marked with a high f0 that is held steady over the syllable, so that f0 range is smaller in these syllables than in unstressed ones. This may be the case in Yidiny and Warlpiri, for example, as we see a confluence of f0 maximum and lower f0 range as correlates to stress. In the case of Burarra, which has no significant effect of f0 maximum, and Ngan'gi, which has a significant negative effect of f0 maximum, this is likely not a stress correlate at all. Instead, this may again be an effect of some phrasal f0 peak occurring elsewhere in words with both high f0 maximum and high f0 range. Future work will focus on integrating phrasal contours with lexical stress in order to tease these two apart.

4.5 Vowel space

Vowels in some Australian languages have been found in some cases to be more peripheral when stressed, and in other cases to be more centralized (Fletcher & Evans 2002, Fletcher & Butcher 2003, 2014) Normalized vowel measurements for each language are given in Figure 4.20 for the purposes of comparing the general shape of vowel spaces cross-linguistically; the side-by-side plots are too small to investigate them in great detail. Larger versions of

these plots are available in Appendix C.

Vowel peripheralization is not a straightforward phonetic factor to measure. More peripheral high vowels have a lower F1, while more peripheral low vowels have higher F1, for example, and a similar trade-off relationship exists for F2 between front and back vowels. To determine a measure that can indicate a vowel's distance from the center of the vowel space regardless of the direction of the distance (higher or lower, fronter or backer), I calculated the Euclidean distance d of each vowel token from the mean of the vowel space for each speaker in each language, using the normalized $F1'$ and $F2'$ values as described in §3.2.

$$d\left(\mu(F'),F'\right) = \sqrt{(F1' - \mu(F1'))^2 + (F2' - \mu(F2'))^2}$$

The result of the above formula is an absolute value that is agnostic to the direction of the deviation from the center of the vowel space. Along with this generalized measure, vowel quality must also be considered to account for average peripherality of each individual vowel in a language. This factor is included in the regression models as a random intercept of segment identity.

Figure 4.21 shows the distribution of Euclidean distance vowels for the three vowel phonemes shared by all the languages included in this dissertation: /a/, /i/, and /u/. The mean values are often around 0.2 units on the ΔF normalization scale, although some individual vowel tokens have Euclidean distances up to 1.5 units, especially for the /i/ vowel which is canonically further from the center of the vowel space than /a/ or /u/.

Most of the languages in this study do not show a significant effect of stress on vowel peripheralization. Only five languages have significant effects of stress in regression model E as shown in Figure 4.22, and the estimate values are very small, from about 0.005 to 0.03 from the intercept. Given that average Euclidean distance values tend to be around 0.5 (as in Fig. 4.21), stressed vowels in these languages are around $1-6\%$ more or less peripheral

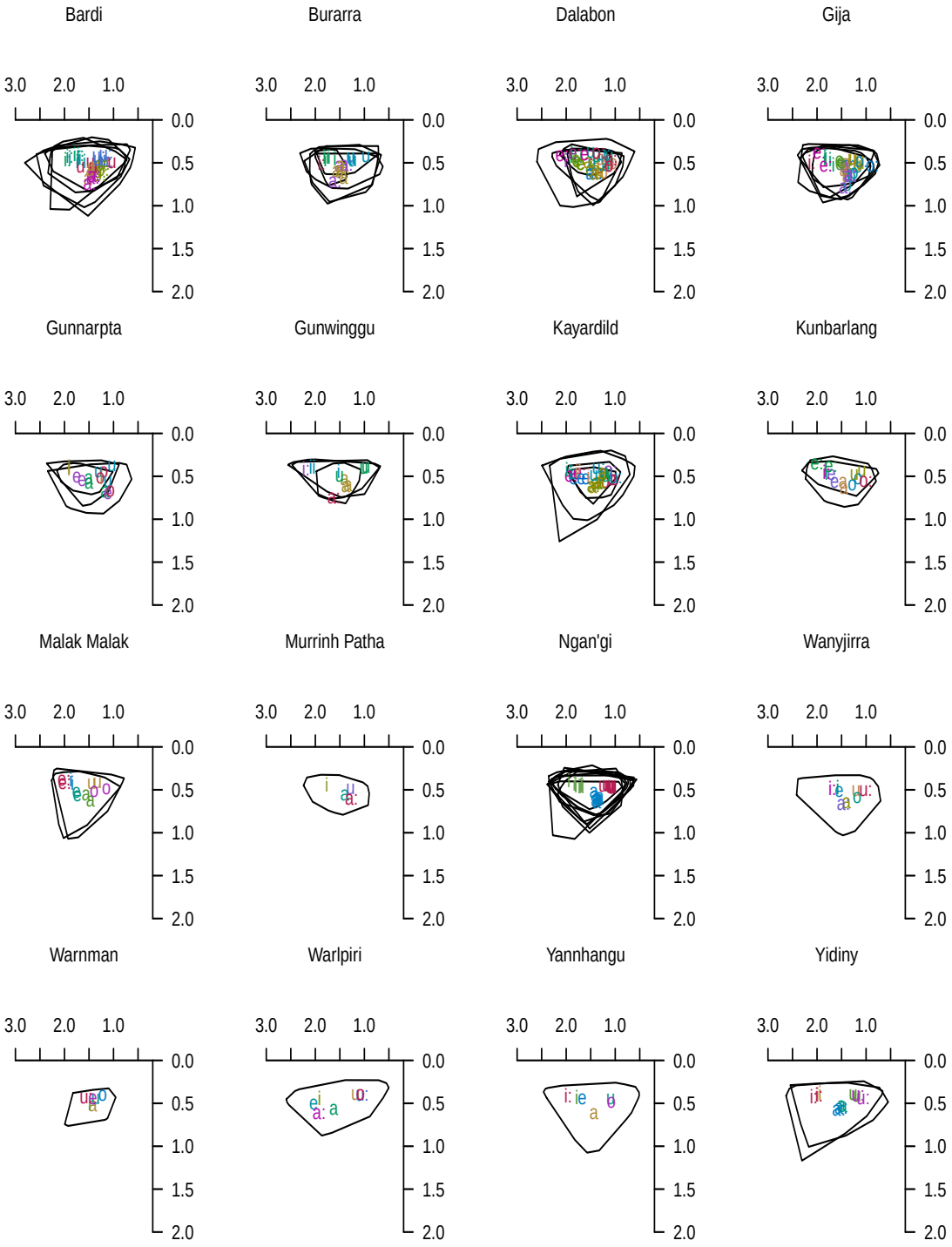

Figure 4.20: Normalized vowel spaces for each language, with polygon for each speaker. The y-axis represents normalized F1 ($F1/\Delta F$), and x-axis represents normalized F2 ($F2/\Delta F$).

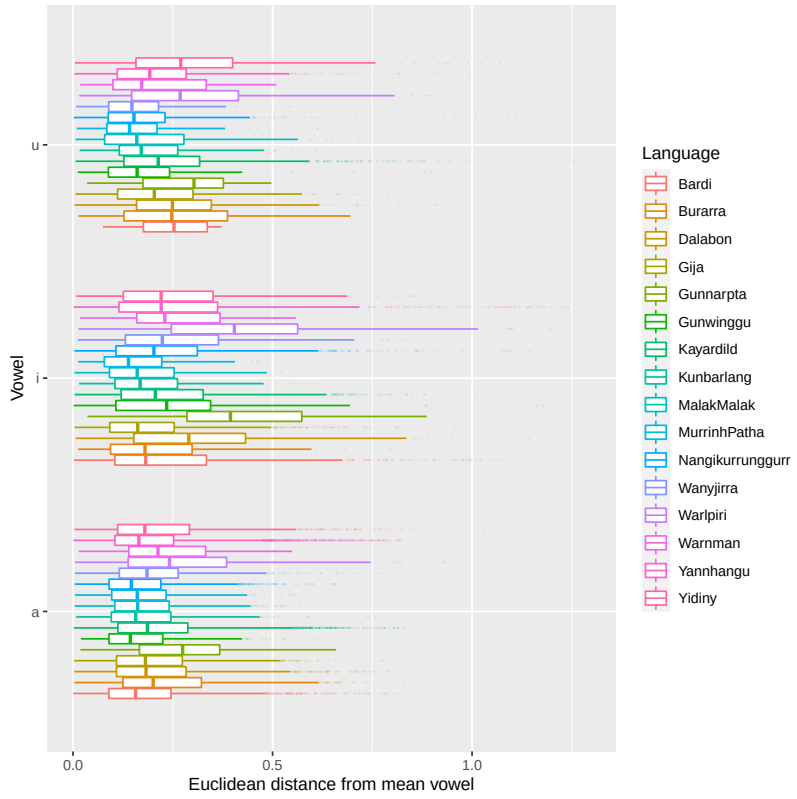

Figure 4.21: Boxplot showing Euclidean distance from the center of the vowel space for /a/, /i/, and /u/ vowel tokens in each language. Units are in the ΔF normalization scale.

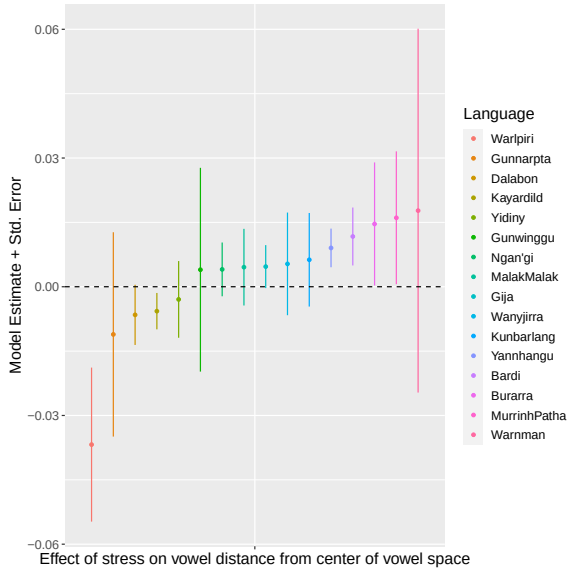

Figure 4.22: Results of regression model E; model estimate and standard error for binary factor 'stress' shown. Unit of measure is Euclidean distance in ΔF normalization scale.

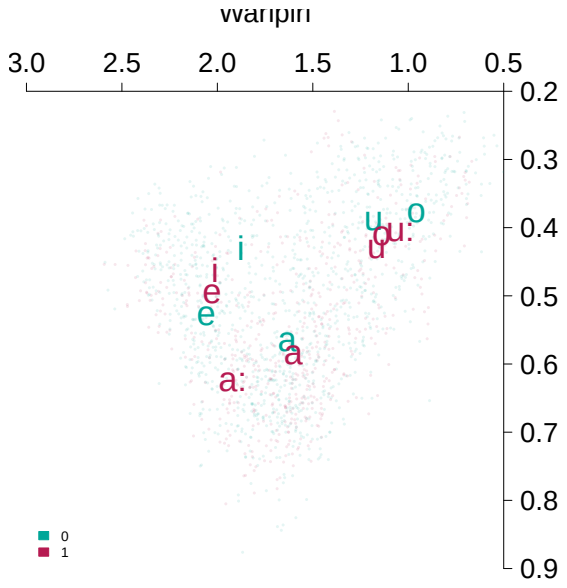

Figure 4.23: Warlpiri vowel space, with stressed and unstressed vowel phonemes separate.

than their unstressed counterparts.

The languages with any significant effect of stress on vowel peripheralization are Warlpiri, Kayardild, Gija, Yan-nhangu, and Bardi. Warlpiri has the largest estimate value at about -0.03, with the negative value indicating that stressed vowels should be more centralized than unstressed vowels. However, this effect is very small, with stressed vowels estimated to be about 6% more centralized than unstressed ones. As shown in Figure 4.23, the effect found by the model seems to reflect the slight centralization of the phonemes /e/, /u/, and /o/, but /i/ seems slightly more peripheral and /a/ is very similar regardless of stress status.

As the other four languages with significant effects have much smaller effect estimates, it stands to reason that these effects will be similarly unclear. While further research may find more robust evidence for these vowel space effects as a correlate of stress, strong conclusions cannot be drawn from the results here.

Language	V. Dur.	Onset Dur.	Post-T. Dur.	Inten.	F0 Max.	F0 Rng.	Vowel
Bardi	+	+		+	+	+	+
Burarra	+						
Gunnartpa	+	+++					
Gija	+	+++			+	+	+
Dalabon		+					
Gunwinggu			+	+	+		
Kunbarlang		+				+	
Kayardild	+		+++	+	+	+	+
Malak Malak	+	++		+	+		
Murrinh Patha			+				
Ngan'gi	+	++		+			
Wanyjirra		+++			+		
Warlpiri	+	+	++		+		+
Warnman	+						
Yan-nhangu	+	+++			+		+
Yidiny	+	+	+	+	+		

Table 4.1: Summary of results for overall language models; + indicates some statistically significant effect that may be attributable to stress. Languages grouped by historical affiliation.

4.6 Summary

Table 4.1 gives a summary of the stress correlate results presented in this chapter for each language. The cells indicate the presence of a significant effect that is likely to indicate a correlation of stress with each acoustic parameter. These effects are only the overall language effects as discussed in this chapter, but in Chapter 5 these effects will be further teased apart based on whether all speakers of the language show the effect, or only some of the speakers.

There is clearly some degree of variation in the correlates of lexical stress across these languages. However, the results summarized in Table 4.1 do not provide a way to measure the degree of difference between languages quantitatively. In order to do this, we need a method of measuring variation at different levels: across languages, within languages across speakers, and within speakers. This method is outlined in Chapter 6.

None of the historically related groups in Table 4.1 have exactly the same correlates found in this investigation. There is, however, some overlap. The Burarra and Gunnartpa data sets share vowel duration as a correlate but not onset consonant duration. As will be discussed in the following chapter, onset consonant duration shows considerable inter-speaker variation, which explains this difference. The Gunwinyguan languages (Dalabon, Gunwinggu, and Kunbarlang) do not share much. Dalabon and Kunbarlang both have effects of onset consonant duration for one consonant group, but for Dalabon it is the glides while in Kunbarlang it is the stop consonants. Gunwinggu has an effect of post-tonic nasal lengthening, as well as f0 maximum, which the other languages in this group do not share. Kunbarlang has an effect of f0 range, but not f0 maximum. Murrinh Patha and Ngan'gi also do not share any correlations with each other, which is not surprising given the idiosyncracies of stress in Murrinh Patha.

The Pama-Nyungan languages have considerable overlap in their stress correlates. All but Wanyjirra has vowel duration as a correlate of stress, and all but Warnman has f0 maximum. All of the Pama-Nyungan languages except Warnman have some effect of onset consonant lengthening, although only Wanyjirra and Yan-nhangu have this effect in all three consonant categories. These two languages are also the only languages to have effects of vowel space on stress. Warlpiri only has an onset lengthening effect for stop consonants, while Yidiny only has an effect for glides. Warlpiri and Yidiny also share an effect of post-tonic lengthening of stop consonants, while Warlpiri additionally has an effect of post-tonic glides.

Ten of the 16 languages in this survey have a correlation of f0 maximum and stress. This provides some supporting evidence for the claim that f0 is an extremely common correlate of stress in Australian languages (Fletcher & Butcher 2014). However, both vowel duration and onset consonant duration are similarly common in this set of languages. Vowel duration correlates with stress in 11 of sixteen languages, while onset consonant duration for at least

one consonant category correlates with stress in eleven languages as well. However, this set of languages is a sample of convenience and not a balanced selection of Australian languages from across the continent. These observations provide interesting evidence for the prevalence of these stress correlates, but further typological work would be needed to draw definitive conclusions.

The results presented in this chapter binned together all speakers of each language in order to look at overall language results and to compare across languages. However, considering the speech of each speaker separately reveals substantial variation in some cases; these results are discussed in the following chapter (Chapter 5).

Chapter 5

Results: Within-language variation

The previous chapter addressed Claim (9.1) of this dissertation by establishing the correlates of stress in each language and discussing the variation across them. Here, within-language cross-speaker variation is examined for the languages in this dissertation, beginning to address Claim (9.2), that the cues to stress will vary across speakers of the same language. A full and detailed investigation of this claim would require both the acoustic analyses presented here as well as separate studies of the sources of sociolinguistic variation in these cues by spending time in each of the individual communities in which these languages are spoken. This type of work was not possible within the scope of the current thesis, and in some cases is not possible at all because some of these languages are no longer spoken. Because of this, only the acoustic studies are presented in this chapter, and any sociolinguistic study of variation in stress cues is left for future research.

In what follows, each of these languages is discussed in turn, except when the data I have for a language only includes speech from one individual; these languages are not included here. For information about all languages in this dissertation and their speakers, see Chapter 2.

As will become clear as each language is discussed in turn, the nature of the data used

in this dissertation project means that not all speakers are equally represented for a given language's data set. In fact, in most cases one speaker has a clear plurality or majority of the data points (measured as number of vowel tokens), while the other speakers have comparatively very little representation. Thus, it is especially important to consider by-speaker variation in light of this sort of skewed data.

5.1 Bardi

There are five speakers in the Bardi data used in this project. Almost half of the data was spoken by speaker DW, while the remaining four each represent less than 20% of the data. This is a common skew of speakers and data, as will be seen in subsequent sections.

Bardi has quite a large effect of stress on vowel duration, as shown on the far right in Figure 5.1. This overall effect overlooks some variation among Bardi speakers, although all five have significant, positive effects of stress on vowel duration. Speaker 3/JS shows an effect almost twice that of the average, while speaker 5/TE has an effect about half the size. As might be expected, the speaker with the highest share of the data (2/DW) has close to the average effect. Since 2/DW accounts for almost half of the data in the full language model, their effects will drive the overall effects more than the other speakers' will.

Bardi was found to have a small but significant effect of stress on vowel intensity. When looking at speakers individually in Figure 5.2, two speakers are clearly driving this overall effect while the other three do not have a significant effect of stress on intensity. Speaker 2/DW is one of the speakers with a significant effect, which explains why the overall effect is significant because of this speaker's share of the data.

Speaker	1/BE	2/DW	3/JS	4/NI	5/TE
Share of data	17.0%	45.4%	14.1%	9.5%	14.0%
Vowel tokens	1527	4060	1262	851	1249

Table 5.1: Share of total data for each Bardi speaker.

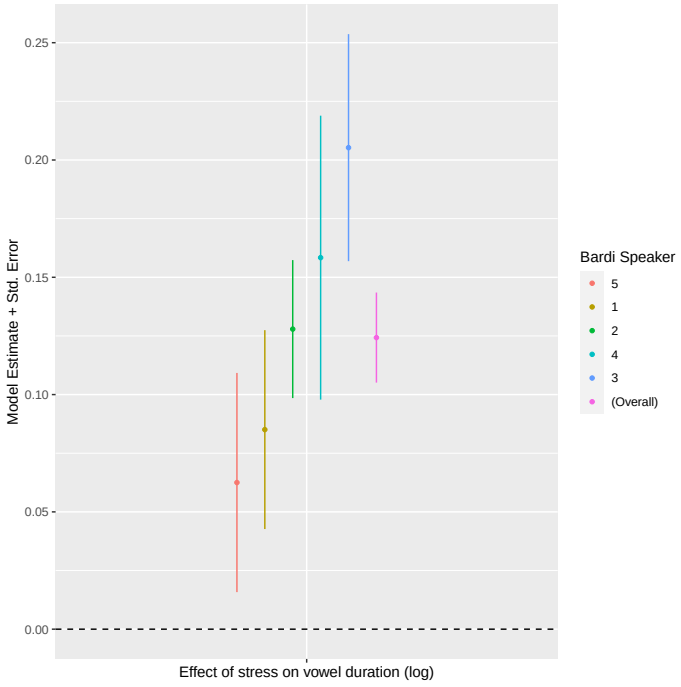

Figure 5.1: Model effect of the fixed binary factor 'stress' on duration of vowels in Bardi. In legend, topmost labels correspond to leftmost dot-whiskers. Lines that cross the zero mark (dark dashed line) represent non-significant model results.

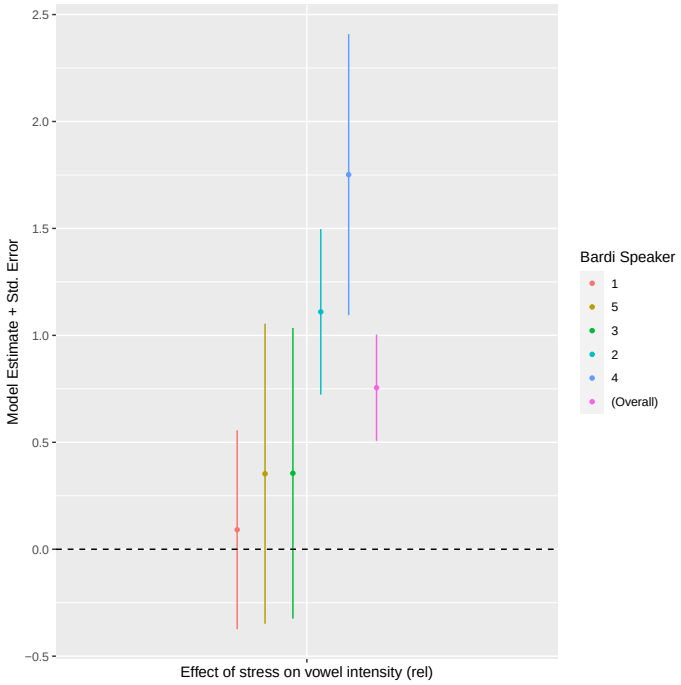

Figure 5.2: Model effect of the fixed binary factor ‘stress’ on intensity of vowels in Bardi. In legend, topmost labels correspond to leftmost dot-whiskers. Lines that cross the zero mark (dark dashed line) represent non-significant model results.

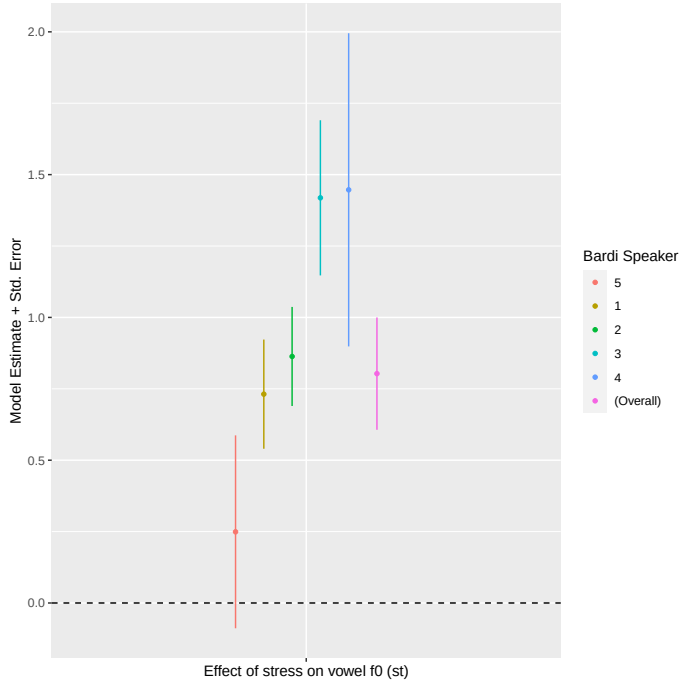

Figure 5.3: Model effect of the fixed binary factor 'stress' on f0 of vowels in Bardi.

The by-speaker effects of stress on f0 are shown in Figure 5.3. Four of five speakers show significant and positive effects in these models, as the overall effect reflects as well. Two speakers, 3/JS and 4/NI, have effect sizes almost twice as large as the overall effect, which again is dominated by data from speaker 2/DW. Speaker 5/TE does not have a significant effect of stress on f0 in this model. It is unclear what the cause of this variation is; without a large population of Bardi speakers, this sort of variation cannot easily be identified as either a substantial speech variant employed by certain groups of speakers, or an anomaly only employed by one or a few. Likewise, it is difficult to know whether this is the sort of variation that is the result of some phonological reanalysis of stress in the mind of the speaker, or the result of language attrition.

Speaker	1/EB	2/HL	3/MB	4/RJ	5/TN	6/01	7/CE	8/JBB
Share of data	47.8%	1.9%	16.3%	7.5%	3.0%	6.1%	4.3%	13.1%

Table 5.2: Share of total data for each Burarra speaker.

5.2 Burarra/Gunnartpa

The Burarra and Gunnartpa datasets (from PARADISEC and ELAR, respectively) are combined here as they include audio from the same language, Burarra, of which Gunnartpa is a dialect. There are eight speakers total in the Burarra data; their shares of the data are given in Table 5.2. Almost half of the data comes from EB, while the remaining speakers account for less than 15% of the data each. Similarly to Bardi, we will expect that speaker 1/EB will largely drive the overall effects in the general language model, and variation will be seen in the other speakers.

Most of the potential correlates to stress investigated in Chapter 4 were not found to be significant in Burarra/Gunnartpa, whether looked at as separate collections or combined as here. As shown in Figure 5.4, most speakers of Burarra/Gunnartpa do not show a significant effect of stress on vowel duration, as the overall effect shows. One speaker (8/JBB) has a significant and positive effect here.

Unlike Bardi, which has very few fluent speakers, the by-speaker variation seen in Burarra/Gunnartpa may be grounds for follow-up study, as the language community is quite robust. Perhaps a variant seen in one speaker in this project is emblematic of a larger sociolinguistic trend that is beyond the scope of a typological study such as this one.

Results are similar for the intensity models. Figure 5.5 shows that most speakers have no significant effect here, but speaker 2/HL has a significant and positive effect. Again, the source(s) of this variation are well suited for future study.

Burarra/Gunnartpa does have an overall effect of stress on f0, in contrast to the factors already considered here. However, as can be seen in Figure 5.6, this overall effect is driven

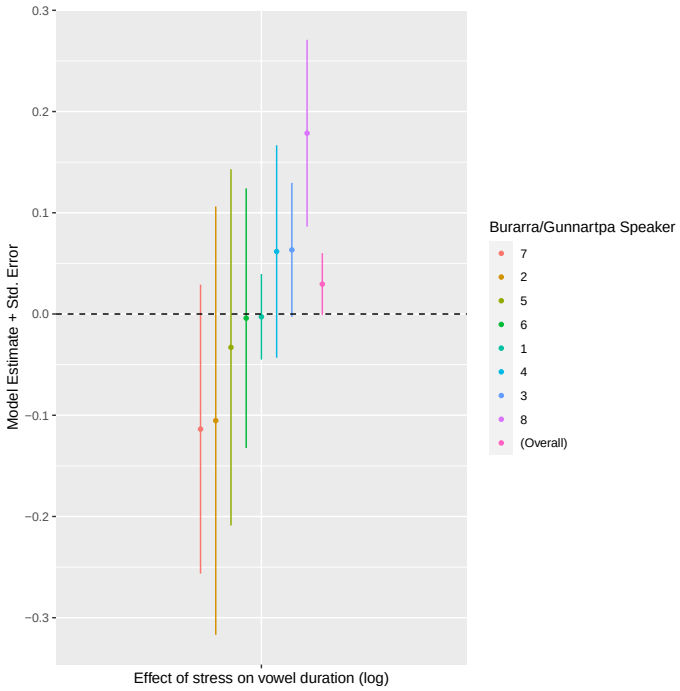

Figure 5.4: Model effect of the fixed binary factor 'stress' on duration of vowels in Burarra/Gunnartpa.

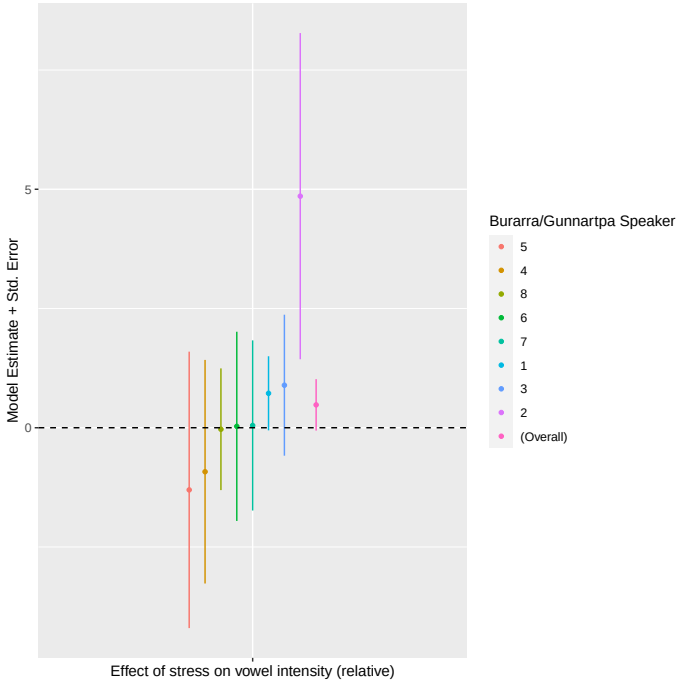

Figure 5.5: Model effect of the fixed binary factor 'stress' on intensity of vowels in Burarra/Gunnartpa.

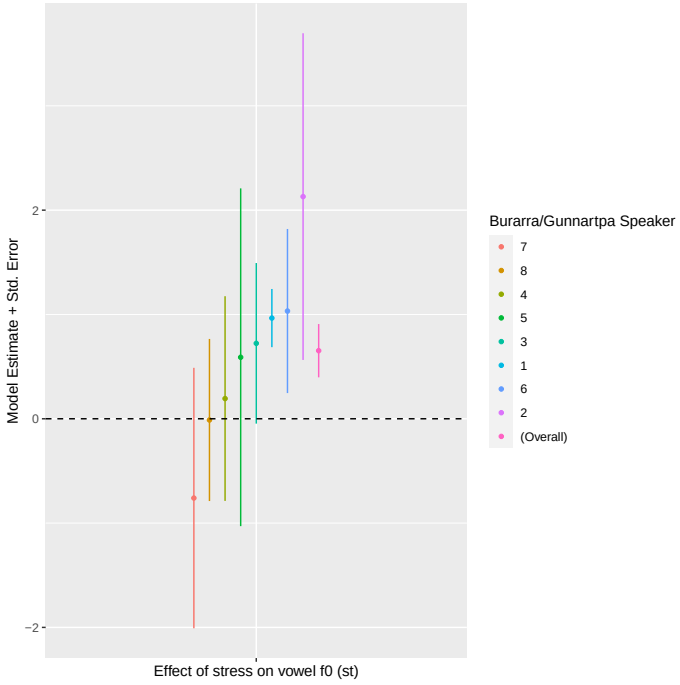

Figure 5.6: Model effect of the fixed binary factor 'stress' on f0 of vowels in Burarra/Gunnartpa.

primarily by the speaker who accounts for the majority of the data, 1/EB. Two other speakers (2/HL and 6/01) share this significant and positive effect, but the other five speakers do not have significant effects here. In this case especially, where there are two multi-speaker groups with differing effects, we want to consider whether sociolinguistic or related factors are at play here.

In the previous chapter (Ch. 4), where the ELAR and PARADISEC archival deposits were considered separately, there were effects of onset consonant duration found for Gunnartpa across all three consonant categories, but in Burarra this was not found for any consonants. These are broken up by speaker for the joined dataset for stops in Fig. 5.7; nasals in Fig. 5.8; and glides in Fig. 5.9.

The inter-speaker results for onset stop consonants only show a significant correlation with stress for one speaker, 8/JBB. The combined data set does not have an overall effect

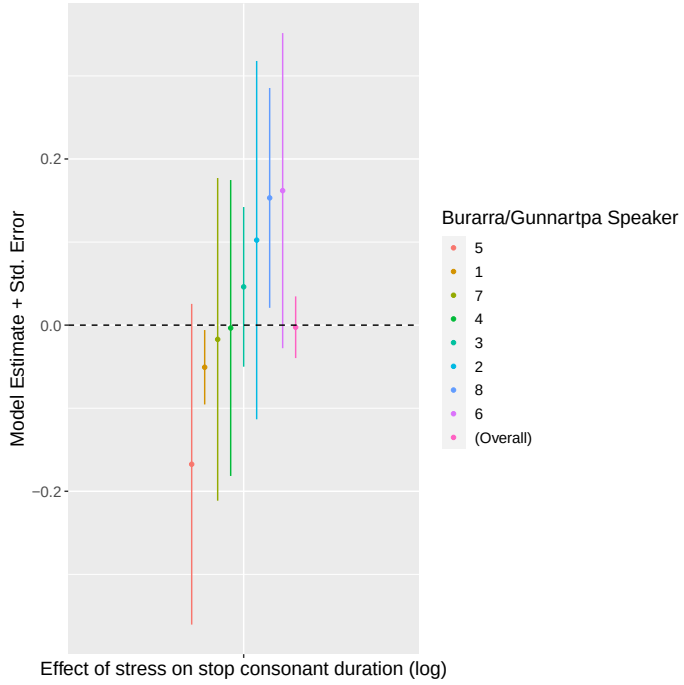

Figure 5.7: Model effect of the fixed binary factor 'stress' on onset stop duration in Burarra/Gunnartpa.

of stress on stop duration. Speaker 8/JBB, while a smaller proportion of the combined data here, accounts for a large proportion of the ELAR Gunnartpa deposit, which likely explains the effect that was seen for Gunnartpa alone. The same situation is observed for nasal onsets in Figure 5.8. Only speaker 8/JBB has a significant effect of stress here, and this is certainly the source of the ELAR Gunnartpa results where the overall effect is not significant.

The effect of onset glide duration does hold for the combined data set. This is carried by significant effects for three speakers: 6/01, 7/CE, and 8/JBB. These are the only three speakers included in the ELAR Gunnartpa data set, while the five speakers without a significant effect are the ones in the PARADISEC Burarra deposit. So, with all the data considered together, Burarra/Gunnartpa has an effect of onset glide lengthening, but not lengthening of other onset consonants. The glide lengthening is a point of interspeaker variation.

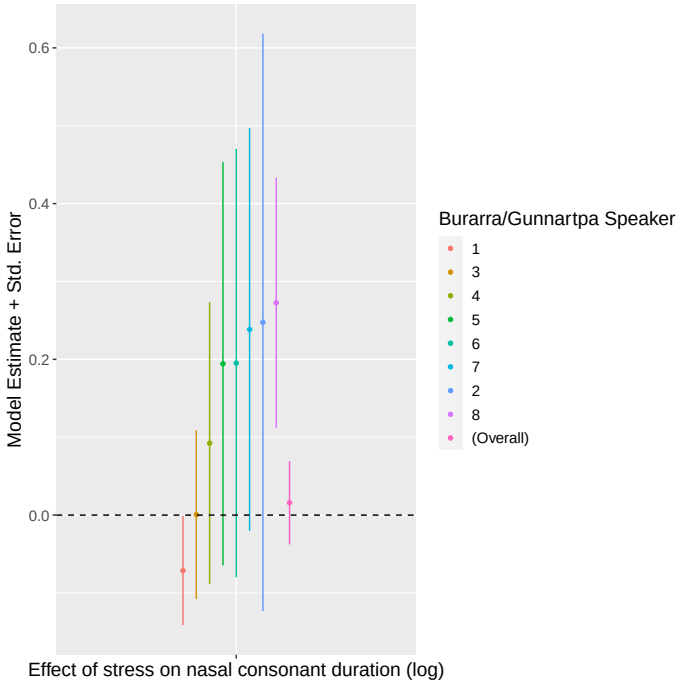

Figure 5.8: Model effect of the fixed binary factor 'stress' on onset nasal duration in Burarra/Gunnartpa.

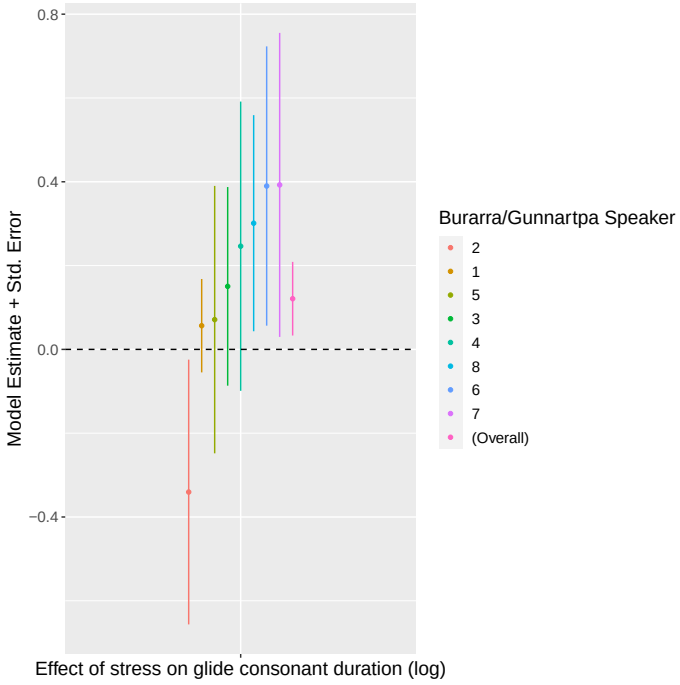

Figure 5.9: Model effect of the fixed binary factor 'stress' on onset glide duration in Burarra/Gunnartpa.

5.3 Dalabon

While the Dalabon data contains audio from five speakers, the share of data held by each speaker is heavily skewed. Table 5.3 gives these shares; speaker MT/3 is the speaker in 79% of the Dalabon audio, and the other four speakers split the remaining 21%. The smallest share is held by speaker JJA/1, who only has 86 vowel tokens in the data used here.

As would be expected given the skewness of the data, the model estimates for Dalabon overall and for speaker MT/3 in Figures 5.10, 5.11, and 5.12 are extremely similar. While individual speakers sometimes vary from this overall measure, such as speaker LB/2 in Figure 5.11, speaker MT/3 is always close to the overall measure because of their predominance in the overall data.

Speaker JJA/1 consistently has very large standard error values. This is because, as shown in Table 5.3, only 86 vowel tokens are present in these data for speaker JJA/1. This is clearly not enough data to draw meaningful conclusions from the regression model, resulting in the standard error values as well as the lack of any significant effects in the models shown.

Keeping these trends in mind, speakers are relatively cohesive in their non-significant results for vowel duration in Figure 5.10 and f0 in Figure 5.12. A split is seen for vowel intensity in Figure 5.11, where three speakers have significant and negative effects as the overall result suggests, while two speakers have no significant effect. However, as discussed in §4.3, a negative effect of intensity is not taken to be a correlation with stress, but likely a word-positional effect instead.

Speaker	1/JJA	2/LB	3/MT	4/ND	5/PA
Share of data	1.2%	6.2%	79.0%	10.7%	2.9%
Vowel tokens	86	432	5459	737	199

Table 5.3: Share of total data for each Dalabon speaker.

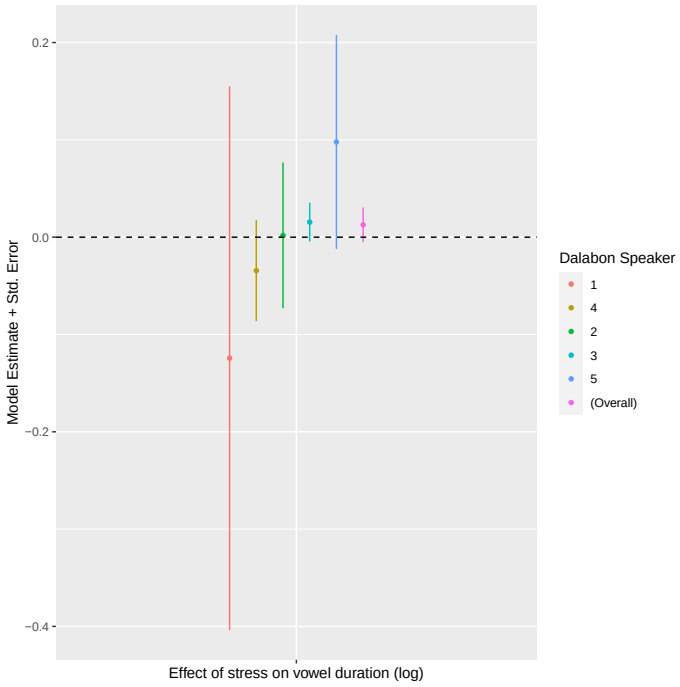

Figure 5.10: Model effect of the fixed binary factor ‘stress’ on duration of vowels in Dalabon.

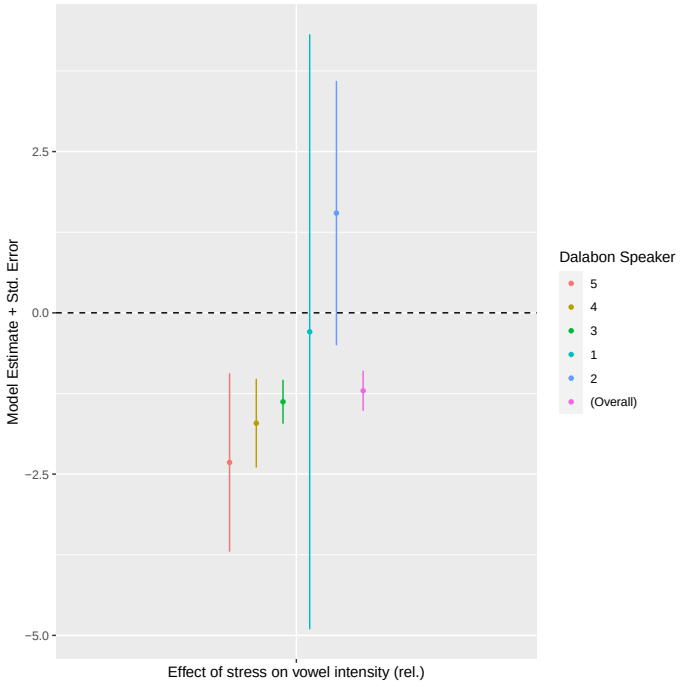

Figure 5.11: Model effect of the fixed binary factor ‘stress’ on intensity of vowels in Dalabon.

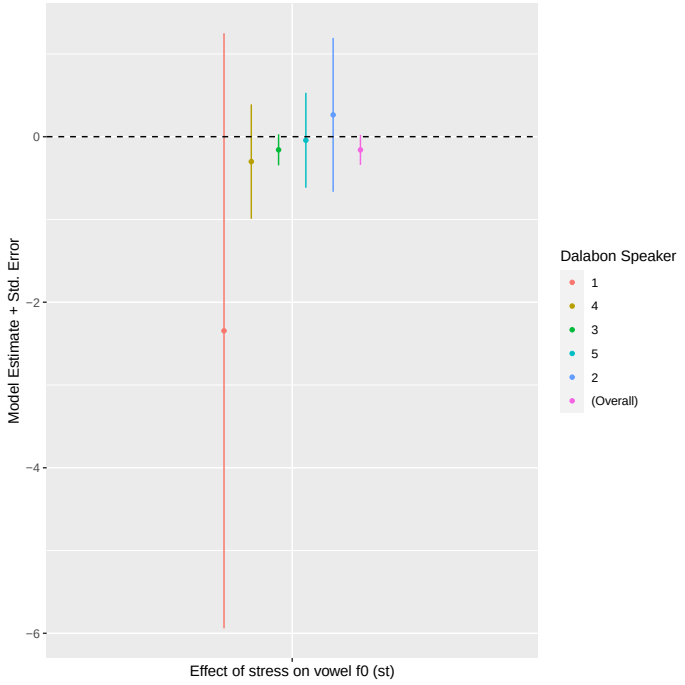

Figure 5.12: Model effect of the fixed binary factor 'stress' on f0 of vowels in Dalabon.

5.4 Gija

There are five speakers included in the Gija data used in this dissertation. The vast majority of the data comes from speaker 2/MT, as shown in Table 5.4, with the other four speakers accounting for 10% or less of the data each. As expected given this skew, the overall model results largely reflect speaker 2/MT's speech, and are less influenced by the speech of the others.

Only two speakers show significant, positive effects of stress on vowel duration, as shown in Figure 5.13. These are speaker 2/MT and speaker 3/PP, the latter of which has

Speaker	1/MJ	2/MT	3/PP	4/RP	5/Y
Share of data	8.2%	68.0%	10.8%	6.8%	6.2%
Vowel tokens	902	7491	1186	747	683

Table 5.4: Share of total data for each Gija speaker.

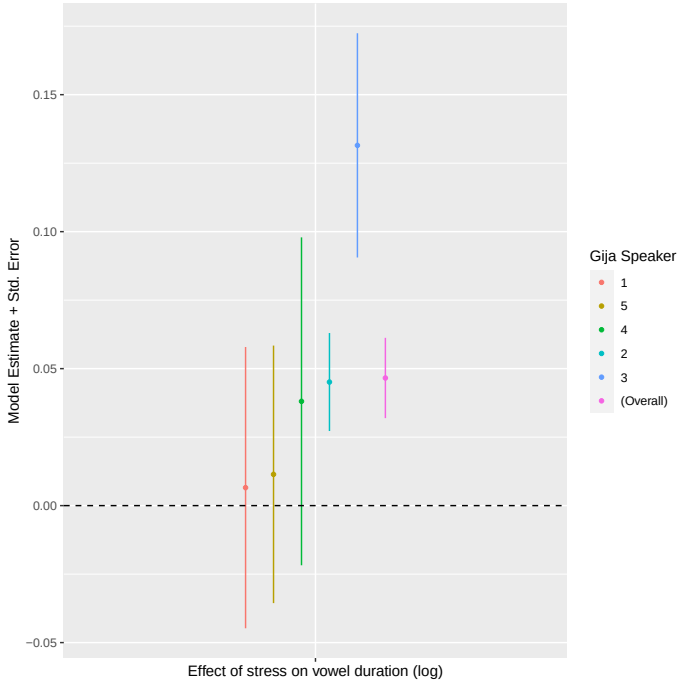

Figure 5.13: Model effect of the fixed binary factor 'stress' on duration of vowels in Gija.

quite a large effect here. The other three speakers do not have a significant effect in their duration models.

The overall effect of stress on intensity in Gija, as most languages in this study, were not found to correlate; instead, some other factor such as word position is likely at play here. Each speaker varies in their intensity models. Three speakers have a significant and negative effect, as the overall result, while the other two do not have significant effects.

F0 is a significant correlate of stress in Gija. The overall effect is that stressed syllables are almost 1 semitone higher than unstressed ones, and all speakers are clusters closely around this point. Speakers 1, 2, and 4 have effects closer to 0.5 semitones, while speakers 5 and 3 have effects just above 1.0 semitone.

The results of these by-speaker investigations reveal that in Gija, f0 is likely the main correlate to lexical stress, while some speakers may additionally use duration to cue prominence. This point is discussed more broadly at the end of this chapter.

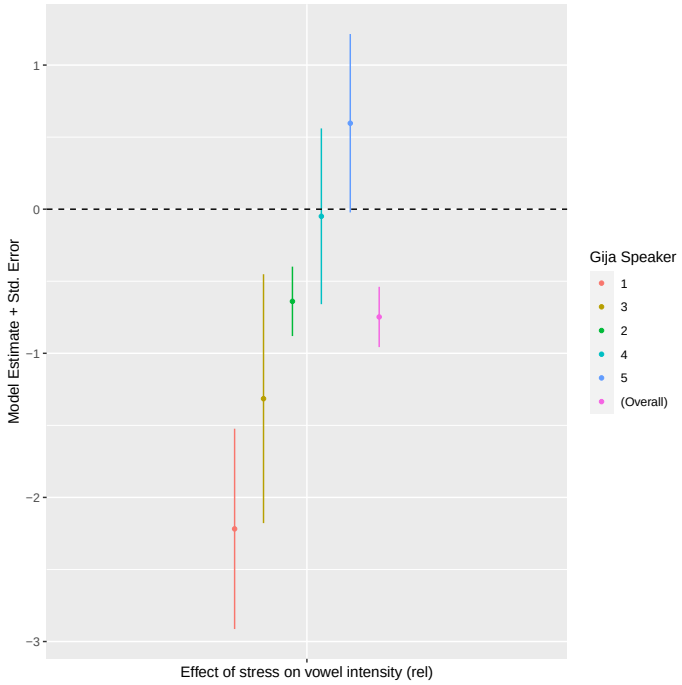

Figure 5.14: Model effect of the fixed binary factor ‘stress’ on intensity of vowels in Gija.

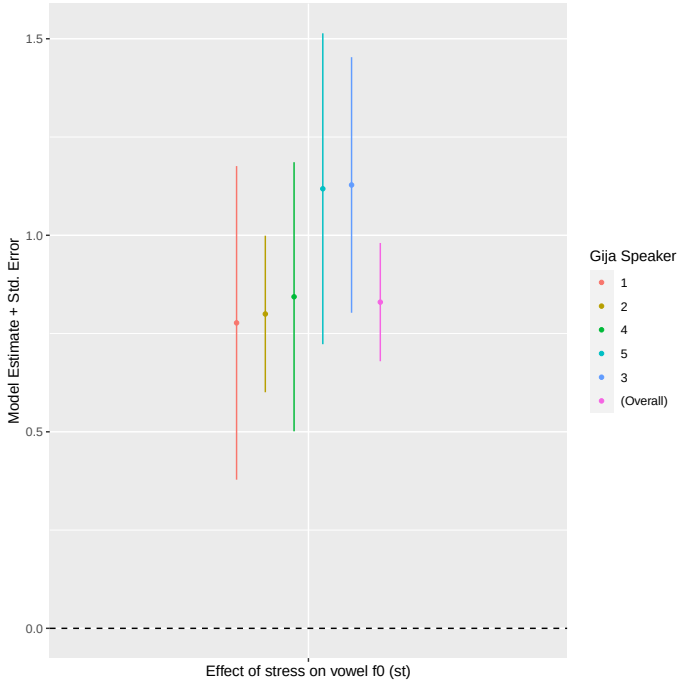

Figure 5.15: Model effect of the fixed binary factor ‘stress’ on f0 of vowels in Gija.

5.5 Gunwinggu

There are three Gunwinggu speakers included in the data used for this project. The data is similarly skewed as in other languages. However, it is important to note that speaker 3/CB, despite representing 11.5% of the data here, only has 79 vowel tokens because the overall Gunwinggu corpus is smaller than many of the other languages. This speaker has very large standard error values in the following model results because of this.

Duration is not an overall correlate with stress in Gunwinggu. This is borne out across speakers as well, though speaker 3/CB trends more positively than the others. This is not meaningful because of the small amount of data for this speaker.

The slight positive effect seen in the intensity model for Gunwinggu overall is heavily influenced by speaker 1/01, who holds 60.5% of the data. The other two speakers in the corpus do not have a significant effect here.

Finally, f0 results for Gunwinggu overall give a significant and positive correlation with stress with an effect close to one semitone. Speakers 1/01 and 2/A both show very similar effect sizes here. Speaker 3/CB does not, but again because of the small amount of data spoken by this speaker no clear conclusions can be drawn. Overall, it seems that f0 might be a consistent correlate with stress across speakers of Gunwinggu.

Speaker	1/01	2/A	3/CB
Share of data	60.5%	28.0%	11.5%
Vowel tokens	417	193	79

Table 5.5: Share of total data for each Gunwinggu speaker.

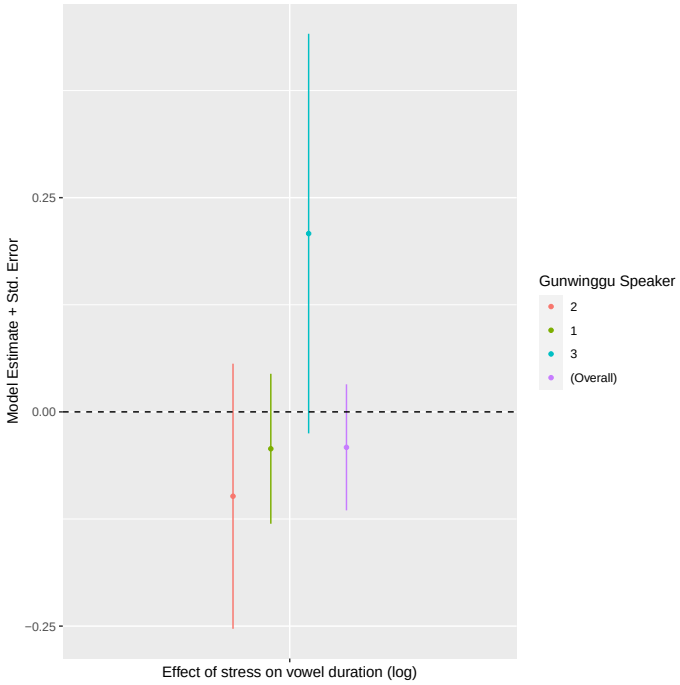

Figure 5.16: Model effect of the fixed binary factor ‘stress’ on duration of vowels in Gunwinggu.

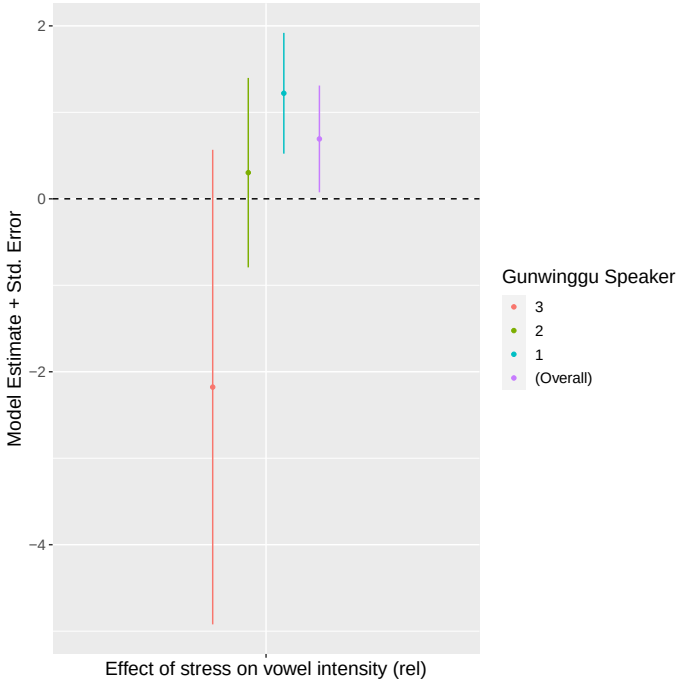

Figure 5.17: Model effect of the fixed binary factor ‘stress’ on intensity of vowels in Gunwinggu.

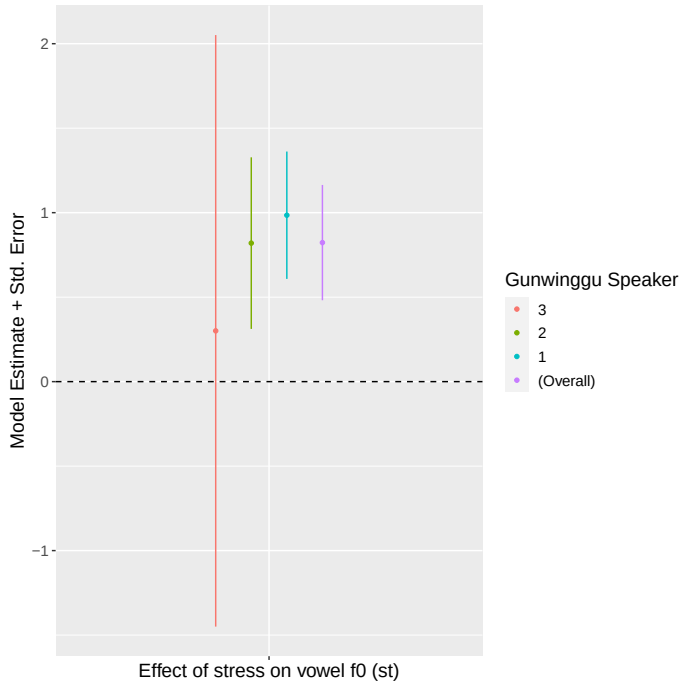

Figure 5.18: Model effect of the fixed binary factor 'stress' on f0 of vowels in Gunwinggu.

5.6 Kayardild

The Kayardild data includes four speakers, but their proportions of the data is massively skewed, as can be seen in Table 5.6. Speaker 2/DN represents 91.6% of the Kayardild data. This has a clear impact on the by-speaker results reported below. In fact, speaker 4/MM is excluded from the following graphs because there are only 18 vowel tokens for this speaker and results cannot be reliable with so little data.

By-speaker duration results are shown in Figure 5.19. Overall, Kayardild has a significant and positive effect of stress on duration, and this is borne out for speakers 1 and 2

Speaker	1/AL	2/DN	3/ET	4/MM
Share of data	7.4%	91.6%	0.8%	0.1%
Vowel tokens	1276	15786	145	18

Table 5.6: Share of total data for each Kayardild speaker.

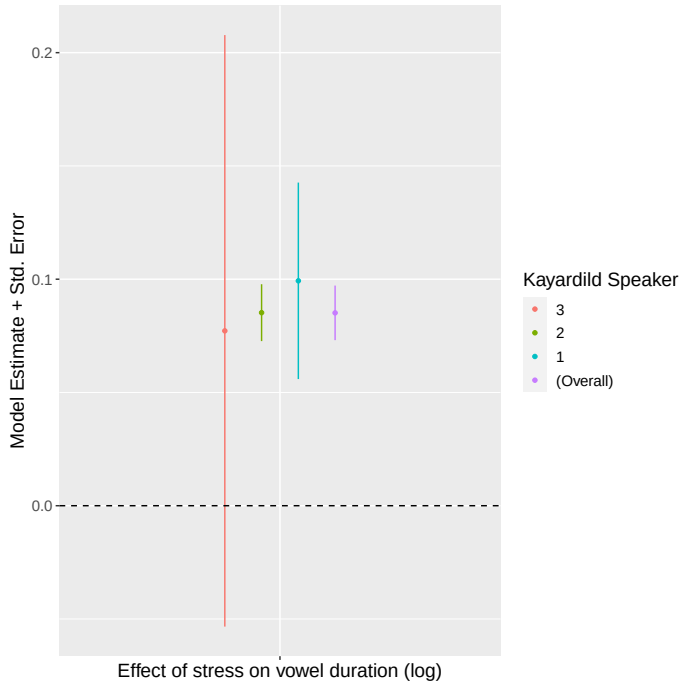

Figure 5.19: Model effect of the fixed binary factor 'stress' on duration of vowels in Kayardild.

here. Speaker 3's estimate value is similar to these other speakers', but with a large standard error and lack of significance as a result. Of course, speaker 3 only has 145 vowel tokens, so these results cannot be interpreted strongly.

Kayardild also has an overall effect of intensity. However, this effect seems to be heavily influenced by speaker 2, and the other two speakers do not have significant results here. As already discussed for intensity, these results are difficult to interpret.

Finally, Kayardild also has an overall effect of stress on f0. This effect is seen for both speakers 2 and 1, suggesting that this would be a consistent effect across more Kayardild speakers. Speaker 3 again has no significant effect, likely influenced by the small set of data available for this speaker's speech.

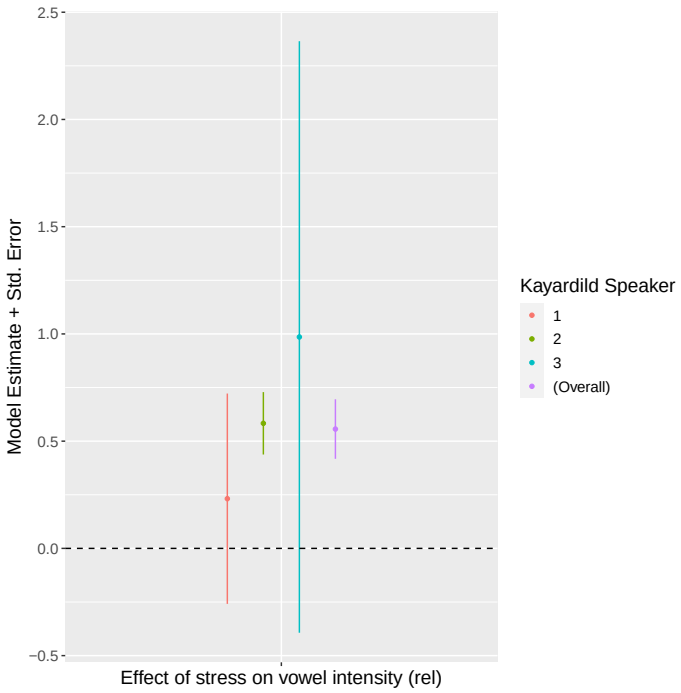

Figure 5.20: Model effect of the fixed binary factor ‘stress’ on intensity of vowels in Kayardild.

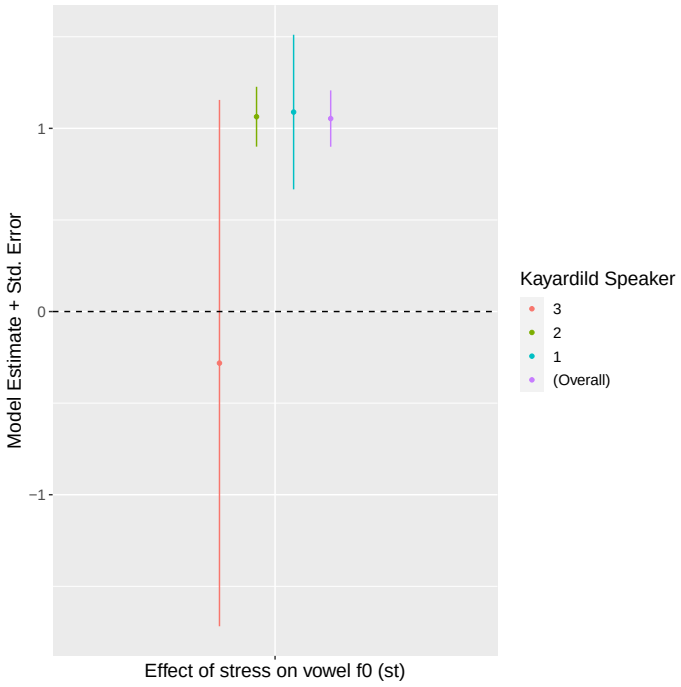

Figure 5.21: Model effect of the fixed binary factor ‘stress’ on f0 of vowels in Kayardild.

5.7 Kunbarlang

There are two Kunbarlang speakers represented in the data used for this dissertation, SM and FA. The vast majority (84%) of the data come from speaker FA, while the remaining 16% come from SM. While most of the other languages with large skews like this also have more speaker variation in the less-represented speakers, both speakers of Kunbarlang have similar estimate values in all cases. However, because speaker FA/1 has so little speech data, their error values are often much larger than speaker SM/2.

In the case of the duration results in Figure 5.22, speaker FA/1 does have a slightly negative estimate value in comparison to speaker SM/2, who has an estimate near zero. However, in both cases and in the overall result, stress is not a significant predictor of vowel duration.

The overall effect of stress on vowel intensity (Figure 5.23) is significant and slightly negative. However, each speaker's individual results is not significant, although the estimate values are similar to the overall result.

Again, estimate values are similar for both Kunbarlang speakers in their effects of stress on f0 in Figure 5.24. However, the larger error values for speaker FA/1 yield an insignificant effect here, while the other speaker and the overall effect are both significant. While it seems likely that more data from this speaker would yield a significant effect with a similar magnitude as that for speaker SM/2, no definite conclusions can be drawn from these results.

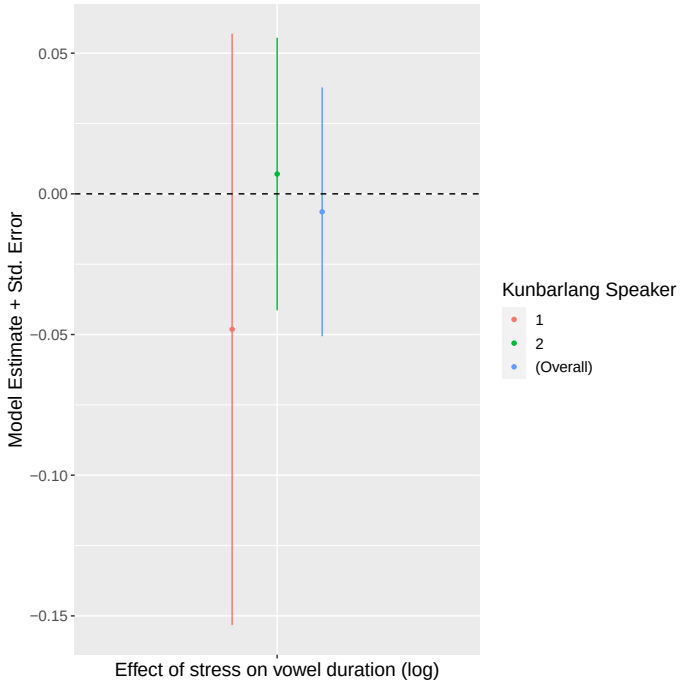

Figure 5.22: Model effect of the fixed binary factor ‘stress’ on duration of vowels in Kunbarlang.

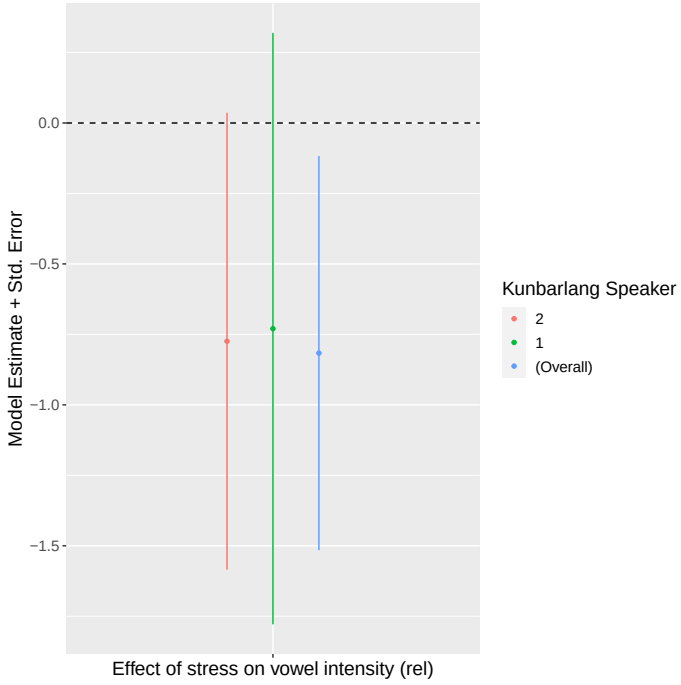

Figure 5.23: Model effect of the fixed binary factor ‘stress’ on intensity of vowels in Kunbarlang.

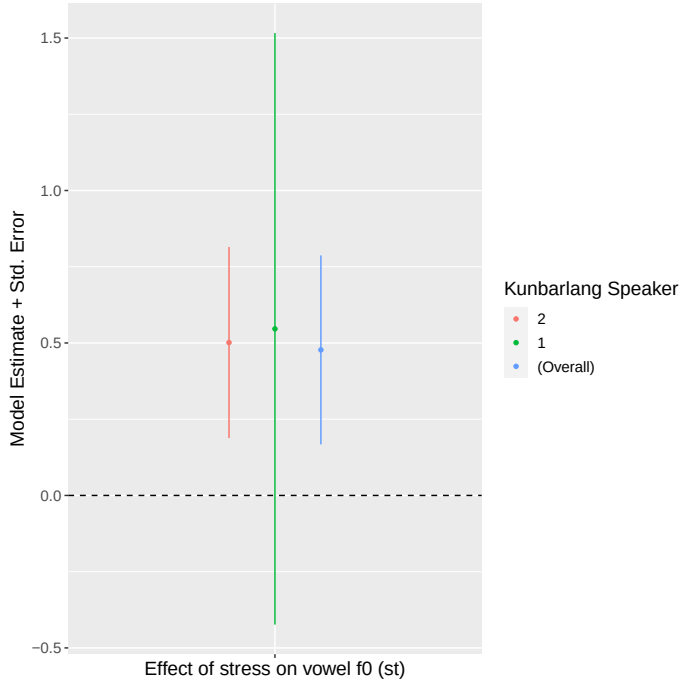

Figure 5.24: Model effect of the fixed binary factor 'stress' on f0 of vowels in Kunbarlang.

5.8 Malak Malak

Malak Malak also has two speakers in the data, K and TW. Speaker K/1 accounts for around 62% of the data in this language, while TW/2 accounts for the remaining 38%, a relatively more equitable split than we have seen in the other languages. Malak Malak shows a significant effect of stress on all of the factors duration, intensity, and f0, and these are also seen in both speakers individually, though in some cases there is more variation than in others.

Figure 5.25 shows the effect of stress on vowel duration for both speakers. Speaker TW/2 has a slightly higher estimate value than speaker K/1, but both of these effects are rather close to one another and the error ranges overlap substantially.

The effects for both speakers of stress on vowel intensity are almost identical. These are shown in Figure 5.26. Both speakers have a slightly positive significant effect here, although see §4.3 for discussion of the interpretation of these results.

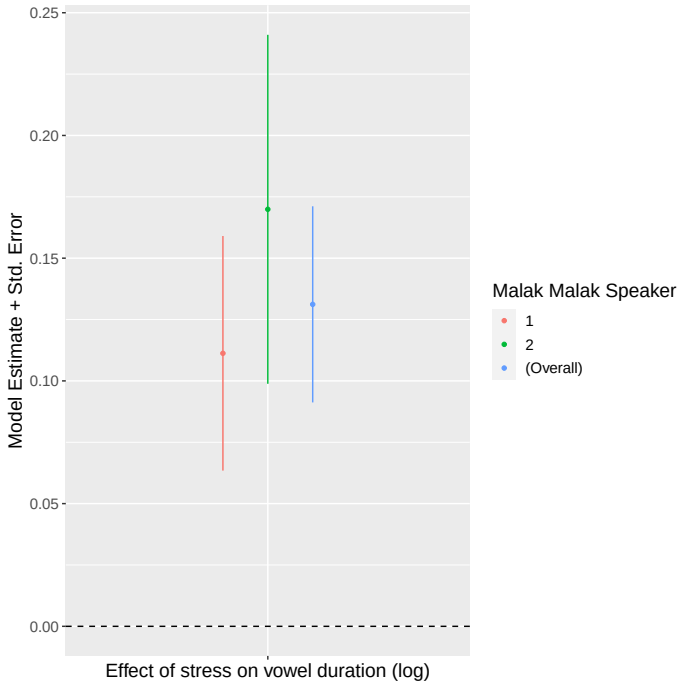

Figure 5.25: Model effect of the fixed binary factor ‘stress’ on duration of vowels in Malak Malak.

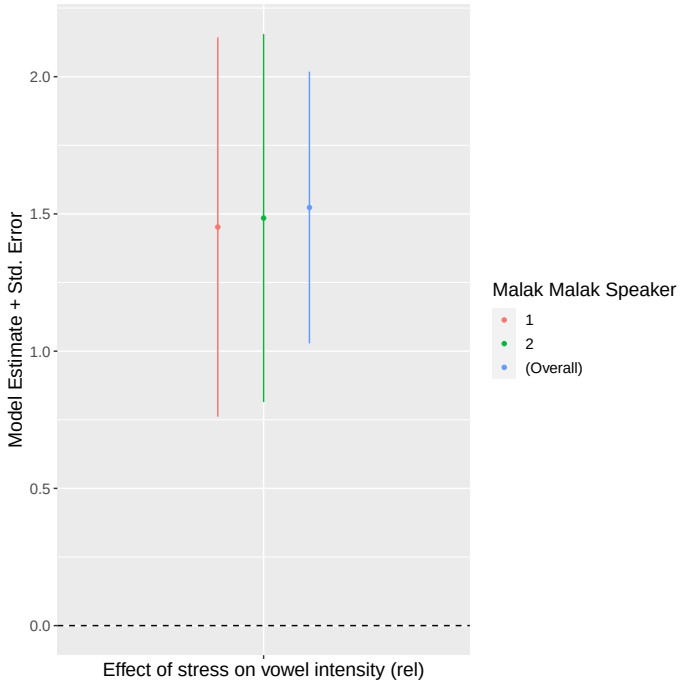

Figure 5.26: Model effect of the fixed binary factor ‘stress’ on intensity of vowels in Malak Malak.

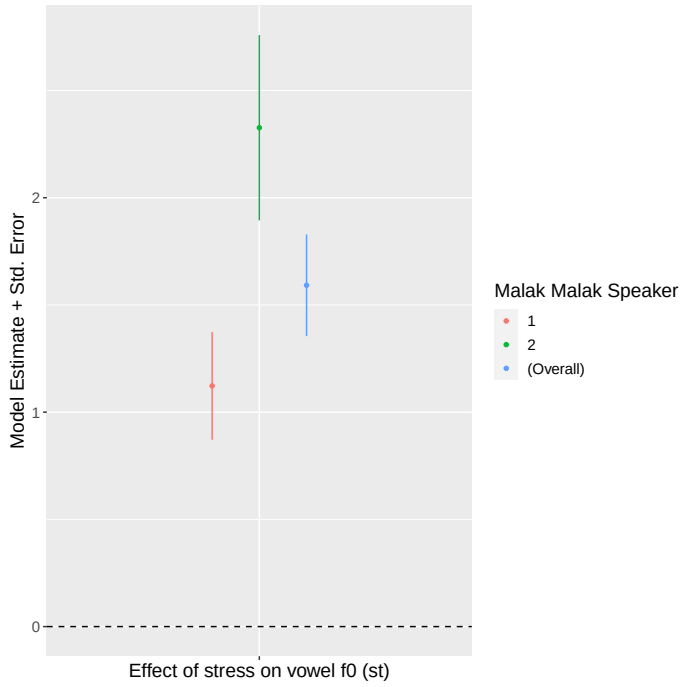

Figure 5.27: Model effect of the fixed binary factor 'stress' on f0 of vowels in Malak Malak.

The two Malak Malak speakers show a bit more variation in their effect sizes of stress on f0 than for the previous phonetic factors (Fig. 5.27). Both speakers have positive and significant effects, but speaker K/1 has an effect of stressed syllables being about 1 semitone higher than unstressed ones, while speaker TW/2 has about twice this effect size, with stressed vowels having over 2 semitones higher f0 than unstressed vowels. Both of these are rather strong effects, but as there are only two speakers here it is not possible to draw broader conclusions as to the nature of this variation. Unfortunately there are very few speakers of Malak Malak remaining, so further study into this question may not be possible.

5.9 Ngan'gi

Nine speakers are represented in the Ngan'gi data. Below is a breakdown of the relative amounts of data for each speaker. The speakers' shares of the data are relatively more equitable than has been seen in many of the other languages in this dissertation. Speakers

Speaker	1	2	3	4	5	6	7	8	9
Share of data	8.7%	17.8%	22.8%	1.9%	6.2%	21.0%	9.4%	6.6%	5.6%

Table 5.7: Share of total data for each Ngan'gi speaker.

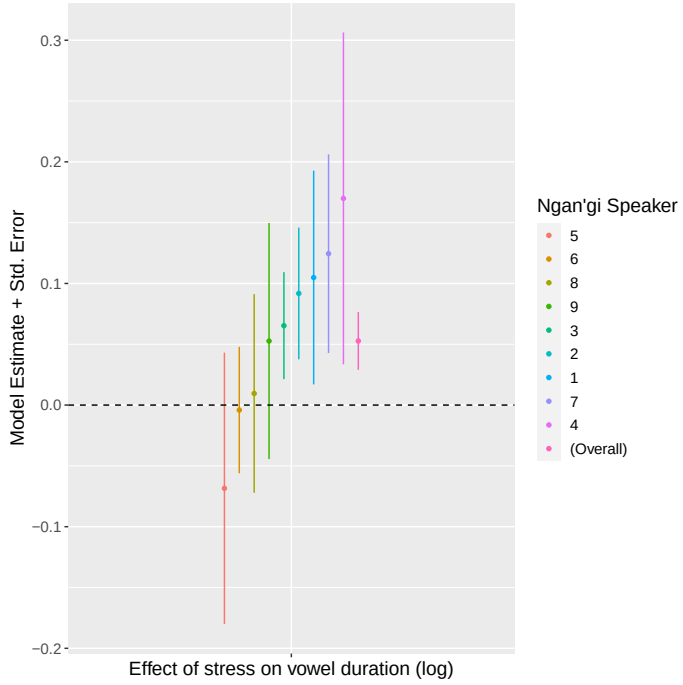

Figure 5.28: Model effect of the fixed binary factor 'stress' on duration of vowels in Ngan'gi.

account for somewhere between 2 and 20% of the data each, and no single speaker has a clear dominance over the dataset in this respect.

As there are so many Ngan'gi speakers included in this study, we can see quite a bit of interspeaker variation in these results. For vowel duration, as in Fig. 5.28, five of nine speakers (4, 7, 1, 2, and 3) have significant effects, as is the overall effect. The other speakers (5, 6, 8, and 9) do not have significant effects of duration, however. As laid out in Figure 5.28, speakers vary along what seems like a gradual cline of difference, although there is a clear difference between those with significant versus non-significant effects.

All Ngan'gi speakers have a significant effect of stress on vowel intensity, as shown in Figure 5.29. Again, there is a clinal progression of effect sizes, from around 1 dB to

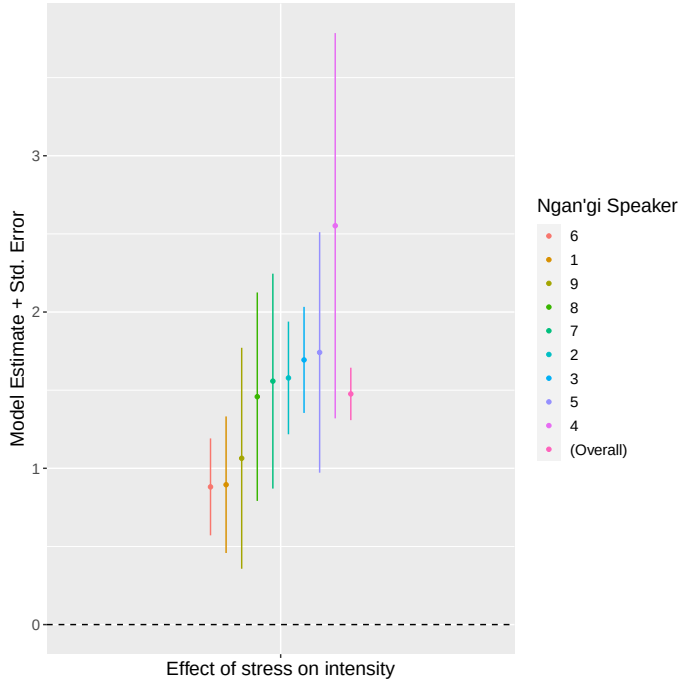

Figure 5.29: Model effect of the fixed binary factor 'stress' on intensity of vowels in Ngan'gi.

around 2.5 dB. These are not especially large effects, however, and as discussed in §4.3 it is difficult to draw strong conclusions about the intensity results in general.

Most Ngan'gi speakers have a positive effect of stress on normalized f0. Three speakers do not have a significant effect here (Speakers 1, 5, and 9). Speaker 8 has the largest effect size at around 1.5 st, and speaker 4 has an effect of about 1 semitone. All other speakers are under a 1 st effect size and have effect sizes that are relatively close to one another.

The Ngan'gi language community numbers in the low hundreds (cf. Chapter 2), and it is likely from these results that there is some sociolinguistic variation in the realization of stress within this community. Further research is needed to determine the sources of this variation.

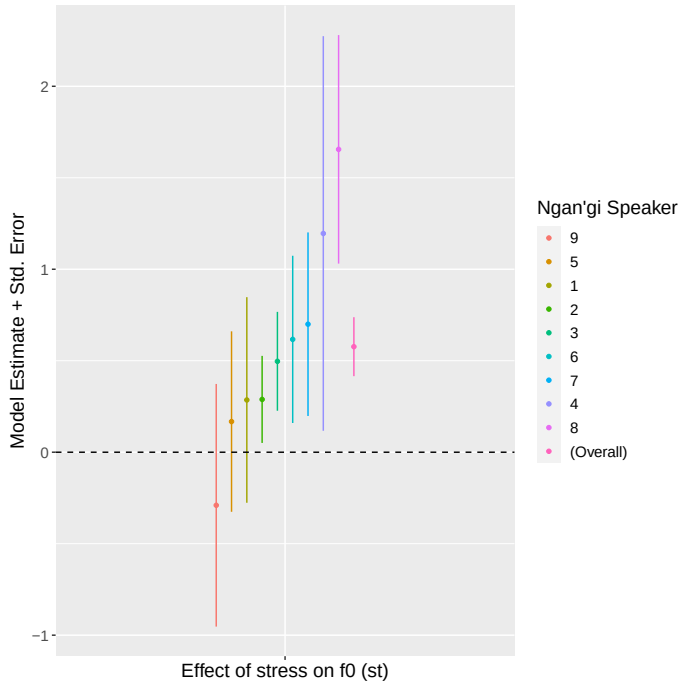

Figure 5.30: Model effect of the fixed binary factor ‘stress’ on f0 of vowels in Ngan’gi.

5.10 Yidiny

There are two Yidiny speakers in the data for this project, DM and TF. DM accounts for around 62% of the Yidiny data, while TF accounts for the remaining 38%. Both speakers are very close in their effects for each of the three factors looked at here, despite the fact that they speak different dialects.

The effect of stress on duration for both Yidiny speakers is significant. Effect sizes do not differ greatly from the overall effect of 0.1 normalized duration units for either speaker.

Intensity is positive and significant in Yidiny, and both speakers have effect sizes very close to the overall effect of about 2 dB. These results are shown in Figure 5.32.

Finally, the effect of stress on f0 is significant and positive for both Yidiny speakers. The effect for DW/1 is around 1.2 semitones, while speaker TF/2 is closer to 1.6 semitones. Both of these are relatively large effects and indicate a strong correlation of f0 with stress

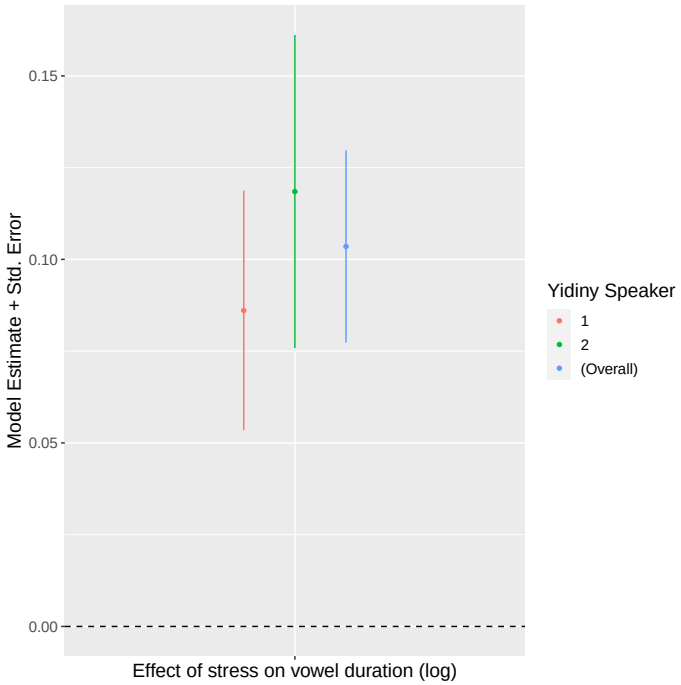

Figure 5.31: Model effect of the fixed binary factor ‘stress’ on duration of vowels in Yidiny.

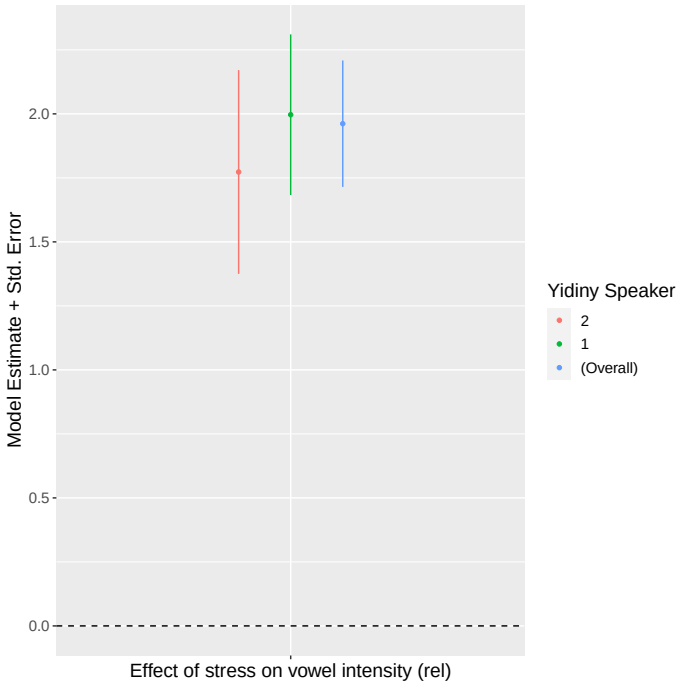

Figure 5.32: Model effect of the fixed binary factor ‘stress’ on intensity of vowels in Yidiny.

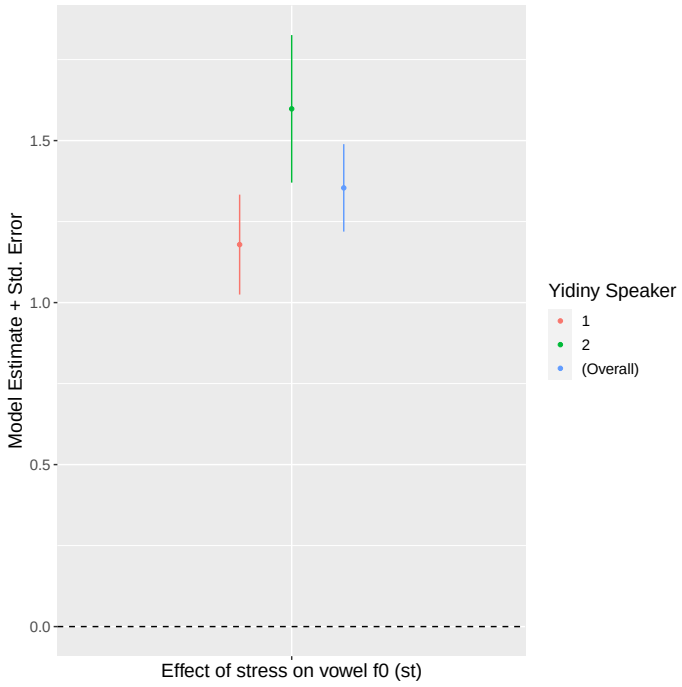

Figure 5.33: Model effect of the fixed binary factor ‘stress’ on f0 of vowels in Yidiny.

in the language.

5.11 Summary

The interspeaker results presented in this chapter have demonstrated that, like all other linguistic processes, the acoustic correlates of stress are not always consistent across speakers of a language. In some cases speakers are consistent with one another in this regard, but some correlates tested show a clear grouping where some speakers use a correlate and others do not. Some correlates also show some interspeaker variation, but the greater proportion of one speaker over the others can sometimes influence the overall language effect disproportionately. Others show a more even split among speakers that are clear indicators of some sort of principled variation within the language. These results address the claim made in (1.2a) in Chapter 1: variation in the cues to stress across speakers is indeed found. However, a more controlled investigation into these data may provide stronger evidence to this point. The data for each speaker as presented here is different, i.e. there are different word types and differing numbers of word tokens from each speaker. A more controlled study, which I leave for future research, would look specifically at word types that are shared by all speakers and compare these directly, thus controlling for environment and frequency effects that are not completely controlled for here. In addition, while some speculation about the sources of this variation has been noted, addressing claim (1.2b), that this interspeaker variation will fall along sociolinguistic lines, is not possible within the scope of this dissertation. However, specific areas to look at in this regard have been pointed out and could be a fruitful area of future study.

The results presented in this chapter, along with results that were not explicitly discussed here but can be found in Appendix C, are merged with the overall language results from Chapter 4 in Table 5.8 for those languages that have more than one speaker in their data set. Broadly, we can see in this summary table that while all acoustic factors may be subject to interspeaker variation in these languages, certain measures such as vowel duration, in-

Language	V. Dur.	Onset Dur.	Post-T. Dur.	Inten.	F0 Max.	F0 Rng.	Vowel
Bardi	+	+		∼	∼	+	+
Burarra	∼	∼					
Gunnartpa	∼	∼					
Gija	∼	∼∼			+	+	+
Dalabon		∼					
Gunwinggu			+	+	∼		
Kunbarlang		+			∼		
Kayardild	+		+ ∼∼	∼	+	+	+
Malak Malak	+	+ ∼		+	+		
Ngan'gi	+	∼∼		+	∼		
Yidiny	+	+	+	+	+		

Table 5.8: Summary of results; + indicates a statistically significant effect overall in the language that may be attributable to stress; ∼ indicates a statistically significant effect among some (but not all) speakers in the language that may be attributable to stress. Languages grouped by historical affiliation.

tensity, and f0 maximum are consistent across speakers in at least half of the languages for which interspeaker variation could be investigated. The consonant duration measures (onset and post-tonic) tended to have more interspeaker variation, although some languages do show this effect consistently across speakers. The interaction of cross-linguistic and inter-speaker results addresses claim (1.3) in Chapter 1: some stress correlates hold for all speakers, while others are apparently less stable and exhibit speaker variation.

One factor to consider with these results is that these languages vary in the number of speakers they have. A language with only two speakers, such as Yidiny for example, may more easily have effects that are consistent across speakers than a language with more speakers. However, this is not the only factor in determining whether a language has consistent or variable patterns here. In Ngan'gi, for example, there are nine speakers, and there is variability in these speakers' use of onset lengthening and f0 maximum, but vowel duration and intensity are fully consistent across speakers. Likewise, in Kunbarlang there are only two speakers, but we still see variability in their use of f0 maximum.

A full understanding of any linguistic phenomenon requires not just looking at lan-

guages as a whole, but also considering the variation across speakers. Including speaker variation in any phonetic study can reveal interesting patterns that might be glossed over when binning all speakers together (Cohn & Renwick 2021). Considering speaker variation in this dissertation is especially important considering the variance in each speaker's proportion of a language's data set; the speaker with the largest proportion of the data is likely to overshadow other speakers' effects.

Now that individual models have been presented both across languages and across speakers within each language, the question arises of how to quantify the variation we have observed. In the following chapter, methods borrowed from the study of genetics are used to address this.

Chapter 6

Quantifying variation with phylogenetic methods

The previous two chapters have presented the results of investigations into the phonetic correlates of lexical stress in each of the sixteen languages investigated here, and for each speaker within each of these languages. This has begun to provide evidence for Claims (9.1), that there is cross-linguistic variation in the cues to stress, and (9.2), that there is inter-speaker variation. However, only broad and qualitative differences have been discussed so far in Chapters 4 and 5; we still want to know whether this observed variation is structured such that we can propose that stress cues undergo regular sound change, as stated in part (c) of Claims (9.1-9.2). In this chapter, a method for quantifying both inter- and intra-language variation in stress correlates is presented: Analysis of Molecular Variance, or AMOVA. This is a model well-established for use in biological phylogenetics, and also used more recently to model types of cultural evolution as well. The model quantifies the sources of variation in the data at three levels: within speakers, across speakers within language, and across languages. Some background on this method is presented in §6.1, and details on the implementation of this model is given in §6.2. The AMOVA results presented in §6.3 find

each of these sources to account for significant portions of the variation.

Using another phylogenetic measure of variation, this chapter also considers pairwise variation across speakers and languages using fixation index, or F_{ST} (6.4). This measure represents the amount of overall variation between each pair of speakers; this information then serves as input to a Neighbor Net network model to visualize relationships between languages based on the acoustics of stress, where close relationships are found between all the Pama Nyungan languages in this study, and some other historical and areal groupings are identified. The chapter ends with some discussion of these results in §6.5.

6.1 Background

6.1.1 The use of phylogenetic methods in linguistics

It has been a longstanding observation that biological and linguistic change seem to follow similar evolutionary principles (Darwin 1871, Croft 2000, Atkinson & Gray 2005). Phylogenetic methods, originally used to model biological evolution, became established as a way of modeling linguistic structure around the turn of the 21st century (Gray & Jordan 2000, Ben Hamed 2015). Phylogenetic models have been applied to historical studies of many language families, including Pama-Nyungan (Bowern & Atkinson 2012), Bantu (Holden & Mace 2003), Semitic (Kitchen et al. 2009), Turkic (Savelyev & Robbeets 2020), Indo-European (Gray & Atkinson 2003), and many others. The focus has largely been on the creation of tree models using Bayesian inference to determine likely historical relationships between groups of languages, but work has also looked at the contribution of specific linguistic phenomena to a model of language relatedness (Bowern 2018a, Macklin-Cordes, Bowern & Round 2021).

6.1.2 Analysis of Molecular Variance

The quantification of variation across all languages is calculated here using the phylogenetic modeling tool Analysis of Molecular Variance (AMOVA), which is used in the field of biology to measure variation among and within genetic population groups. The model output gives the percentage of variation in the data that comes from (a) within populations (for our purposes, within speakers); (b) across populations within groups (across speakers within languages); and (c) across groups (here, groups are languages). Another way of measuring amount of variation across languages is the F_{ST} statistic, which gives the percentage of difference between each pairwise grouping of speakers/languages.

The Analysis of Molecular Variance was first presented in Excoffier, Smouse & Quattro (1992) for studying variance among DNA haplotypes within a single biological species (479). The method is similar to an analysis of variance (ANOVA), with modifications made to be tailored to genetic population variation and to consider variation at three levels: within populations, across populations within larger groupings, and finally across groupings. In this initial demonstration, Excoffier, Smouse & Quattro (1992) looked at groups of ten human genetic populations, grouped in pairs into five geographical regions. The within-population variance was thus variation within one genetic population of humans; the cross-population, within-grouping variation was the variation across each pair of populations within their region; and the cross-grouping variation was the variation across the regional groupings. Most of the variation (about 75%) was found within populations, about 20% was found to be attributable to differences across groupings, and around 3% of the variation was attributable to differences across populations within larger groupings (Excoffier, Smouse & Quattro 1992: 486).

Excoffier, Smouse & Quattro (1992) are quick to note that AMOVA can be a useful tool for studies of variation beyond just DNA haplotypic variation within biological species,

calling it a "coherent and flexible framework for the statistical analysis of molecular data" (479). This model was soon used for studying variance in other types of biological genetic data (Michalakis & Excoffier 1996). More recently, this model has been applied to studies of variance in aspects of human culture, specifically music (Rzeszutek, Savage & Brown 2012) and folktales (Ross, Greenhill & Atkinson 2013). It has not to my knowledge been used for linguistic data in any published work. This is because of the focus on tree-based modeling for groups of related languages using linguistic factors that have been established to undergo regular historical change. In this project, I want to first establish that stress variation exists in a systematic way such that we can expect it to undergo regular change, thus motivating an investigation into the structure of this variation, which AMOVA is particularly well-suited for.

For the purposes of using this tool from biology on linguistic data, an analogy from biological to linguistic data must be established. We will use the example of AMOVA data for determining variation among colonies of bacteria for this exercise. In biology, each population is essentially one colony of bacteria, e.g. in a Petri dish; each coded line within the population is a genetic sequence for one bacterium from this colony. In language, one population is taken to be all speech coming from one individual; that is, the collection of each speaker's utterances is taken to be a 'colony.' Each coded member of the sample is taken to be a relatively short utterance of 100 words (see the following section for more discussion on this point). These populations are then grouped together into what in biology may be colonies of bacteria that are the same species; for language, we will group together speakers (our samples) into languages. Thus we have a similar stratification of within-population variation (variation across utterances); cross-population within-group variation (cross-speaker within-language); and cross-group variation (cross-linguistic).

6.2 How AMOVA is implemented

Analysis of Molecular Variance can be implemented using the phylogenetics program Arlequin (Excoffier, Laval & Schneider 2005). This program takes a specific type of project file with a `.arp` extension. The structure of this input file is similar to other file structures used for phylogenetic modeling programs. There are two mandatory sections, 'Profile' and 'Data — Samples', with optional sections specifying a distance matrix, a list of haplotypes, genetic structure, and Mantel tests. Arlequin will generate this information automatically unless these sections are specified; for the present implementation of AMOVA, the automatic generation option for the distance matrix was chosen, while genetic structure was specified in order to define speakers as part of a language group.

In the 'Profile' section, formatting characters are specified such as the locus separator, which is usually whitespace, and the character to indicate missing data, usually '?'. This section also requires specifying the number of samples and the type of data; for this implementation of AMOVA, the data type was set as 'standard' haplotype data, meaning the Arlequin program was expecting multi-state coded characters, i.e. groups binned into categories notated '0', '1', '2', etc.

The 'Data' section includes all other required and optional sections of the input file. The one mandatory section here is 'Samples,' which is the main data portion of the file. For biological implementations of this model, a sample would consist of the genetic sequences of the members of different biological populations. For my purposes here, each sample is a speaker, and each line in the sample represents a 100-word utterance sampled from that speaker's audio data. It is important to have multiple coded lines in each sample for two reasons. First, within-speaker variation will be measured based on the variation in the members of each sample, so having one line per sample would gloss over within-speaker variation and artificially inflate cross-speaker and cross-language results. Second, binning

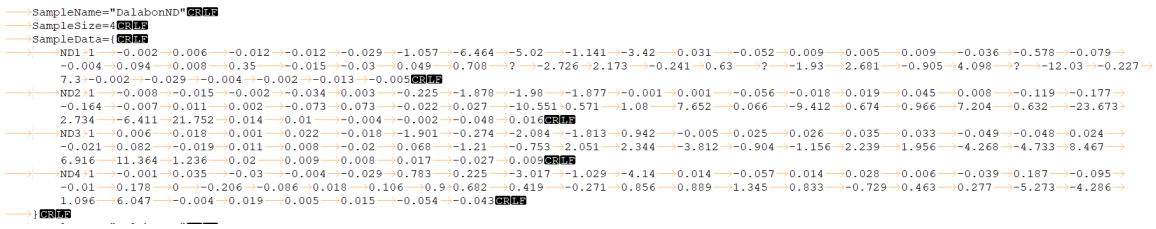

Figure 6.1: Example of one sample in an Arlequin input file using raw data measurements.

all word tokens together for each speaker would not be very informative; across time and especially across audio files, the environment in which the speaker is speaking may change substantially, including their distance from the microphone as well as levels of background noise. By breaking up each speaker's data into smaller chunks, the likelihood that conditions will change to a large degree is reduced. The decision to use 100-word utterances (instead of 50 or 200, for example) was arbitrary on my part; I leave an investigation into the optimal size of utterances for this type of modeling to further research.

There were a total of 472 of these utterance chunks in the input file, across 50 speakers of 16 languages. As an example of what these data look like, Figure 6.1 shows the coded sample for Dalabon speaker ND. This speaker has around 400 word tokens in their audio data, which were split up into four lines in the sample coding. Each character on the line represents the arithmetic difference between the average X measurement of stressed Y syllables and the average X measurement of unstressed Y syllables, where X is one of the acoustic measurements listed in (1) that was investigated as a potential correlate of stress, and Y is one of the vowel qualities /a/, /i/, or /u/, or a consonant category 'stop', 'nasal', or 'glide', depending on X.

1. Measurements included in Arlequin input file:

   (a) DUR: vowel duration

   (b) INT: vowel intensity

	/a/	/i/	/u/
DUR	DUR-A	DUR-I	DUR-U
INT	INT-A	INT-I	INT-U
F0MAX	F0MAX-A	F0MAX-I	F0MAX-U
F0MIN	F0MIN-A	F0MIN-I	F0MIN-U
F0RNG	F0RNG-A	F0RNG-I	F0RNG-U
F1	F1-A	F1-I	F1-U
F2	F2-A	F2-I	F2-U
EU	EU-A	EU-I	EU-U

Table 6.1: Vowel characters coded for AMOVA model input.

(c) F0MAX: f0 maximum

(d) F0MIN: f0 minimum

(e) F0RNG: f0 range

(f) F1: first formant

(g) F2: second formant

(h) EU: Euclidean distance for vowel space

(i) OD: onset consonant duration

(j) PTD: post tonic consonant duration

The first eight measurements listed in (1), all of which are vowel measurements, were crossed with the vowels /a/, /i/, and /u/, to get a total of 24 vowel-related characters in the Arlequin input file, as enumerated in Table 6.1. These three vowels are the only ones considered here because they are the three vowels that are shared by all sixteen languages being analyzed. Including vowels that are not in all of these languages would cause problems for the AMOVA results, because there would be large amounts of missing data in the input file.

The last two measurements listed in (1) are durations of onset consonants and post-tonic consonants respectively. These were crossed with the three categories that consonants were binned into when investigating consonant durations as correlates of stress: stops, nasals, and

	STOP	NASAL	GLIDE
OD	OD-STOP	OD-NASAL	OD-GLIDE
PTD	PTD-STOP	PTD-NASAL	PTD-GLIDE

Table 6.2: Consonant characters coded for AMOVA model input.

glides. This results in six characters as listed in Table 6.2. There is a total of 30 characters listed for each 100-word sample for each speaker in each language in this study.

The differences in stressed versus unstressed vowels for each of these characters will usually be a positive number when the stressed vowels are more prominent on average than the unstressed ones for a particular utterance. They will usually be zero or a negative number when there is no stress-related difference for that character. To highlight this generalization in the model, and also for ease of use with the Arlequin program, multistate character coding was used. Positive numbers, which correspond to stress correlation, were coded as '2'; negative numbers were coded as '1'; and zero values were coded as '0'. True zeroes were rare, with fewer than 200 coded for in the input file with over 14,000 individual items coded across all speakers and languages. Figure 6.2 shows the same sample for Dalabon speaker ND as in Fig. 6.1, now with this multistate character coding.

In some cases, especially for the consonant-related characters, there may not be a value for a certain character. This is because some 100-word chunks may not have nasals in onset position in both stressed and unstressed syllables, for example. These instances are marked as missing data with a '?' symbol, as can be seen in a few instances in Fig. 6.2. No coded character had greater than 25% missing data across all speakers, and 27 of 30 characters

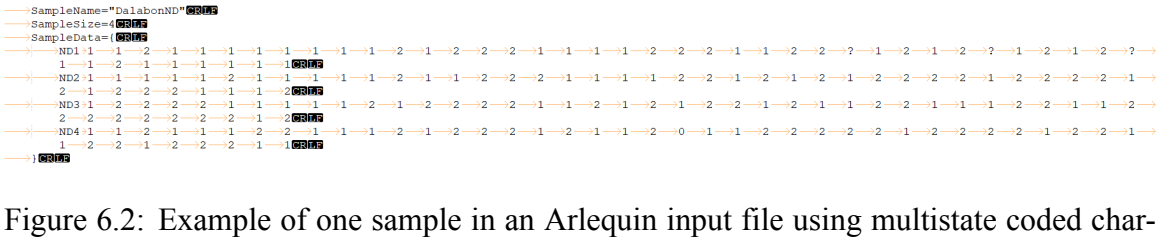

Figure 6.2: Example of one sample in an Arlequin input file using multistate coded characters.

had less than 5% missing data.

Finally, the 'Genetic Structure' section of the Arlequin input file was also specified. This section allows for specification of each speaker into a group, in this case a language, so that the AMOVA results can also include cross-linguistic variation. There were 50 speakers across 16 languages, as already described in Chapter 2. Other optional sections, including the distance matrix, were calculated automatically in Arlequin in order to run the AMOVA.

The final coded Arlequin input file is available in full online at `https://doi.org/10.5281/zenodo.6354656`.

6.3 AMOVA Results

The results of the AMOVA modeling are given in Table 6.3. The variation at all three levels the model considers is significant. The largest proportion of the variation is within speakers (87.2%), cross-linguistic variation accounts for 7.6% of the variation, and inter-speaker within-language variation for 5.3% of the variation in the data. These sorts of numbers are typical for AMOVA results in both genetic and cultural populations to which it has been applied (e.g. 2.06% between populations (music) in Rzeszutek, Savage & Brown (2012); 9.1% between populations (folktales) in Ross, Greenhill & Atkinson (2013); also see Lewontin (1972), Rosenberg et al. (2002) for summaries of human genetic variation). However, the question naturally arises why so much of the variation is intraspeaker, and whether the small numbers for interspeaker and cross-linguistic variation are in fact meaningful for the quantification of linguistic difference in the phonetic correlates of stress. I argue that these

Source of variation	d.f.	Sum squares	Var. components	% of var.	p-value
Across languages	14	369.48	0.49	7.55%	< 0.001
Across speakers, Within language	35	261.99	0.34	5.26%	< 0.001
Within speakers	422	2390.35	5.66	87.19%	< 0.001

Table 6.3: AMOVA results.

results are meaningful for the following reasons. First, and probably most important, variation in any of these phonetic factors is likely to be highest when a factor is not a correlate of stress, and that variation should be seated comfortably within speaker. For example, Ngan'gi has no effects of stress on post-tonic consonant duration. Thus, we would expect that the coding for all three post-tonic consonant duration characters should vary essentially at random; thus, the measure of variance in all three of these characters will be quite high. Subsequent work could potentially address this issue by coding all non-significant correlates as '0'. Second, it is unlikely that all the correlates found in the regression results (cf. Table 4.1) would be considered the primary correlates of stress in the minds of speakers; however, with perception experiments being unavailable for the current study, we do not have a way to identify these except by picking out correlates with significant but small effect sizes. If a stress correlate is secondary in the mind of a speaker, the obligatoriness of the correlation holding for any given 100-word span of utterance may vary, thus increasing the within-speaker variation further. Thus we can expect that any results of this type will have large amounts of within-speaker variation, but that does not discount the statistical significance of the other levels of variation.

The across-speaker within-language results indicate that about 5.2% of the variation in the data comes from interspeaker variation. This result is more or less borne out in the discussions of speaker variation in Chapter 5. There are clearly interspeaker effects in the use of certain stress correlates in the languages in this study, just as there are interspeaker effects in any other linguistic phenomenon (cf. Cohn & Renwick (2021)). This source of variation could be a sign of changes in progress, or of stable sociolinguistic variation within the language community.

Finally, cross-linguistic variation is found to account for about 7.6% of the variation in the data. This source of variation can be the result of historical changes, or of significant cross-linguistic differences between unrelated languages. While the AMOVA model does

not point out the specific sites of variation that contribute to this overall number, it does indicate that the languages in this study are varying with respect to their stress correlates separately from the variation that comes from intra- and inter-speaker sources. In order to look more specifically at the differences between each pair of languages, we can look at the F_{ST} statistic, which is discussed in the following section.

6.4 Pairwise fixation index

Another way of looking at differences between and within languages is a pairwise fixation index, or F_{ST}. This is a measure of genetic distance between populations that ranges from 0 to 1. Values closer to 0 indicate that genetic distance is smaller and the populations being compared are more similar. A value of 0 means that the populations are identical, while a value of 1 indicates no similarities whatsoever between populations. In biology, values between 0 and 0.25 are expected between two populations that have some amount of interbreeding and thus are genetically quite similar (*Ecology Disrupted* 2021).

Pairwise F_{ST} values were calculated in Arlequin for each population in the AMOVA input file; that is, these values were calculated for each speaker-speaker pair both within and across languages. Our expectations for F_{ST} values does not match up completely with the expectations for these values in biology. In biology, a value over 0.25 likely indicates that there is little contact and little genetic mixture between populations, but we may not expect the same standard cutoff point to be meaningful for language data. All of the similarity values are likely to be on the lower end of the F_{ST} scale, because there is likely to be a lot of overlap between speakers, especially when neither speaker is using some acoustic factor to mark stress and there is random-like variation. The fixation indices will then need to be evaluated on a smaller scale than biological data might.

Figure 6.3 gives a visualization of all the speaker-speaker pairwise F_{ST} values across all languages. The histogram plot shows that these values do indeed trend low, and most

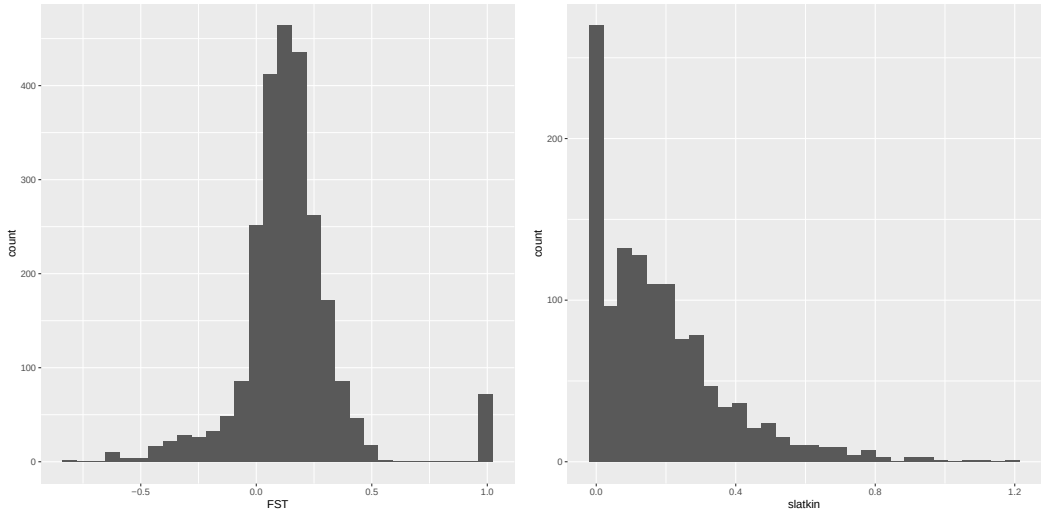

Figure 6.3: Histograms showing the distribution of pairwise F_{ST} values across all languages (n=2500). On left: raw F_{ST} values. On right: Slatkin-corrected F_{ST} values where all negative numbers go to zero.

values are close to 0.2. Some of the values generated by Arlequin are negative, but for the purposes of interpreting F_{ST} a negative number is equivalent to zero. In the two plots in Figure 6.3, the left shows all the raw values, while the right shows the corrected F_{ST} values corrected using Slatkin (1995)'s linearized F_{ST} measure.

The pairwise fixation index values were used as an input to a Neighbor Net network plot. All non-significant values were excluded and coded as '0'. Remaining Slatkin linearized F_{ST} values were binned into ten 0.1-sized bins plus one bin for zero values. This allows for the R package `phangorn` (Schliep 2011) to recognize the coding and create a distance matrix for input into the network model. For readability and to more easily compare across languages, all speakers of a language were averaged together for this part. The distance matrix was calculated as Hamming distances (Hamming 1950) between the characters.

The output of the Neighbor Net is shown in Figure 6.4. The structure of these languages does not look treelike as one might expect from clearly historically determined phenomena,

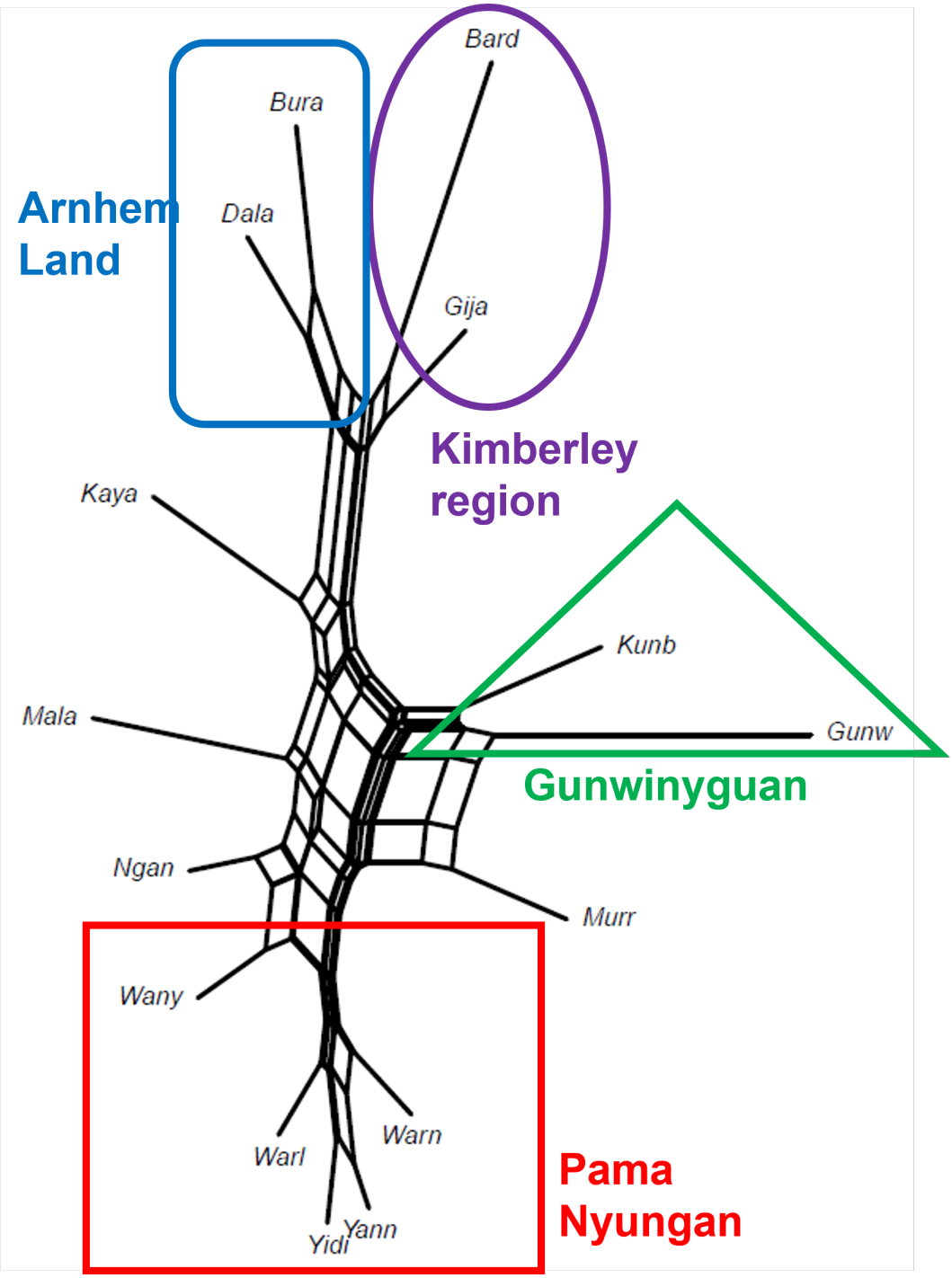

Figure 6.4: Neighbor Net based on Slatkin linearized F_{ST} values.

but of course not all of these languages are thought to be historically related to one another. However, there are some notable groupings of languages in this network. The bottom of the NeighborNet shows a clear grouping of the Pama Nyungan languages (the red box in Fig. 6.4). Wanyjirra seems to be an outgroup here, but Warlpiri, Yidiny, Yannhangu, and Warnman all group together more closely and separately from the other languages here. Another historically related group that shows some clustering here is Gunwinyguan, although only Kunbarlang and Gunwinggu are close together in Fig. 6.4 (the green triangle). Dalabon is not close to the other Gunwinyguan langauges, instead forming a group with Burarra in the blue box on the upper left. These languages are both spoken in Arnhem Land, so the connection may be areal here; however, many other languages here are also spoken in Arnhem Land, so it is not a completely straightforward connection. Finally, Bardi and Gija are close together in Fig. 6.4, in the purple oval. These languages are both spoken in the Kimberley region, in this case the only two languages from this region. It is possible that this grouping results from some areal effects on the phonetic cues to stress. The remaining languages in Fig. 6.4— Kayardild, Malak Malak, Ngan'gi, and Murrinh Patha— are not clearly grouped with any other languages in the NeighborNet.

6.5 Summary

This chapter has presented an investigation into the population structure of the dissertation languages based on the acoustic factors related to lexical stress. The language data used in this project was broken up into many 100-word samples in order to create the appropriate data type for the modeling and to ensure that comparison is done with samples of the same size. Using Analysis of Molecular Variance, we were able to identify the sources of this variation. Most of the variation we see in these data comes from within each speaker; as discussed, we expect this to be the case because of the nature of the data being investigated. The AMOVA model also found significant amount of variation attributable to both cross-

speaker, within language variation, and cross-linguistic variation.

The AMOVA results presented here address claims (1.1c) and (1.2c) from Chapter 1. Through this study of population structure, we have seen that there is significant variation that is attributable to cross-linguistic variation (claim (1.1c)) to the exclusion of other sources of variation; and that there is significant variation across speakers of the same language (claim (1.2c)) to the exclusion of other sources.

This variation was investigated further using fixation indices, or F_{ST}. This gave us pairwise values for each speaker pair indicating the similarities in the data between them. These values were generally small, most values falling in the lower 50% of the value range. The results were used as input to a Neighbor Net visualization of the relationships between languages. The result showed some interesting relationships between some languages. Most strikingly, the Pama Nyungan languages all grouped together in the network (Fig. 6.4). Furthermore, we saw two of three Gunwinyguan languages grouped together as well, and potential areal patterns seen in the Kimberley region and in a subset of the languages spoken in Arnhem Land.

The F_{ST} results address claim (1.1b) from Ch. 1: "closely related languages will be more likely to share cues to prominence than languages that are more distantly related." The grouping of Pama Nyungan and Gunwinyguan languages here support this claim. However, we also observe some apparent areal patterns as well, suggesting there may be contact-induced change in some cases in the phonetic cues to stress.

The results presented in this chapter indicate that variation in the phonetic correlates of lexical stress is structured similarly to other types of principled linguistic change. Significant variation is found between languages as well as across speakers of the same language, and similar patterning is found between historically related languages as well as unrelated languages that are in contact with one another. These findings are promising for the pursuit of research in prosodic-phonetic change in Australian languages and more generally as

well.

Chapter 7

Phrasal prosody

In this chapter, we turn to a topic not yet investigated in this dissertation: phrasal prosody. While the processing and analysis of lexical stress correlates in Chapters 3-6 was simplified by the fact that most Australian languages have consistent initial stress placement (Goedemans 2010, Fletcher & Butcher 2014), this is not the case for categories of phrasal contours in these languages. For this reason we must first create hypotheses about what these contour categories are before folding them into a broader analysis of prosodic change. This is not a simple task, especially without close knowledge of the languages at hand. This chapter presents a case study in using an automated method for prosodic contour identification, first proposed by Kaland (2021), but this is only the first step in a strong theory of prosodic contours in these languages, and further work is required to test the hypotheses established here. The larger question of including phrasal contours and other higher-level prosody in a model of prosodic change goes beyond the scope of this dissertation, and I leave it for future research, but I present the results of this automatic analysis of prosody as a proof of concept for conducting such analyses on archival audio.

Identifying types of phrasal prosody is a particularly difficult problem in language documentation work for two reasons. First, prosodic work requires specialized data collection

methods and command of the prosodic literature (Gussenhoven 2004, Himmelmann 2008). For this reason prosodic documentation has not historically been a focus of general language documentation efforts; it is not until recent decades that recording equipment and analysis tools have become accessible and portable enough for this to be the case (Macaulay 2021). Second, prosodic work usually requires deep knowledge of the language's syntax and discourse structure. Intuitive analyses are subject to error based on perceptual biases of non-native speakers (Kaland 2021, Xu 2011). All of this is even more difficult when working with archival materials without access to native speakers' intuitions and judgments about the meaning that prosodic structures encode (Simard & Schultze-Berndt 2011). As a result, detailed intonational studies for endangered languages are rare.

Phrase-level prosodic description in Australian languages is even more rare than word-level stress descriptions, reflective of the broader trend of this type across language documentation work in all areas of the world (Fletcher & Butcher 2014, Macaulay 2021). However, some work exists for languages such as Kayardild (Fletcher, Evans & Round 2002), Arrernte (Tabain 2016), Jaminjung (Simard 2010), Bininj Gun-wok (Bishop 2003), Djambarrpuyngu (Jepson 2019), and Dalabon (Ross, Fletcher & Nordlinger 2016), among others. Fletcher, Evans & Round (2002) note that while phrasal prosody in Bininj Gun-wok has been found to function exclusively as a way to demarcate the edges of phrasal units (Bishop 2003), Kayardild seems to use phrasal prosody for both this demarcative function as well as phrasal prominence functions such as focus marking. This finding underscores the importance of looking at phrasal prosody in more Australian languages, as there is clearly variation that has yet to be surveyed in a comprehensive way.

Prosodic phenomena are as important an aspect of a language as any other linguistic phenomenon, one of many crucial topics to create a full understanding of a language's phonological system. However, prosodic description has not become a standard aspect of language documentation in the same way that segmental description is (Himmelmann

2008, Macaulay 2021). A major roadblock to describing the prosody of a language is the formation of initial hypotheses. With initial ideas about what prosodic contours to look out for, along with some knowledge of the language, it becomes easier to construct targeted sentences and texts to get more examples of a particular phrase type, and to get multiple speakers to say the same sentences in the same context to study prosodic contours more carefully (Himmelmann 2008). One possibility to create these initial hypotheses without investing a lot of one's own research time, which is often limited due to funding constraints, is to first run a bottom-up, automatic categorization algorithm to identify phrasal contours. Besides saving time for the researcher to identify likely phrase types, an automatic categorization algorithm has the notable benefit of sidestepping non-native speaker biases in identifying prosodic phenomena.

This chapter presents results of an automatic analysis of phrasal prosody in the Australian languages included in this dissertation. To maximize comparability across languages, the process is kept as automatic as possible, only considering existing phonological accounts of phrasal prosody in late stages of analysis, whenever such accounts exist. In what follows, each of the sixteen languages in this study is considered in turn.

These results are just the beginnings of an analysis of phrasal prosody in the project languages— further work would be needed to test the hypothesized f0 contour patterns directly, ideally with targeted perceptual experiments, in order to draw more definitive conclusions. The following analyses contribute a first look at the phrasal prosody of the dissertation languages, in many cases the first analysis of this type for the language. A cluster analysis is run, using the methods laid out in Kaland (2021) for automatic prosodic categorization.

7.1 Background

A small body of recent work has explored the capabilities of automatic methods in identifying prosodic phenomena such as tone and phrasal prosody. This includes Dockum (2017) and Grabowski & McPherson (2019), who have tested methods for automatic categorization of tone contrasts in a variety of languages. Cole & Shattuck-Hufnagel (2016) present Rapid Prosodic Transcription (RPT), which is a fast and easy way of identifying basic prosodic features. This is not the best method for all types of language data and research situations, as it requires training native speakers of the language to do the transcriptions. For my purposes in this dissertation, for example, native speakers are not readily available for this kind of task, and in some cases no native speakers remain for the language, so we require a method that identifies prosodic contours based only on audio data.

Relatively little work has considered methods for automatic categorization of f0 contours used in phrasal prosody, with the notable exception of Kaland (2021), whose toolkit is used for the analysis in this chapter. Using automatic tools for categorization of phrase types, Kaland argues, eliminates effects of non-native impressionistic categorizations and facilitates the identification of subtle contour categories that non-native researchers are likely to miss. While a full and detailed description of phrasal prosody in a given language would be best supported by targeted perceptual experiments, such work is best performed only after a preliminary analysis to establish hypotheses to test, and these hypotheses could be formed without the interference of non-native speaker biases via the use of automatic methods.

Kaland (2021) presents two case studies to demonstrate the performance of the contour clustering toolkit: tone contours in Zhagawa elicited speech, and phrase contours in Papuan Malay spontaneous speech field recordings. The data used for the Papuan Malay case studies consisted of natural speech recordings made in a field setting, with intonational phrases

segmented by a researcher familiar with the language. This is exactly the type of data that I have in this dissertation project, as I am working with archival materials with transcriptions made by the researcher who recorded the audio, often in consultation with native speakers of the language.

7.2 Methods

The contour clustering toolkit described in Kaland (2021) is available at the following URL: `https://constantijnkaland.github.io/contourclustering/`. The scripts included in the toolkit were used for this analysis without any modification to the underlying script, with settings as noted in this section.

This analysis was performed as automatically as possible. One consequence of this is that phrase boundaries were defined as any period of speech between two pauses, as indicated in the original transcripts in the archival materials. This is certainly an imperfect metric for phrasehood; however, I have no reason to believe that this should preclude using it for preliminary analysis, as a time-saving way of investigating phrasal phenomena in this large multilingual corpus. This allows the method to be maximally applicable to research on any language and comparable across languages with varying amounts of prosodic research already done on them.

The scripts available in the toolkit from Kaland (2021) were used with minimal modifications. First I ran the 'time-series f0 measurements' Praat script, which takes f0 measurements at 20 regular intervals across each phrase (Boersma & Weenink 2018). The script defines a 'phrase' as any interval containing text in the specified TextGrid tier, and I define these intervals as an utterance with silence or non-speech intervals before and after it. These intervals were (in most cases; see language descriptions in Chapter 2) set by the researchers who made each language deposit while doing their transcriptions of audio in ELAN. The default settings from Kaland (2021) were kept the same, except that minimum duration was

set to 100 ms instead of 0.1 ms, because these intervals were too short for Praat to take 20 measurements. Pitch minimum was set at 75 Hz, and pitch maximum at 600 Hz, time step was 0.01, and stylization resolution was 2 st.

The clustering method used for this analysis is complete-linkage clustering. This is an agglomerative method with iterative pairwise clustering, similar to another common clustering method, UPGMA (unweighted pair group method with arithmetic mean) (Sokal & Michener 1958). Clusters are formed by grouping elements together that have the minimum distance between them. This is done in an iterative fashion to build up larger clusters, which means that a new distance value needs to be calculated for the smaller clusters in order to continue the method. For complete-linkage clustering, this new distance is calculated as the maximum distance between elements of the cluster. For example, if the first cluster has elements (a, b) and we want to determine the distance between this cluster and a new element c, we choose the larger of the distances between a and c and between b and c. This is what distinguishes complete-linkage clustering from other agglomerative clustering methods; UPGMA clustering calculates this new value as the mean distance between elements of the cluster, and single-linkage clustering uses the minimum distance of the cluster elements.

Contour clustering analysis was run in R using Kaland (2021)'s graphical user interface clustering script. Data was pruned based on the automatic suggestions for subsetting when the number of clusters was set at $n = 25$, as was done in Kaland (2021). After pruning, the number of clusters was set at $n = 2$, and then investigated at each $n + 1$ until I judged that additional clusters were uninformative. This was usually because the cluster added when looking at $n + 1$ clusters had a contour that was extremely similar to a contour in an existing cluster. This is the same sort of judgment recommended in Kaland (2021) to settle on the number of clusters in a relatively objective way apart from knowledge of other work on prosody of the language.

F0 measurements were standardized by speaker. Each file was normalized by subtracting the mean f0 from each measurement; therefore, in the results that follow, the zero mark indicates average f0 and positive and negative numbers are higher or lower deviations from that mean.

7.3 Results by language

7.3.1 Bardi

The Bardi data was normalized by filename (the proxy for speaker used in the contour clustering script) by subtracting the mean, thus centering all measurements around zero. Data was subsetted based on the automatic subsetting at 25 clusters, as recommended in Kaland (2021). After subsetting, one notable cluster with $n = 5$ remained, and clusters were explored both with and without this cluster because of its small size. After subsetting, the original $n = 1940$ contours (utterances) was cut down to $n = 1852$ (95.5%) with the small cluster, or $n = 1847$ (95.2%) without it. There were nine rounds of data subsetting before no automatically-detected subsets remained.

The cluster dendrogram for Bardi is shown in Figure 7.1. The branching here shows a major split separating roughly one third of the data from the rest, a split that corresponds to the two major clusters in Figure 7.2. The larger of the two clusters (on the right in Fig. 7.2) corresponds to a mid tone that falls over the utterance, while the smaller cluster (left in Fig. 7.2) corresponds to a higher-than-average tone that falls over the course of the utterance.

The contours in Figure 7.2 show the major split at the top of this dendrogram. The more common contour (Cluster 2 in the figure) shows a relatively flat contour, beginning at average f0 and lowering a bit over the course of the phrase. The smaller cluster (Cluster 1) shows a higher-than-average f0 at the beginning of the phrase, which lowers to average by

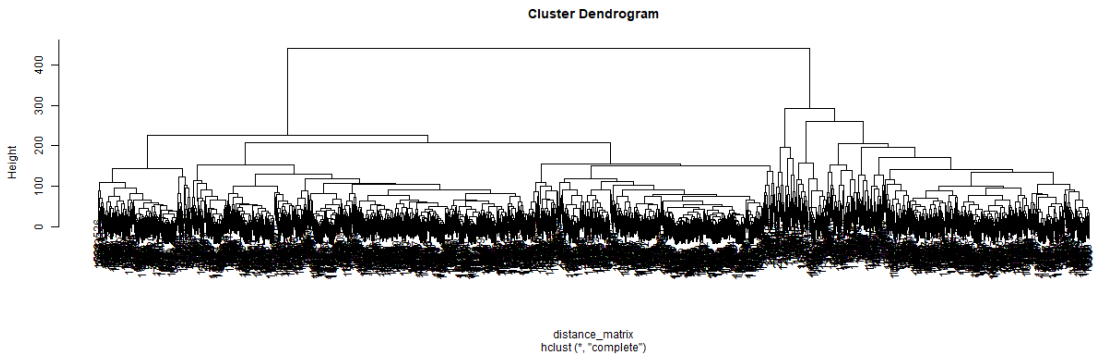

Figure 7.1: Cluster dendrogram for Bardi, after subsetting of data.

the end.

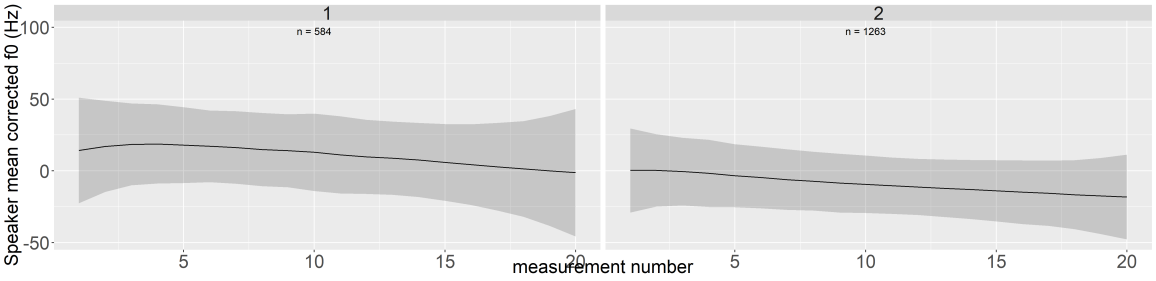

Figure 7.2: Two major contour clusters in Bardi.

However, looking at a two-way split in contour clusters misses a lot of meaningful variation in Bardi phrases. Figure 7.3 shows the contour clusters detected when there are eight clusters specified, and more distinct phrase types make themselves clear. The flat contour with average f0 remains (Cluster 3 in the figure), now one of three relatively flat contours along with a low-f0 and high-f0 contour (Clusters 4 and 3 respectively). These are joined by two falling contours, one mid-to-low (Cluster 5) and one a more dramatic (but much less frequent) high-to-low contour (Cluster 6). Cluster 7 shows a high f0 gesture that is maintained throughout the phrase, followed by the low boundary tone at the end, while Cluster 8 shows an extra-high contour that seems relatively stable across the phrase, though this cluster seems to have more variation than the others. The only rising tone is seen in Cluster 2, where a slightly above average f0 rises at the end of the phrase.

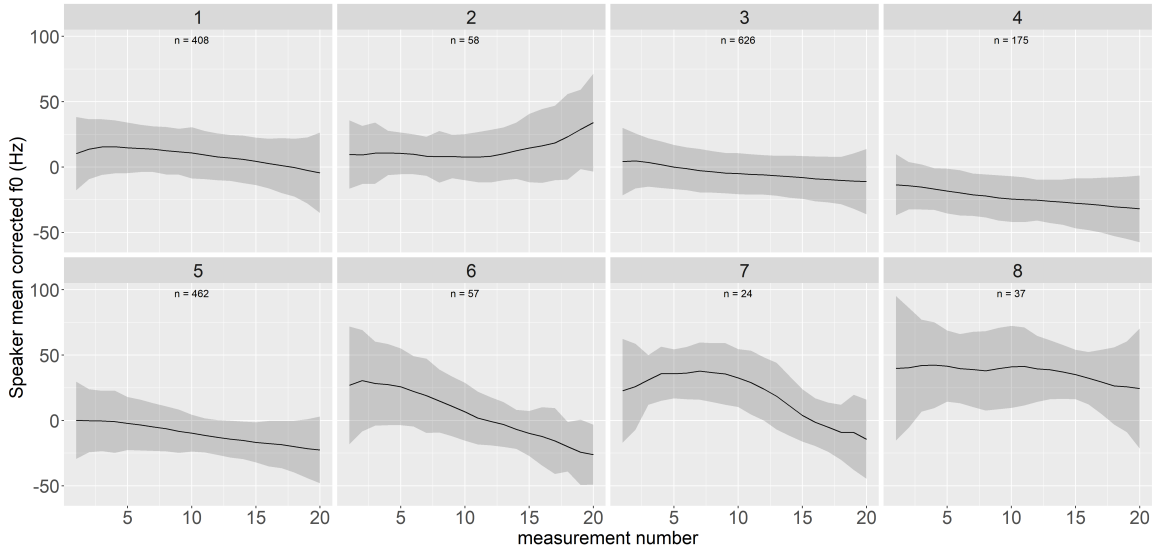

Figure 7.3: Eight contour clusters in Bardi.

7.3.2 Burarra

The Burarra audio files included 1565 segmented utterances. After data subsetting, 1484 utterances (94.8%) remained in the clustering analysis. The dendrogram for Burarra is given in Figure 7.4; there are some clear outliers near the top of this dendrogram, which are among the utterances that were excluded in data subsetting.

In some cases, such as in the AMOVA analysis presented in Chapter 6, I have combined Burarra and Gunnartpa data sets, as these are both from the same language, despite being deposited in different archives and being recorded in different settings at different times. For the purposes of this phrasal contour analysis, I have not done this. This is only because the phrasal contour script has a file size limit of 5 MB, and combining these two data sets would require first pruning the data extensively myself before input into the script. When needed for other languages, I have done this, but for the Burarra-Gunnartpa case there is a reasonable grouping of data into two groups that I have taken advantage of.

Nine clusters were settled upon, shown in Figure 7.5. Most of these clusters involve a

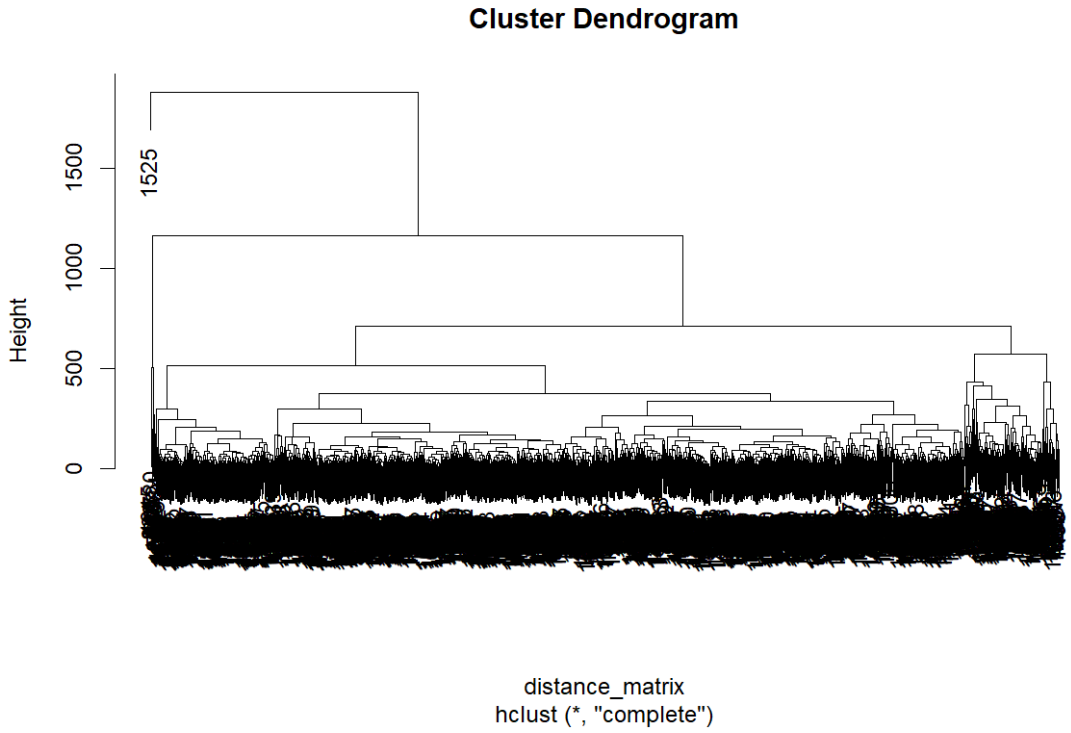

Figure 7.4: Dendrogram for Burarra.

phrase-final decline of differing degrees. Clusters (1) and (2) in Figure 7.5 seem to show a declarative phrase type, with the difference between them being that (1) begins with a slight high tone while (2) begins at about average f0. Cluster (3) shows a flat contour with a small f0 rise at the end of the phrase. Likewise, cluster (6) also has a rise in f0, in this case at the very beginning of the phrase, with f0 remaining stable at an above-average tone for the remainder of the phrase.

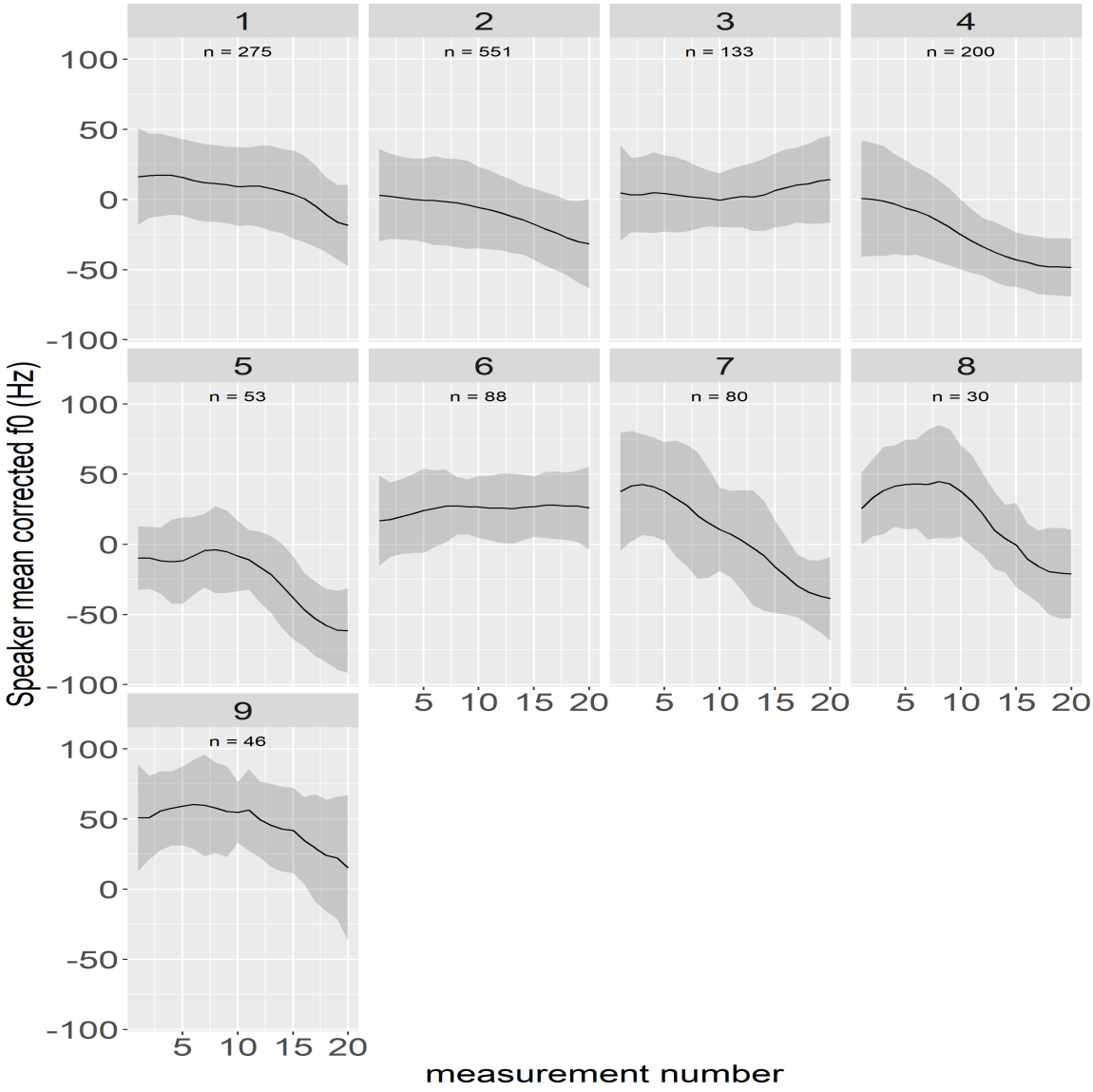

Figure 7.5: Phrasal clusters for Burarra.

The remaining five clusters all show a more dramatic f0 fall at the end of the phrase. Cluster (4) shows a steady decline over the phrase, which begins at about average f0 and ends quite low. Cluster (5), on the other hand, stays at average pitch until about halfway through the phrase, with the last half showing a steep decline in f0. Clusters (7), (8), and (9), in contrast, all begin with high f0. The contour of cluster (9) is similar to that of (5),

with f0 remaining stable until a fall about halfway through the phrase; the distinction here is that (9) begins high and falls to about average f0. Finally, cluster (7) begins with a high tone that falls to a low pitch steadily over the phrase, while cluster (8) shows a high tone that peaks mid-phrase and then similarly falls to a low f0. These clusters may both represent phrases with a focused element at different positions in the phrase.

7.3.3 Gunnartpa

The Gunnartpa data set was smaller than the Burarra data, with only 216 phrasal contours before subsetting and 208 (96.3%) after subsetting. The dendrogram is given in Fig. 7.6.

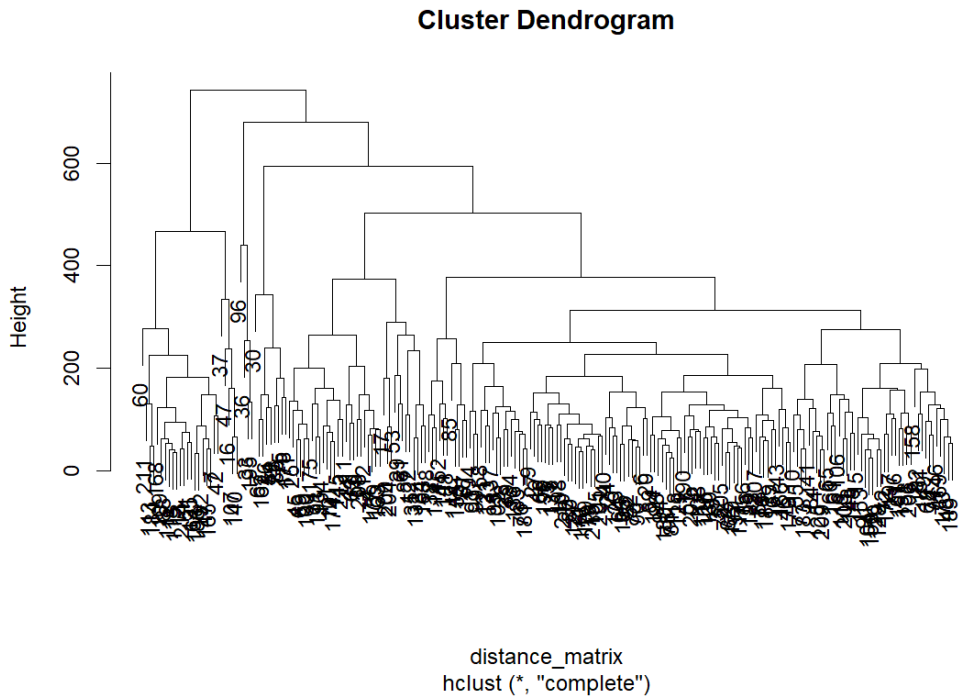

Figure 7.6: Gunnartpa dendrogram.

Eight clusters were identified, shown here in Figure 7.7. These clusters are smaller than the clusters of other languages looked at here, because the overall data set is smaller. The largest cluster is (2), which shows a likely declarative phrase type with about average f0

and a slight decline over the course of the phrase. Cluster (5) is similar, but shows a stable (low) f0 for the first half of the phrase, with a decline after that; this is similar to cluster (5) in the Burarra data set (Fig. 7.5). The Gunnartpa data set also has a small cluster in (6) that shows somewhat steady pitch for half of the phrase followed by the f0 rise. Cluster (7) also has a rise, peaking in the middle of the phrase, with f0 starting and ending at about average on either side. Cluster (8) begins with a high tone, with a fall by about halfway through the phrase and remaining steady for the remainder of the phrase.

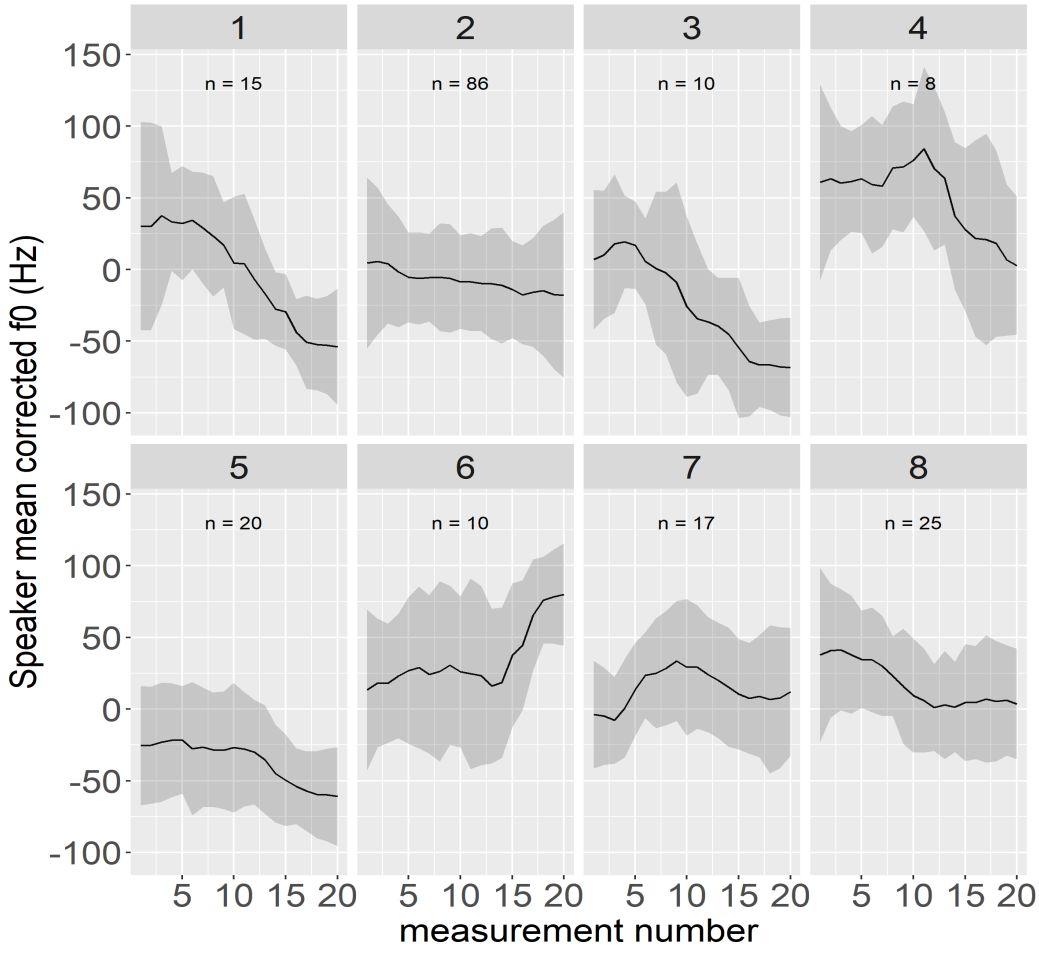

Figure 7.7: Phrasal clusters for Gunnartpa.

The remaining three clusters in Gunnartpa have a pitch fall over the phrase. Cluster (1) in Figure 7.7 begins with a high tone and falls to a low tone at the end, while cluster (3)

begins with about average pitch or perhaps a slight rise, also ending with a low tone. And finally, cluster (4) has high f0 for the first half of the phrase, with a fall to about average f0 at the end.

7.3.4 Dalabon

The Dalabon dataset (cf. §2.2) is very large, and required some pruning before running the cluster analysis script to meet the 5 MB file size maximum. The CSV file generated with Kaland's Praat script was trimmed first for 'jump kill effect' values between 0.9 and 1.1, which is how the data would have been trimmed using the script anyway. Also excluded were single-word phrases; while a single word can be a phonological and syntactic phrase in Dalabon, it is likely not the case for all single-word phrases in the transcriptions used, so it was decided that this was a reasonable group of phrases to exclude for the sake of shrinking the file size. After these exclusions, the resulting data file was 4.997 MB and contour analysis was run.

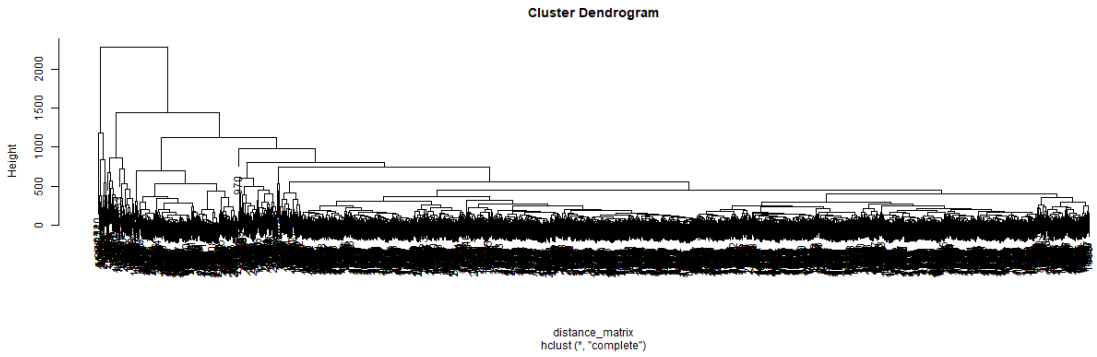

Figure 7.8: Cluster dendrogram for Dalabon.

The uploaded Dalabon data included 2,176 contours, with 2,123 (97.6%) left after subsetting within the GUI script. The dendrogram of clusters, shown in Figure 7.8 is heavily skewed. Because of this, the largest cluster of around $n = 1700$ did not split until the number of clusters was $n = 10$. This major split, clusters (1) and (2) in Figure 7.9, reflects a

contour that starts mid and falls toward the end, and a phrase that remains level throughout, respectively.

The remaining clusters are substantially smaller than these two large clusters, with numbers falling from $n > 800$ to $n < 150$. The next largest cluster is (4), which shows a level, low contour across the phrase. This is very similar to cluster (7), which is also low and level, but with average f0 hovering at around 100 Hz below average in contrast to cluster (4)'s 50 Hz below.

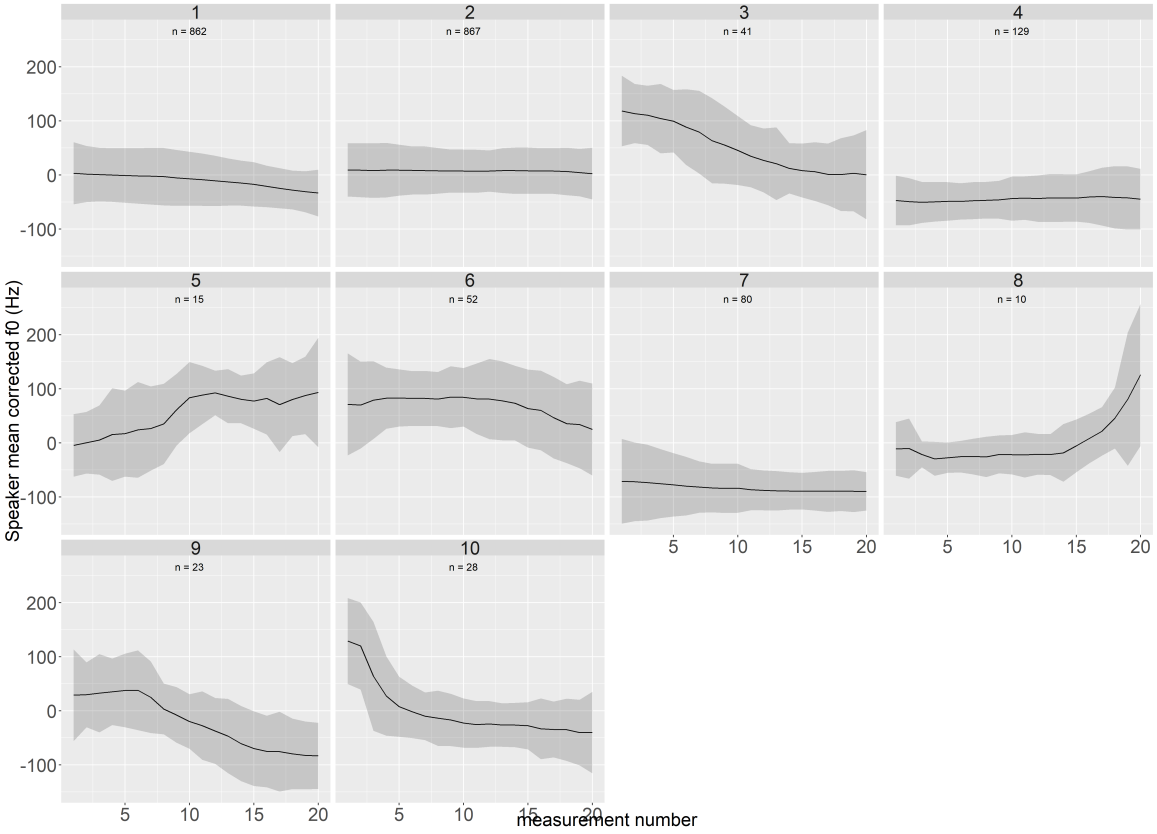

Figure 7.9: Cluster plots for Dalabon where $n = 10$.

Three of the smaller contours— (3), (6), and (10)— begin with high f0. In cluster (10), this high tone falls immediately after phrase onset, likely representing a boundary tone or a pitch reset from a previous phrase. Cluster (3) starts off very high and drops steadily across the phrase, ending at about average f0. And cluster (6) shows a steadier high contour, with

f0 remaining relatively level until a slight lowering at the end of the phrase.

The remaining clusters— (5), (8), and (9)— begin at about average f0. In cluster (9) this average f0 falls steadily starting about one-third of the way into the phrase. Clusters (5) and (8) rise by the end of the phrase, but in (5) this is a more gradual rise, while cluster (8) has a sharp rise right at the end. This latter contour likely reflects an uptick in f0 leading up to the following phrase. It is unlikely that this pitch rise reflects a boundary tone, given that this particular cluster is extremely small at $n = 10$.

7.3.5 Gija

There were 889 phrasal contours in the Gija data, and 864 of these (97.2%) remained after data subsetting. Four clusters were settled on, shown in Figure 7.10. The first two clusters— (1) and (2)— make up the vast majority of the phrases, and both seem to represent a standard declarative phrase type. Cluster (1) begins with about average f0 and falls to a low tone phrase finally. Cluster (2), on the other hand, begins slightly above average f0 and falls to slightly below average, with the overall linear slope being flatter than in cluster (1). Cluster (2) is the larger with $n = 520$. Cluster (1) may represent a stronger low boundary tone indicating the end of a group of phrases or something similar.

Cluster (3) also has a phrase final low tone, but the contour begins with high tone that falls steeply to average f0 about a quarter of the way through the phrase. Cluster (4) shows a pitch rise, with steady f0 at about average for the first three-fourths of the phrase, and a steep rise at the end.

7.3.6 Gunwinggu

The Gunwinggu data is small relative to the other languages here. There are 186 contours in the data, with 149 (80.1%) left after subsetting. Presumably as a consequence of fewer data points, the clusters were smaller and more variable than for other languages in this chapter.

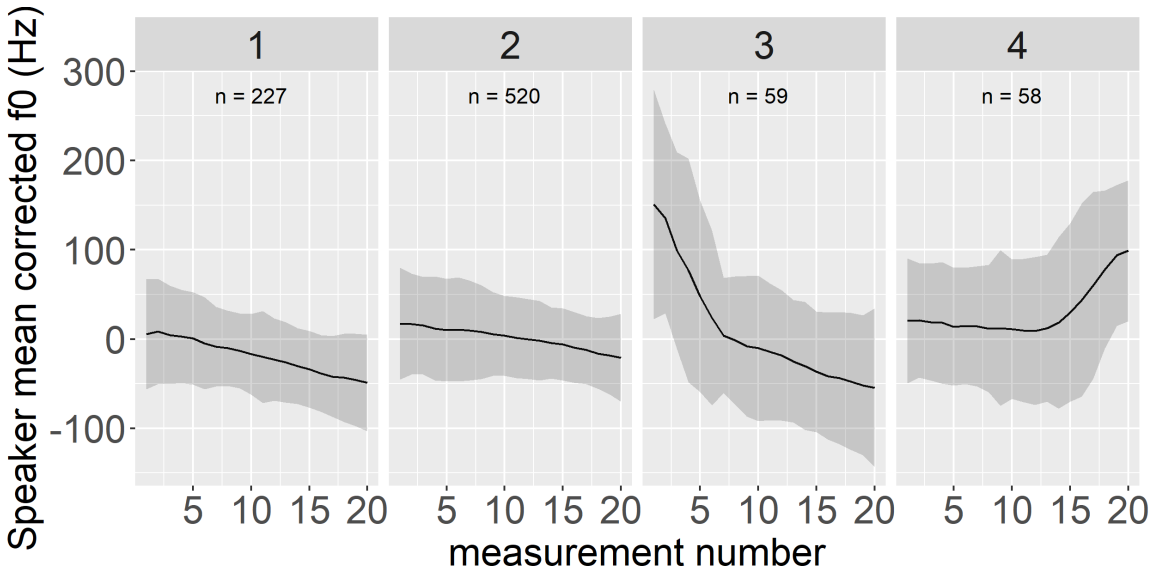

Figure 7.10: Phrasal clusters for Gija.

Four clusters were settled upon, shown in Figure 7.11.

Clusters (1) and (2) are by far the largest clusters and are both declarative type contours. Cluster (1) starts off at average f0 and has a fall at the end of the phrase, while cluster (2) is more stable throughout, with a less extreme fall near the end. Clusters (3) and (4) have much steeper declinations in pitch and also start off with high tones. Cluster (3) seems to maintain the high f0 until the very end, while in cluster (4) this declination is steadier across the phrase. However, these clusters are both quite small ($n < 10$), so it is difficult with so little data to determine if these clusters are truly meaningful and would be more frequent in a larger data set, or if they are simply flukes.

7.3.7 Kayardild

The Kayardild data set is very large and required substantial pre-input subsetting. After upload, there were 2,603 phrase contours in the data, and 2,390 of these (91.8%) were left after in-script subsetting. There were twelve phrase clusters that were determined to be worth discussion, as shown here in Figure 7.12. The largest of these clusters is (1) with

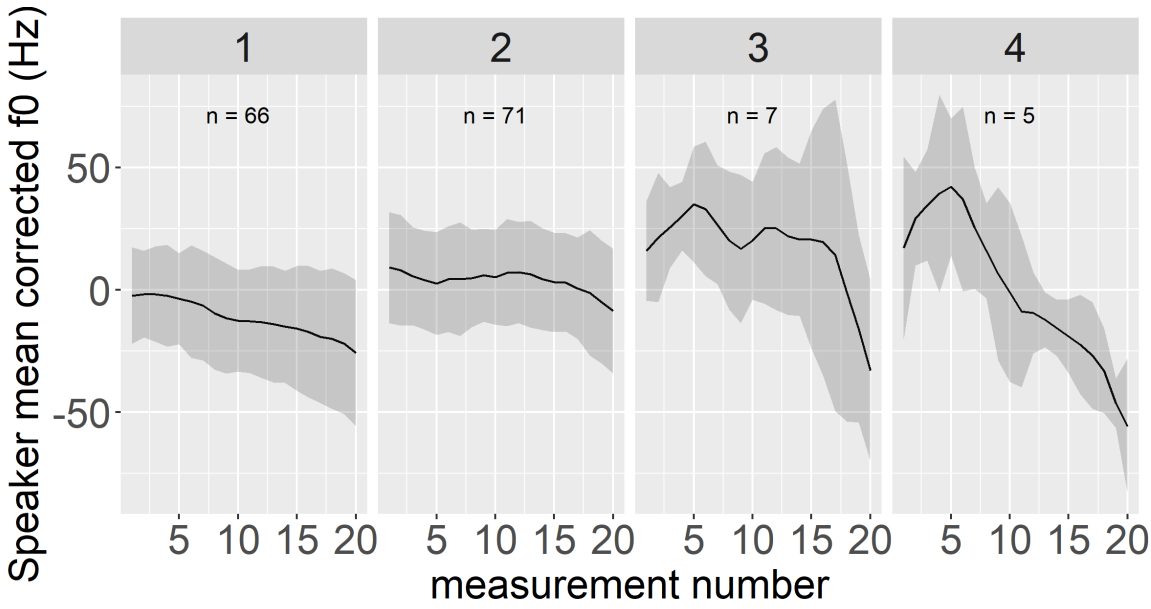

Figure 7.11: Phrasal clusters for Gunwinggu.

$n = 620$, which seems to represent a standard declarative phrase type, with average f0 that falls slightly by the end of the phrase. This is a similar contour to cluster (3), which has a smaller f0 fall over the phrase and also tends to have just below-average f0. Cluster (9) also shows this kind of contour, but f0 is substantially below average here. These contours can likely be grouped together as declarative phrases at different points in the utterance, along with cluster (10), which has steady and high f0 until a steeper fall at the end of the phrase.

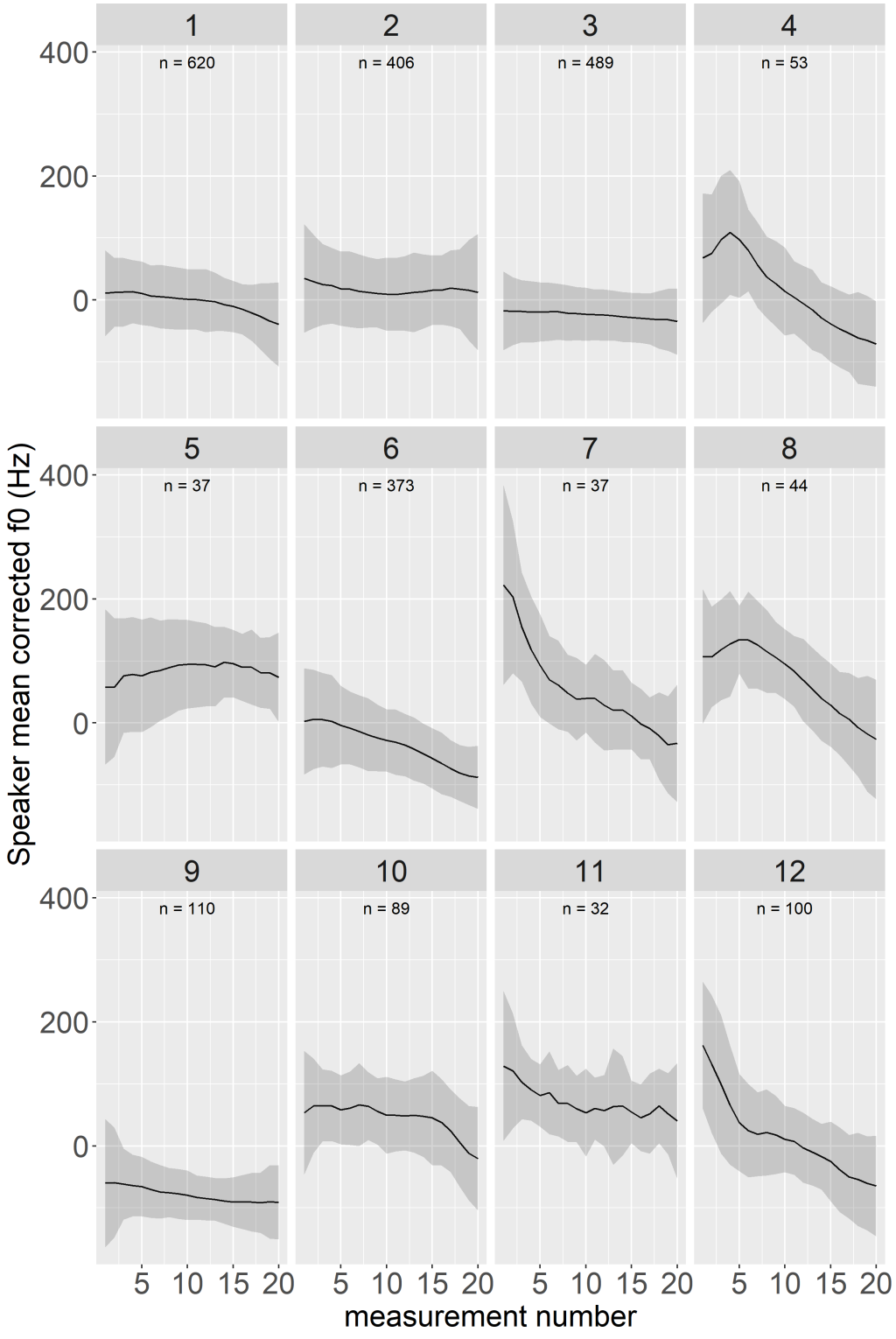

Figure 7.12: Phrasal clusters for Kayardild.

Cluster (2) starts off with a slight high tone that falls to a steady, about-average f0 for the remainder of the phrase. This kind of contour is not seen in the other clusters except perhaps in cluster (11), which has higher f0 overall than (2) and also has a steeper fall at the beginning of the phrase. Cluster (5) is perhaps this inverse of this contour, beginning with relatively lower f0 with a rise, then evening out to steady and high pitch for the remainder of the phrase.

The remaining clusters all show steeper falls phrase finally than the ones already mentioned. Cluster (6) is the largest of these with $n = 373$. The contour begins with average f0 and falls to a low tone by the end of the phrase. The other remaining clusters begin with a higher f0 and end with a low tone, like cluster (7), which falls quite steeply across the phrase. This is similar to the contour in (12), which shows some sort of intermediate low point after the initial fall. This is a substantial cluster with $n = 100$ and would certainly be worth further investigation in Kayardild.

Clusters (4) and (8) both have intermediate high f0 peaks after the start of the phrase. Both then have a fall throughout the phrase, but cluster (4) ends with a substantially lower f0 than cluster (8) does.

7.3.8 Kunbarlang

Kunbarlang had 259 contours in the input data, with 95% (246) left after subsetting. Similarly to Gunwinggu, clusters were quite small because of the small number of contours. The largest cluster in Figure 7.13 is cluster (2), which shows a slightly below-average f0 that declines a bit over the phrase, likely a declarative phrase type. The second-largest cluster is (3), which has a slight rise halfway through the phrase and a subsequent pitch fall at the end. A similar contour exists in cluster (4), with the distinction being that (4) starts off low and falls even lower, while (3) starts off at about average f0.

The remaining three clusters for Kunbarlang are small, but were judged to indicate

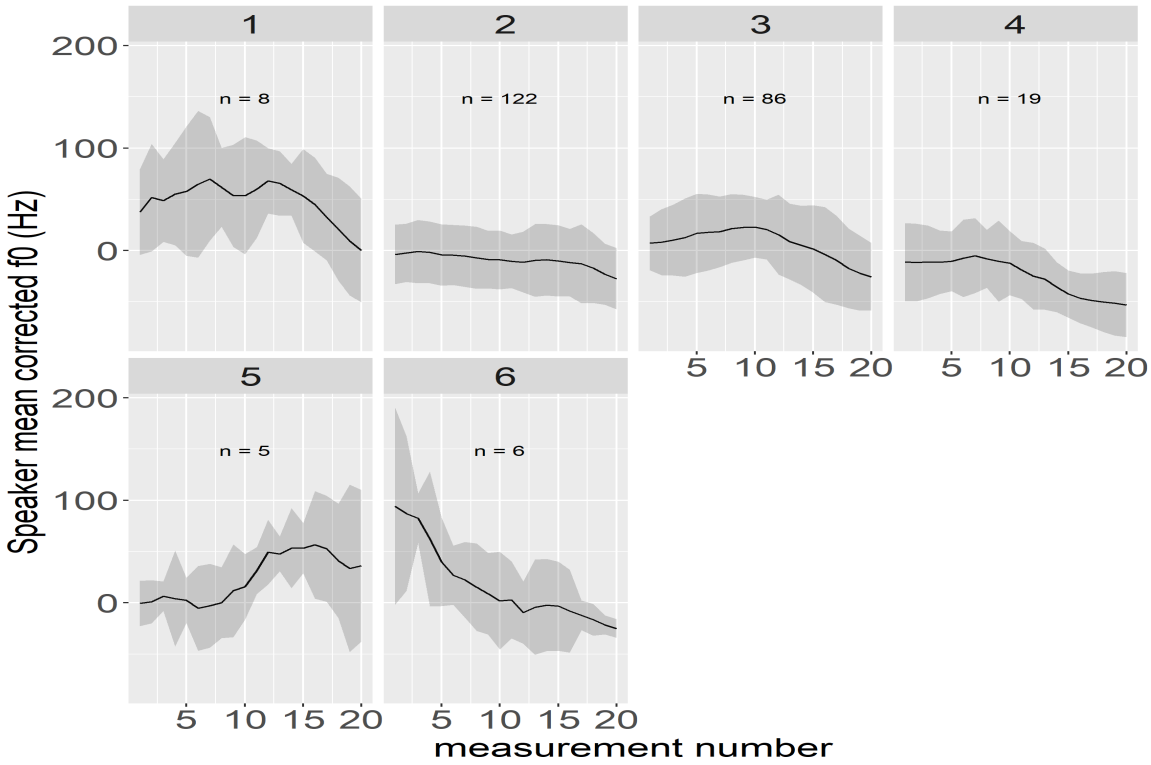

Figure 7.13: Phrasal clusters for Kunbarlang.

distinct contour types that are worth discussing and investigating further in Kunbarlang if the opportunity arises. Cluster (1) shows a similar contour to (3) and (4), with a slight rise followed by a fall, but the f0 level begins quite high and falls to about average. Cluster (6) also shows a fall, in this case from a high tone at the very beginning of the phrase that falls to below-average f0 by the end. Cluster (5) shows the only pitch rise in these clusters, going from average f0 to a high pitch about halfway through the phrase.

7.3.9 Malak Malak

The Malak Malak data consists of 748 contours, most of which (99.1%, $n = 741$) remained after initial subsetting out the small clusters. The dendrogram in Figure 7.14 is quite skewed near the root, but begins to show more evenly split clusters lower down in the structure. Even so, almost half of the phrases (49.3%, $n = 365$) are contained in one cluster, Cluster

1 in Figure 7.15. This large cluster did not break up until the number of clusters was quite high, at which point there were more small clusters than I judged to be reasonable. For this reason, the number of clusters I settled on for Malak Malak was $n = 6$.

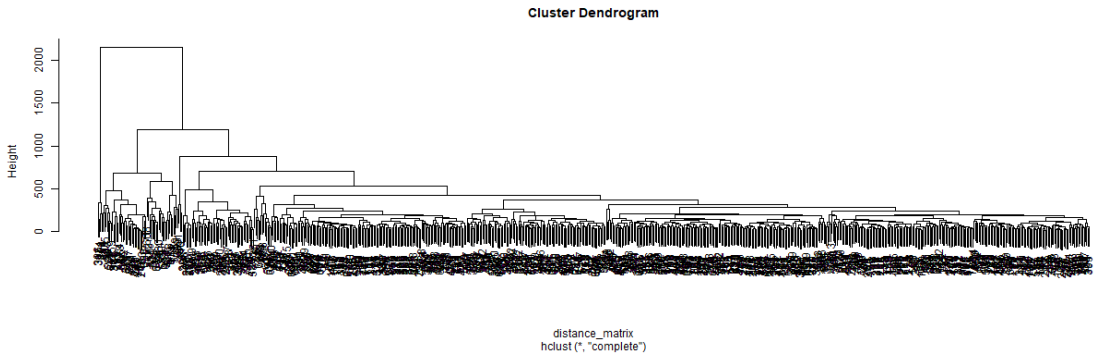

Figure 7.14: Cluster dendrogram for Malak Malak.

The largest cluster in Figure 7.15, Cluster (1), shows a steady pitch level at around average f0 across the phrase, with only a slight lowering toward the end of the phrase. Cluster (3) is about half the size of (1), and it shows a contour that starts off with level f0 as the former does, but with a sharper lowering of pitch in the last fourth of the phrase. This cluster may consist of phrases with final low boundary tones, in contrast to Cluster (1) which does not have this boundary tone. Likewise, Cluster (2) in Malak Malak starts the contour with high f0 that falls steadily to the end of the phrase, which may indicate an initial boundary tone.

The remaining Clusters (4), (5), and (6) are much smaller than the first three ($n = 16$, $n = 21$, and $n = 31$, respectively). However, these were chosen to be included in the final cluster results because of their distinctness from the clusters already discussed. Cluster (4), for instance, shows a contour with low f0 throughout, with a slight fall toward the end of the phrase. This type of contour may indicate a particular discourse function or something similar. Cluster (5) shows a phrase type with a rising intonation. And Cluster (4), the smallest here, seems more variable than the others, but seems to be clustering together a

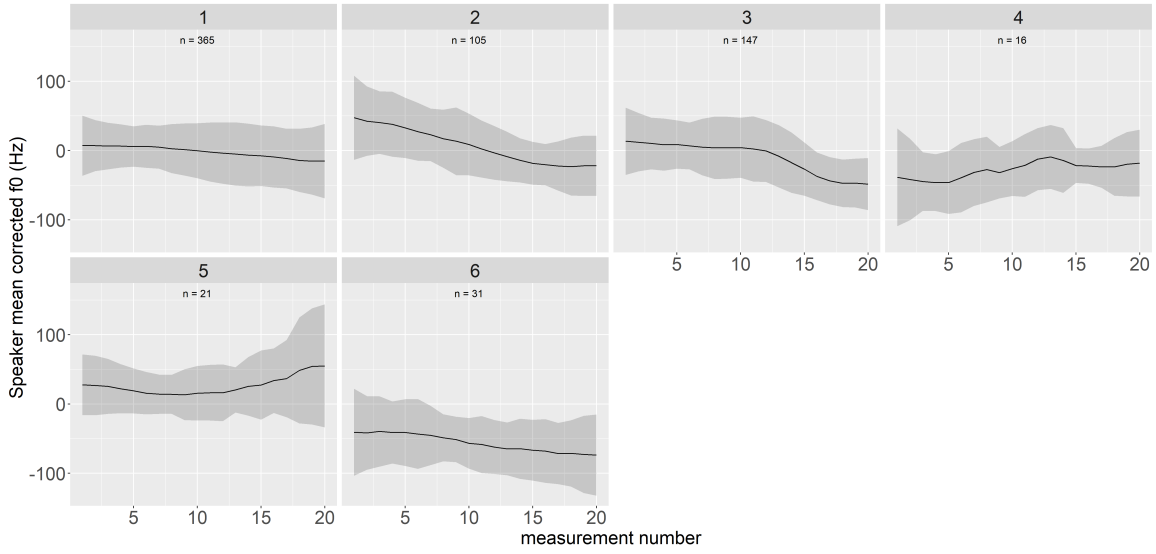

Figure 7.15: Cluster analysis for Malak Malak ($n = 6$).

group of phrases that have low tone that rises at some non-initial point in the phrase— in some cases this seems to be medial, in other cases final.

7.3.10 Murrinh Patha

There were 257 contours in the Murrinh Patha data, with 242 (94.2%) left after subsetting. The dendrogram, shown in Figure 7.16, is perhaps the most evenly split that we have seen so far. Apart from some clear outliers that were excluded via data subsetting, there is a close-to-even split near the top of the structure. It is possible that Murrinh Patha speakers happen to use a wider variety of phrase types in their speech, but I do not think this is the most likely explanation. The Murrinh Patha audio data from the PARADISEC deposit used here was collected as part of a structured task for the Social Cognition Project, which is geared toward "gathering enriched language data for descriptive, comparative and documentary purposes" (Barth 2009). This involves speech situations like retelling of stories and conversational problem-solving, but the content, and therefore the types of phrases, are influenced by the task.

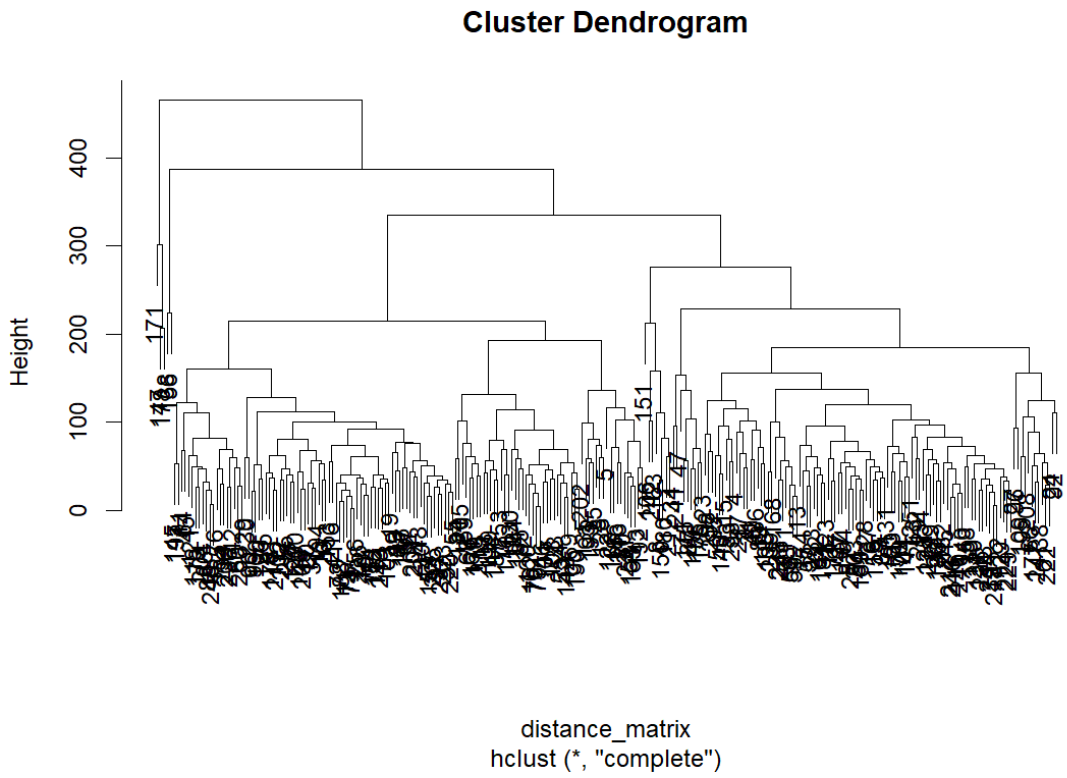

Figure 7.16: Dendrogram for Murrinh Patha.

There are phrasal contours with both rises and falls in the clustering results shown in Figure 7.17. Clusters (2) and (4) seem most like to be declarative phrase types, as they are the most frequent and have the least dramatic pitch excursions. Cluster (2) has low pitch throughout, with a slight rise in the middle of the phrase that falls again at the end. Cluster (4), on the other hand, holds steady at about average f0 until the end, where there is a fall.

Clusters (1) and (3) also have a fall, but both start off with slightly above-average pitch. Cluster (1) seems like it may have two high f0 peaks through the phrase, ending with a fall. Cluster (3) has a steady fall, starting with above-average f0 and falling to a low pitch at the end.

Pitch rises are seen in the remaining two clusters. Cluster (5) seems to start at average pitch, with a low tone close to the beginning of the phrase. This then rises at the end of

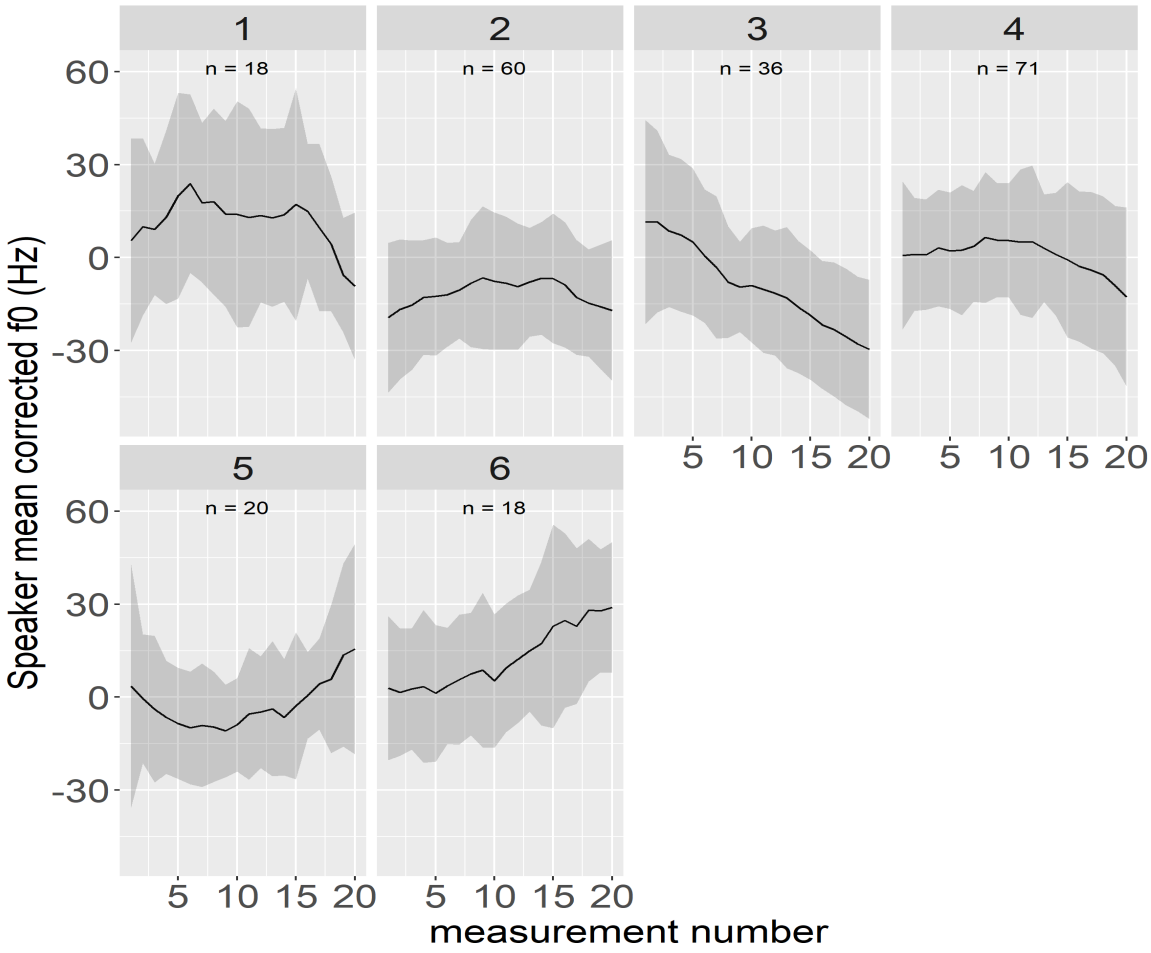

Figure 7.17: Phrasal clusters for Murrinh Patha.

the phrase to slightly above average f0. Cluster (6) does not have this low tone, showing a steadier rise from average f0 at the beginning to the high tone at the end of the phrase.

The data for Murrinh Patha, while still spontaneous speech, was structured more closely than the other language data included here, and this has clearly impacted the structure of the clustering results in this prosodic analysis. This makes a case for structuring data collection for phrasal phenomena in order to see a variety of phrase types more clearly; however, as the other results in this chapter demonstrate, the bottom-up clustering method can and does detect different phrase types in more skewed data, as long as the researcher is aware that the skew is expected.

7.3.11 Ngan'gi

The Ngan'gi data set was quite large and needed to be trimmed before inputting into the clustering script. To reduce the file size to 5 MB, the 'jump kill effect' cutoff was changed from 10% (0.9-1.1) to 8% (0.8-1.08). After pre-input trimming, there were 2,526 contours uploaded. After subsetting smaller clusters out, 91.6% (2,315) remained.

Ten clusters were settled on (see Figure 7.18). The vast majority of phrases ($n = 1198$) have a standard declarative-type contour in (1), beginning a little above average f0 and falling throughout the phrase to end a little below average. The second-largest cluster is (2) with $n = 357$, and this contour shows a slight rising intonation over the phrase, beginning at average f0 and only rising a little by the end. Because of the higher numbers in this cluster relative to the remaining eight, it seems likely that this is a major phrase type contour, such as a question, or introducing a new topic in the discourse.

Most of the remaining clusters in Fig. 7.18 end with a falling tone: (3), (4), (5), (6), (8), and (10). Clusters (3) and (8) show a steady decline over the phrase, but in (3) the contour starts off high and falls to slightly below average f0, while in (8) pitch starts off about average and falls quite low. Clusters (4), (5), and (6) all have some medial f0 peak in the phrase, followed by a pitch fall. Cluster (6) has the earliest of these peaks, right at the beginning of the phrase. Cluster (5)'s medial peak is a bit later, and cluster (4) has a peak right in the middle of the phrase. These peaks may represent focal words at different positions in the phrase, or similar sorts of contours at the beginning for phrases of differing lengths, since duration is normalized for the purposes of this clustering. Cluster (10) also has some sort of medial rise in f0, but it is not altogether clear that this is the same kind of contour as these other three.

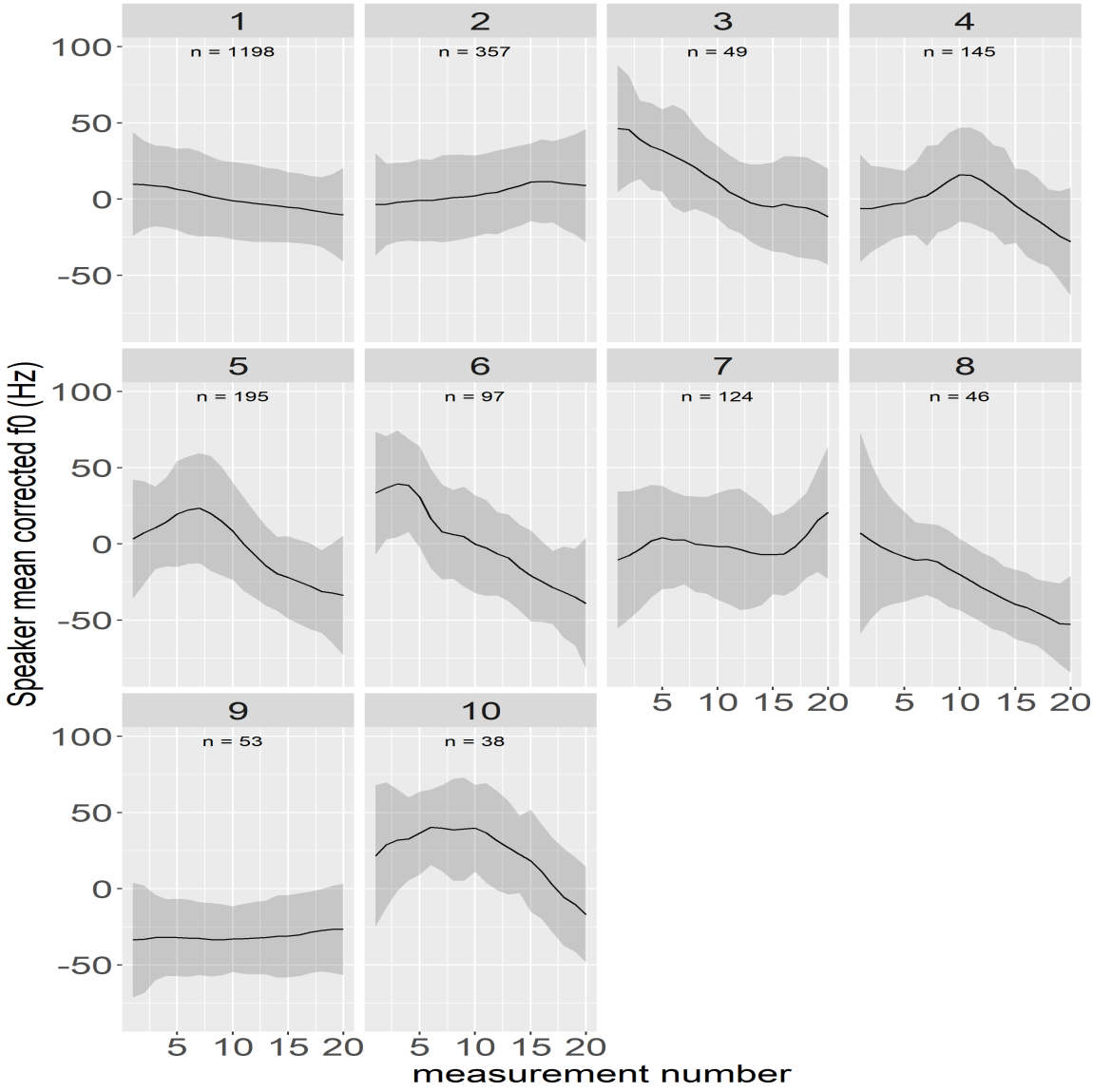

Figure 7.18: Phrasal clusters for Ngan'gi.

7.3.12 Wanyjirra

There were 230 contours in the Wanyjirra data, and 89.6% (206) were left after subsetting. Six clusters were settled upon, shown here in Figure 7.19.

Cluster (1) is the largest here with $n = 57$, and it shows a basic declarative phrase type

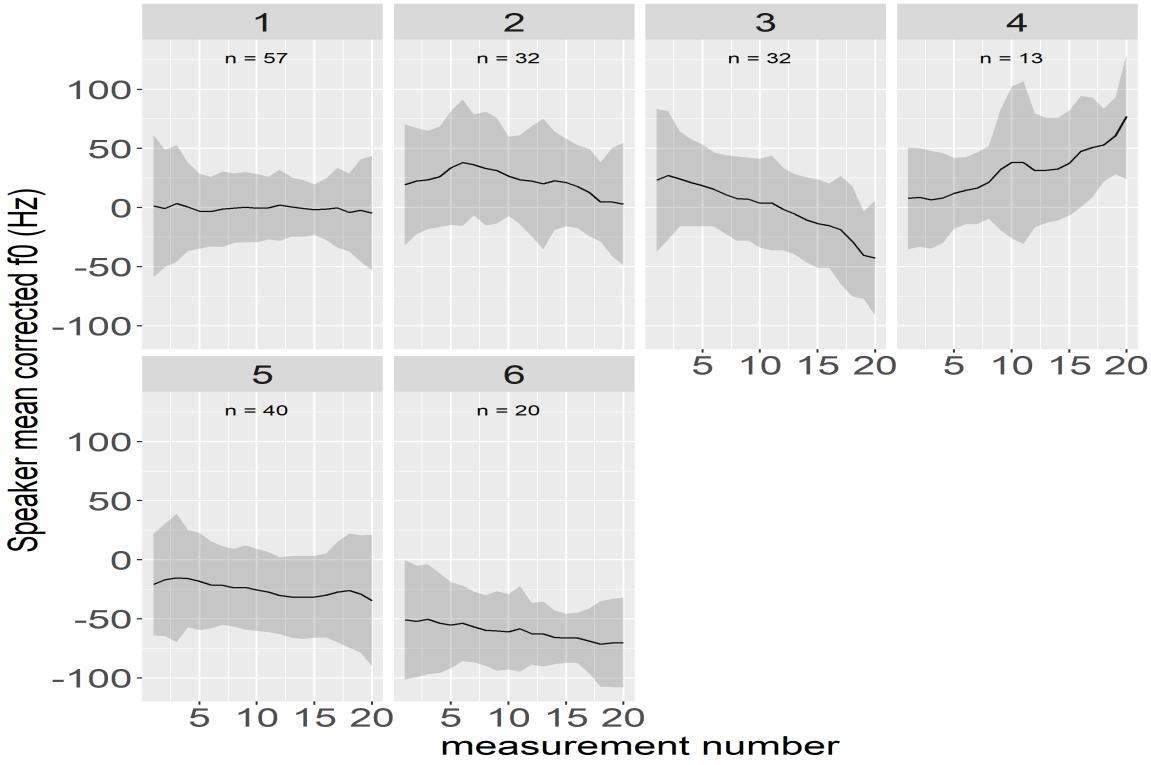

Figure 7.19: Phrasal clusters for Wanyjirra.

with generally steady and average f0 across the entire phrase. Clusters (5) and (6) also show likely declarative phrases, both with low starting f0 which may indicate that these phrases occur later on in the course of an utterance. Cluster (5) starts off just below average f0, with a slight fall, while Cluster (6) begins with quite low f0 and falls even lower throughout.

Clusters (2) and (3) also have a phrase-final fall, but both start off with above-average f0. Cluster (2) seems to have a pitch rise after the beginning of the phrase, with f0 falling back to about average by the end. Cluster (3), on the other hand, starts off with above average f0, and falls steadily through the phrase, ending with a low tone.

Finally, cluster (4) is the one contour that has a final pitch rise. Pitch starts off at about average, and rises throughout, with perhaps a slight rise in the middle of the phrase. This is the smallest of the clusters with $n = 13$, but the distinctiveness of the contour relative to the others in Fig. 7.19 is likely to indicate a specific phrase type in Wanyjirra.

7.3.13 Warlpiri

There were 250 contours in the uploaded Warlpiri data, with 235 (94%) left after subsetting. Five clusters were settled upon, as shown in Figure 7.20. Phrase types were skewed in the expected way based on the type of speech, as is reflected in the skewed cluster sizes here.

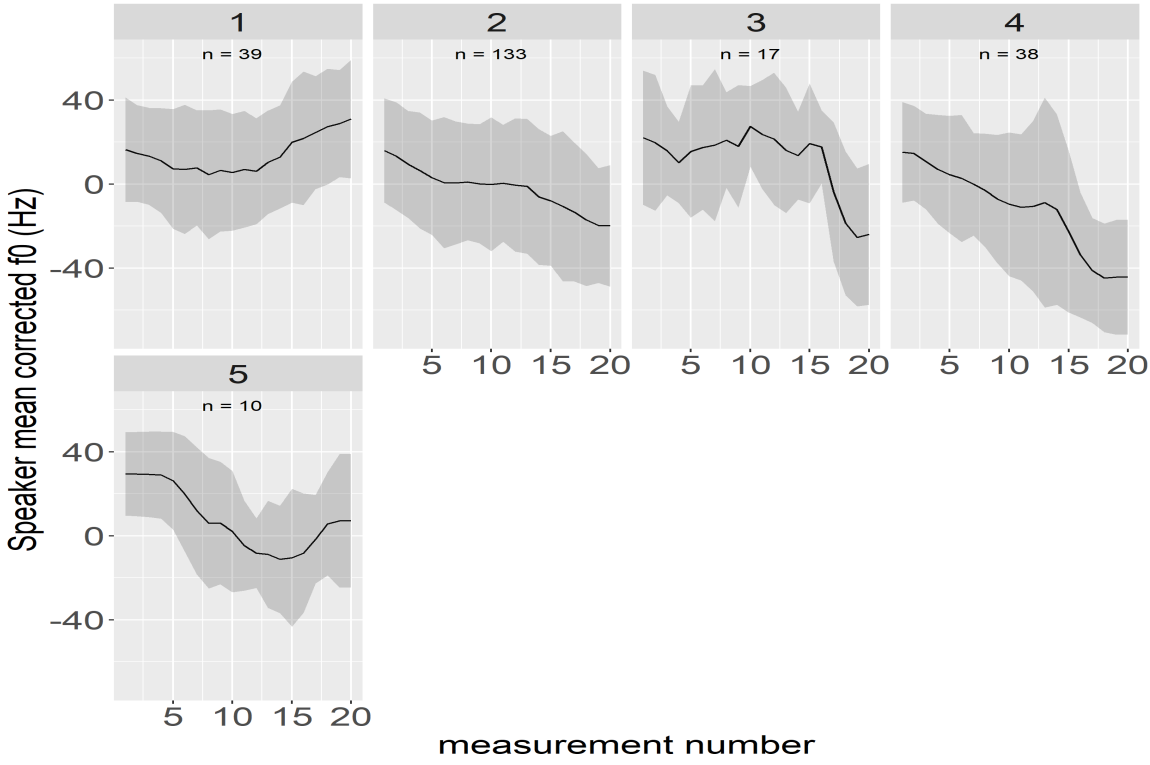

Figure 7.20: Phrasal clusters for Warlpiri.

The largest cluster is (2) with $n = 133$. This contour seems to be a declarative phrase type, which begins with f0 slightly above average and has a fall to low f0 by the end of the phrase. Cluster (4) also has a falling f0 like this, but the pitch excursion is larger and the phrase ends with much lower f0 than this larger cluster. Cluster (4) also seems to have a pitch rise right before this phrase-final fall.

Cluster (3) also has a phrase-final fall, which comes after a somewhat steady high tone for the first three-fourths of the phrase (although note the variability in the contour here).

The last fourth of the phrase has a steep pitch fall from above average to low f0.

Clusters (1) and (5) have phrase-final rises. In cluster (1), pitch starts off at above average, followed by a slight fall in the middle, ending with quite high f0. Cluster (5), on the other hand, starts off with high f0. This is followed by a low tone about three-fourths through the phrase, ending with a rise that ends the contour at about average f0. While this is a smaller cluster with $n = 10$, the contour is distinct enough that I judged it worth discussing here.

7.3.14 Warnman

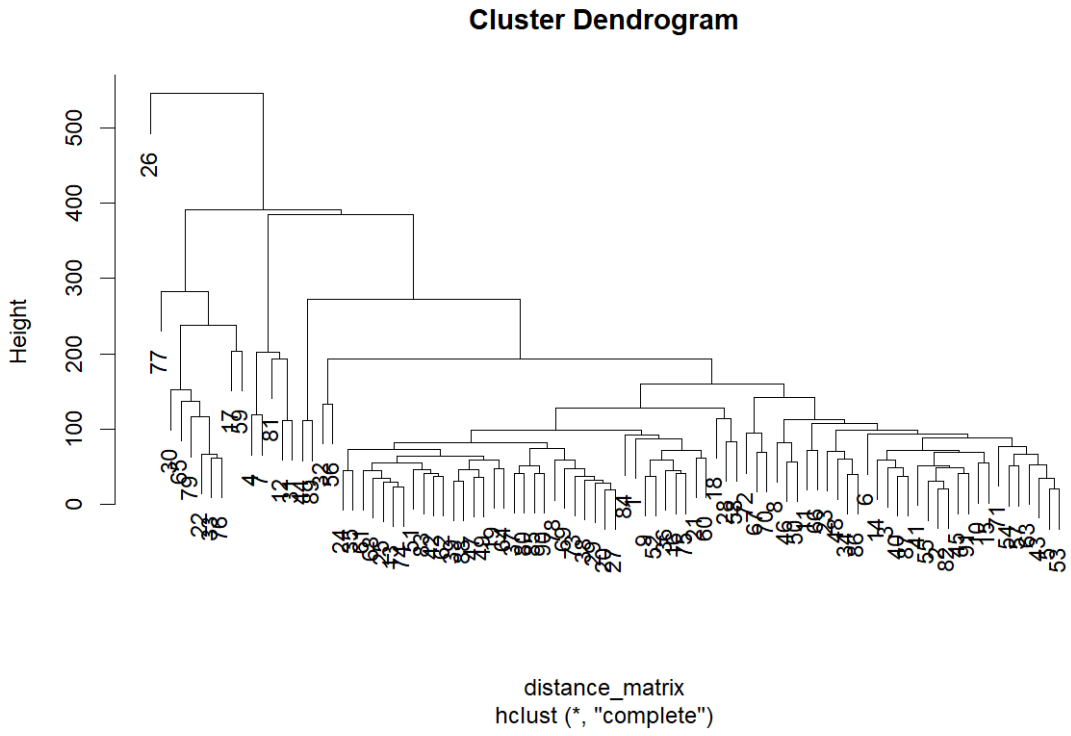

Figure 7.21: Dendrogram for Warnman phrase contours.

There were only 91 phrases in the uploaded Warnman data, the smallest data set of these languages. This is less of an issue when looking at word-level phenomena, as 91 phrases still amounts to hundreds of words, but for this clustering analysis the number is

likely not sufficient. The clustering algorithm did not work well on such a small data set, leaving many very small clusters that were excluded using the script's automatic subsetting. Only 65.9% of the uploaded contours (60 phrases) remained after data subsetting. These small clusters can be seen fairly clearly in the dendrogram in Figure 7.21, which has many outliers and small cluster groups.

Because of this small set of data, the results for Warnman are unlikely to be reliable; for thoroughness, two Warnman phrasal clusters are given in Figure 7.22. Both of these contours seem to show a relatively steady f0 at about average level throughout the phrase. It seems from this small amount of data that both of these clusters represent the declarative sentences, which we expect to be most frequent in natural speech data.

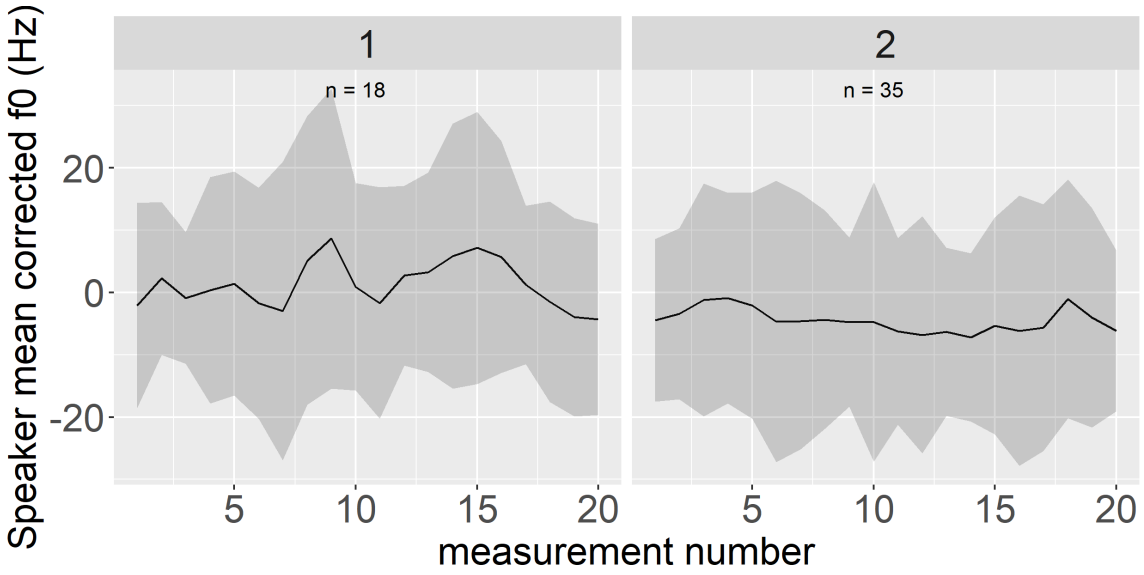

Figure 7.22: Two Warnman contour clusters.

7.3.15 Yannhangu

Yannhangu had 623 contours identified by Kaland's Praat script, and 91% ($n = 567$) were left after the initial subsetting of small clusters. The dendrogram shows the expected skew toward declarative phrase types (Figure 7.23).

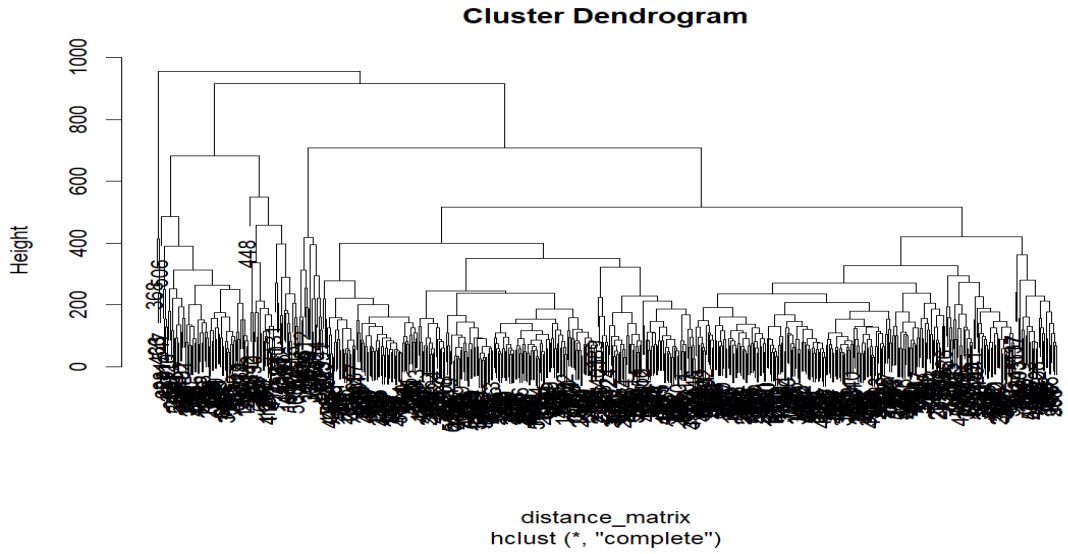

Figure 7.23: Dendrogram for Yannhangu phrase contours.

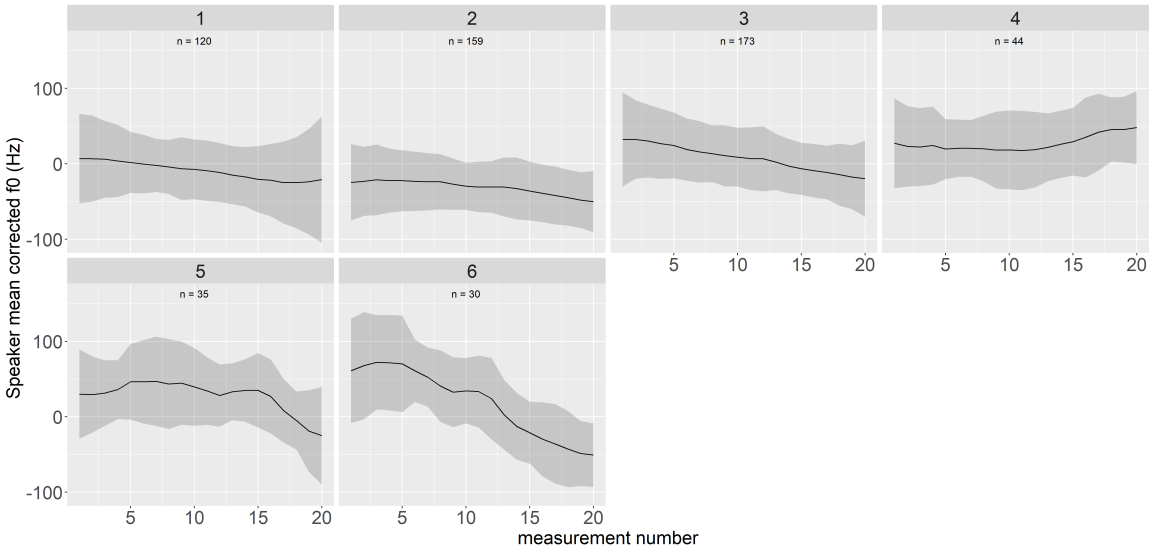

Figure 7.24: Cluster analysis for Yannhangu (n=6).

Six clusters was determined to be the most informative number of contour types; these are shown in Figure 7.24. Clusters (1) through (4) are fairly similar, with a few differences between them. Cluster (1) starts at average pitch and declines a little across the phrase.

Cluster (2) starts a bit below average but then has a contour similar to cluster (1). Cluster (3) starts with a high tone and declines through the phrase, while Cluster (4) starts near average or a little higher, and then raises at the end.

Clusters (5) and (6) show more dramatic pitch excursions. Cluster (5) captures phrase types that may be a bit variable in the first part of the phrase, sometimes with a higher pitch about one quarter through, and a steep decline at the end. Cluster (6), on the other hand, begins with a high tone and also has a steep fall at the end of the phrase.

7.3.16 Yidiny

There were 659 contours in the uploaded Yidiny data, with 91.2% of these (601) left after subsetting. There were five clusters settled on, as shown in Figure 7.25.

All but one of these clusters shows a falling intonation. The largest cluster is (2) with $n = 268$; this cluster starts off with about average f0 and falls steadily throughout the phrase. The same basic contour is seen in clusters (4) and (5) as well, but cluster (4) starts off with slightly below average f0, and cluster (5) with quite low f0. We can hypothesize that these all represent the basic declarative phrase type, but at different points in the narrative.

Cluster (1) also has a phrase-final falling tone, but the pitch throughout the phrase seems to remain almost stable at above-average f0 for most of the phrase, only falling toward the end, so this seems to be a different contour than the ones already discussed.

Finally, cluster (3) shows a rising tone. This phrase contour begins at just below-average f0, rising throughout the phrase and ending slightly above average. It is likely that this is a different phrase type, such as an interrogative, in contrast to the other phrase contours picked out by the clustering algorithm.

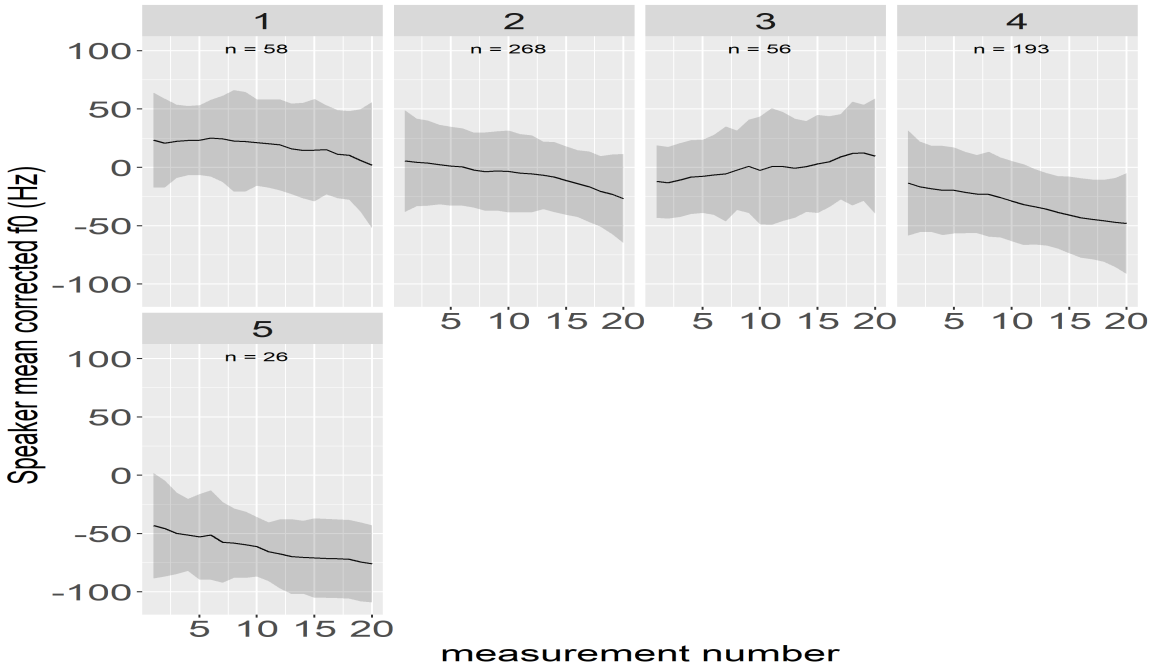

Figure 7.25: Phrasal clusters for Yidiny.

7.4 Summary & Discussion

Phrasal prosody is a complex subject that requires close investigation into a language's narrative and discourse structures. While a detailed investigation into phrasal prosody in these languages lies beyond the scope of this dissertation, the preliminary analysis using automated clustering methods just presented provides an idea of the types of phrasal contours that make up each language's data, and provides a test case for using the methods put forth by Kaland (2021) to establish initial hypotheses about phrase types.

One commonality across all of these languages is that the largest proportion of the data consisted of what looks like a declarative phrase type, with a phrase-final fall. In some cases, multiple clusters of this type were identified, differentiated by the beginning f0— this is likely because of a general pitch declination over the course of a larger utterance or breath group. This phrase type was by far the most frequent across all of these languages, even in Murrinh Patha, where the data collection methods were more structured and balanced in

phrase types than the others. These results suggest perhaps a reassuring fact about studying phonetics with spontaneous speech data: the vast majority of phrases are likely to be the basic, declarative type, because of the nature of narrative speech as being mostly declarative storytelling and informing.

Other phrase types, such as those indicating questions or introducing new topics, are less frequent than declarative phrases but are still picked up in the contour clustering analysis. The use conditions and detailed descriptions of these contours especially requires follow-up investigations into the language by someone who has detailed knowledge of the language, ideally a native speaker or at least an experienced researcher and/or learner of the language. These are the cases where the automated analysis of phrasal contours can serve as a source of initial hypotheses about phrase types, which can then be followed up with more structured data collection to test those hypotheses. Conversely, a language that already has some prosodic analysis done in the traditional way, using a combination of impressionistic hypotheses and native speaker intuitions, can make use of this automatic analysis to look at frequencies of these phrase types in different types of speech, and potentially find less-common phrase types that are not as readily available to the researcher's ear or the native speaker's intuitions.

The results presented here add important information about these languages' prosodic structure and speak to the typological questions posed by Fletcher, Evans & Round (2002), who identified a distinction among Australian languages that use phrasal prosody mainly to demarcate boundaries versus those that use phrasal prosody for other focus-marking functions and thus have pitch events at various locations within the phrase rather than only at the left or right edges. The most frequent phrase types did in fact show boundary-marking pitch events— mainly final falls or rises, and initial high tones. In some of these languages the boundary-marking phrase types are the only ones detected by the clustering algorithm. However, in other languages (see as an example Ngan'gi) pitch peaks seem to occur at

more varying positions within the phrase, suggesting that prosodic contours are serving more than just a demarcative function in these languages. Further work is still needed to synthesize these observations of prosodic variation into a comprehensive typology of Australian phrasal prosody, but this chapter certainly supports the claim that phrasal prosody is not uniform across the continent.

Chapter 8

Archival phonetics

Traditional phonetics training focuses heavily on the availability of highly controlled experiments and quiet recording locations, but such settings are difficult, if not impossible, to come by when working in the field. Recordings collected in the field are inherently noisier than those made in a phonetics booth; fieldwork materials tend to consist of natural speech rather than careful experimental setups, and field locations are often outdoors or in other environments with a lot of background noise. Increasingly, however, linguists have had success using field recordings to investigate a variety of phonetic questions. Recent work using archival field recordings has investigated a variety of acoustic measures, including duration-based contrasts such as voice onset time and phonemic vowel length (A. C. Yu 2008, DiCanio & Whalen 2015, Lawyer 2015, DiCanio et al. 2015); f0 contrasts such as tone, stress, and other prosodic phenomena (Tuttle 2003, Coombs 2013, Gordon 2015, Babinski 2021b); vowel space and quality (Blankenship 2002, de Carvalho 2010, Esposito 2010, Keating et al. 2010, Garellek & Keating 2011); among others (Kakadelis 2018, Tang & Bennett 2018, Hall et al. 2019, Whalen & McDonough 2019). While using noisy audio and/or natural speech recordings may yield somewhat different results than highly controlled lab-collected data, both data collection methods come with a set of potential bi-

ases. Fieldwork materials often show higher variation because of the collection setting, but lab-collected materials might show too little variation resulting from unnatural, overly formal speech production. Despite these biases, neither situation should preclude the use of such data to draw meaningful conclusions (cf. Beckman 1997, Xu 2011, Anand, Chung & Wagers 2011).

This chapter provides a discussion of the use of archival materials in phonetics, in particular materials consisting of mostly spontaneous speech and collected in a wide variety of field conditions. The skills involved in accessing and processing materials available in language archives are important and will only become more in demand as language archives grow and languages cease to be spoken. What follows is meant to serve as a practical guide for those looking to do archival phonetic work; its structure is based in part on a symposium talk given at the Linguistic Society of America Annual Meeting in January 2022 (Babinski 2022).

The entirety of this dissertation could be considered a case study in using archival materials for research on prosody, phonetics, typology, etc. This chapter boils down the most relevant methodological points, with reference to other areas of this work that provide more specific detail. This dissertation project has required data processing methods to mitigate the effects of background noise and differences in recording setup on relevant phonetic information. The variation in the state of these materials reflects the often-uncontrollable nature of recording audio in the field, but careful collection practices combined with post-hoc digital signal processing techniques can mitigate data loss substantially.

8.1 Background

Naturalistic speech and experimental speech obviously differ along many dimensions. However, both types of data provide crucial information about language. Cohn & Renwick (2021) emphasize the complementarity of these two types of data for approaching a research

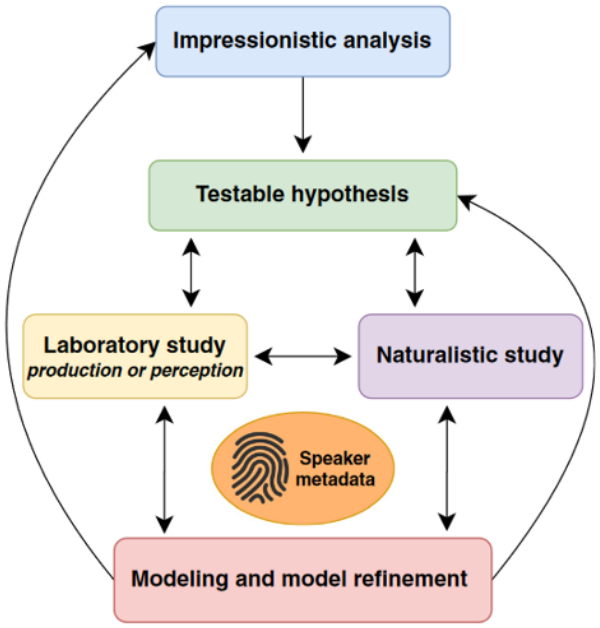

Figure 1: Proposed methodological framework for phonological study and modeling; arrows indicate directions of feedback between studies and methods.

Figure 8.1: Schematic of aspects of phonological study from Cohn & Renwick (2021: 103).

question. The schematic in Figure 8.1 shows the aspects that go into understanding of any phonological phenomenon. Study of naturalistic speech data, which has been the focus of this dissertation, can serve as input to determining testable hypotheses, constructing experimental studies, and creating theoretical models; likewise, all of these aspects of research can themselves serve as input to a naturalistic study. This is all to say that one should not shy away from using natural speech data for phonetic and phonological research, as this is a crucial component to full understanding of a linguistic phenomenon and can yield important results that lead to testable hypotheses, follow-up experiments, or even full theoretical models.

8.1.1 Two types of 'noise'

It is important for this discussion to distinguish between two types of what we might call 'noise' in natural speech audio. One of these is background noise, while the other may

more accurately be called variability in linguistic conditions. The first of these, background noise, is the kind of extraneous sound that recording devices pick up on but that obscure the linguistic signal and that we want to eliminate altogether. While of course making clean recordings without background noise is the ideal, it is often unavoidable especially in field settings. Solutions for dealing with this type of noise require methods of cleaning up our recordings so that the linguistic signal is clearer.

The other type of noise, variability in linguistic conditions, results from the inherent multidimensionality of language and will be present in any type of natural speech. Experimental conditions are usually constructed in order to keep all other linguistic phenomena the same, only varying the factor(s) that are of research interest for the study. When working with natural narrative and conversational speech recordings, utterances are not controlled in this way. Methods for dealing with this type of noise include isolating the linguistic phenomena that we want to look at through data normalization and statistical methods.

8.1.2 Archives and endangered languages

Establishing best practices for the use of archival language data is especially relevant for work with endangered languages, for which any source of audio recording is highly valuable and for which collecting new audio is much more difficult than it is for well-resourced global languages. Additionally, languages that are no longer spoken may be documented only via archival materials, as is the case for many of the languages in this dissertation.

Since it is the focus of this dissertation, research in phonetic typology is the focus of this guide. However, it is important to note that these same methods are extremely useful in the creation of language learning materials in language reclamation projects. Specifically, the creation of word- and segment-aligned transcriptions can serve as the basis for language textbooks and dictionaries, and utterance-aligned transcripts can be used for the creation of multimedia stories and oral histories. One example of archival materials being used in this

way is the Kullili Dictionary mobile app, which is based on archival language recordings from Kullili elders in the 1960s or earlier. This language is no longer spoken and these resources are being used as a part of a broader language reclamation effort.

8.2 Data acquisition

The gold standard for making good language recordings requires a high-quality recorder such as the Zoom H4n; a quiet and controlled recording location; and strict control over the speaker's distance from the microphone. This will produce maximally consistent and quality recordings. However, these conditions can really only be met when one has access to lab-like conditions in a university building or something similar. Collecting linguistic data in a fieldwork setting rarely lends itself to the highly controlled, lab-based settings of data collection at an academic institution such as sound-dampening booths or otherwise quiet locations. Furthermore, travelling to a field location may not be possible, whether because of a lack of funding or health concerns, especially in light of the recent COVID-19 pandemic and the tendency for endangered language communities to be skewed toward an older age demographic that may be at increased risk.

Data acquisition from non-lab based sources present their own caveats, but none of the sources discussed here should be excluded as a potential source of data as long as the proper considerations and controls are taken into account. Recording audio remotely, for example, prevents the use of high-quality recorders such as the Zoom H4n. However, recent work has tested the quality of various remote recording methods and laid out expectations in terms of the accuracy of the acoustic measurements from such recordings. Furthermore, this may be a good option when travel to a field location is not possible, especially if one already has existing community relationships. A working remote-recording protocol with community contacts can serve as a way of collecting recordings that is efficient in terms of time, funding, carbon footprint, and potential health risks. Remote recording is discussed

in §8.2.1.

Another potential data source for language recordings is the language archive, which is discussed in §8.2.2 below. The deposits made by language researchers over the years to these repositories can be a great resource especially for those interested in comparative and typological language work, since audio from many languages can be accessed relatively quickly. The recordings available from these sources potentially have the same problems as data collected in the field oneself, except that the recording conditions cannot be controlled for and cannot be changed, since all the recordings have already been made. Thus it is even more crucial to make use of post-hoc digital signal processing methods to clean and control the phonetic signal in these data.

While most of this discussion focuses on acquiring audio materials for academic research purposes, the same methods can be hugely beneficial in language reclamation projects as well. Using one of these methods of data collection can save time and be more cost effective for those working within language communities to create digital dictionaries, textbooks, storybooks, and mobile language apps. Remote recording is an especially useful tool for these purposes, as fine-grained phonetic reliability may not be as much of a priority for these sorts of projects.

8.2.1 Remote recording methods

Recording language remotely can be a cost-saving way of conducting fieldwork, and it is especially relevant in the wake of the COVID-19 pandemic that brought on widespread travel restrictions and health concerns for those hoping to work with language communities in remote locations or with a large proportion of older speakers. However, this usually means that recordings will be made on a phone, tablet, or computer, rather than with gold standard recording equipment, a particular concern for those conducting phonetics research. Recent work (Freeman & De Decker 2021, Sanker et al. 2021, Zhang et al. 2021) has investigated

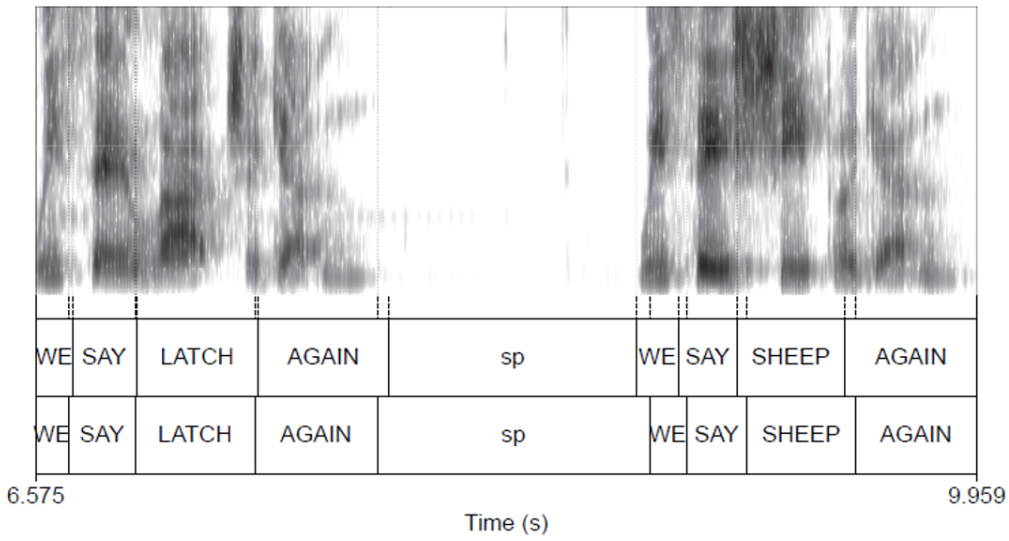

FIGURE 3. Comparison of word-level alignments from the Skype condition (top tier) and the gold standard H4n (bottom tier) for speaker CS. Audio from the Skype condition.

Figure 8.2: Figure from Sanker et al. (2021: e372).

the viability of different remote recording methods for various types of phonetic inquiry, making the process of selecting a remote recording method more straightforward. This section discusses specifically the results found in the Sanker et al. (2021) paper, although all of these works draw similar conclusions.

Sanker et al. (2021) tested ten different remote recording methods: Android recording, external microphone, internal laptop microphone, iPad, iPhone, Skype, Zoom Meetings, Facebook Messenger, Cleanfeed, and Audacity. These were compared against a gold standard Zoom H4n recording of the same real-life utterances to determine potential issues in that recording method.

One of the differences between remote and gold-standard recording methods was a discrepancy in the segmentation boundaries placed by a forced alignment algorithm. An example of this issue is shown in Figure 8.2, where the same audio recording is segmented differently when recorded on Skype (the top tier) versus using the gold standard Zoom H4n recorder (bottom tier). In this case, the difference stems from the fact that Skype record-

ings (similarly to some of the other recording methods tested) are compressed in a way that changes the signal slightly and results in these differences, which could cause errors in measurements on specific phones, or measures of duration. Other recording methods, such as iPad recordings, allow for more background noise in the recording itself, obscuring phonetic cues to consonant-vowel transitions, for example. The forced aligner then has a more difficult time finding these cues in the signal and places the segment boundary incorrectly.

While all methods had some digital artefacts of recording, as discussed via this example, broadly they performed well outside of some small effects of segment duration, vowel formants, and signal-to-noise ratios, as well as large discrepancies in center of gravity measurements. Crucially, however, recording methods always maintained contrasts between phonemes, despite potentially having all measurements shifted from the gold standard. Only one recording method, Facebook Messenger, was so problematic that the authors recommend it be avoided, especially since there are fewer customizable settings available to the user. Cleanfeed gave the best results (closest to the gold standard), but Skype and Zoom Meetings have the advantage of having a video option and also likely being more familiar to language consultants. The major takeaway from this work, however, is that documenting your recording setup, in all cases but especially when recording remotely, is extremely important for others to understand potential confounds of your results.

8.2.2 Using archival materials

Another option other than collecting one's own field data is to make use of existing linguistic data that has been deposited into a language archive. There are many digital language archives with deposits of language materials, including ELAR, PARADISEC, AILLA, CLA, AIATSIS, PARADISEC, and more. These repositories are a great resource, especially for those looking at multiple languages at once for a comparative or typological project. They are all relatively accessible, but each archive, and each collection in that

archive, will have its own idiosyncrasies that can take some work to figure out. This section lays out some of the main points from an audit of language archives considering their usability and accessibility (Lake et al. 2022); and one considering the usability of specific archive collections within archives (Babinski et al. 2022).

Archive use and usability

Digital language archives differ in their web structures and site organization. They may have different search functions and capability, different protocols for accessing collections, and different accessibility of the site and collections overall. Lake et al. (2022) conducted an audit of about 50 digital language archives to investigate their performance with respect to their accessibility, discoverability, and functionality. One important finding in terms of archive accessibility is that most of the digital archives surveyed had few or no site interface languages besides English and perhaps a few other major world languages. This is a problem especially for endangered language collections, where community members in particular may not be able to access the materials on their own language because of this issue. Running the websites through Google Translate is not a satisfactory alternative to having full site translations available, because the automatic translations are often incomplete and can remove functionality of certain links and lists on the website.

The discoverability of materials in digital language archives varies widely, which is an important consideration when beginning to search for archival materials to use. Archives vary as to the search parameters available, e.g. whether a user can search for specific file types (like ELAN files), specific types of recording such as narrative or elicitation, or whether search functions within a collection were available at all. Metadata is also often sparse or missing in the web interface; while there is often a text file within a collection laying out the collection's contents, when this information is not available on the website itself it can be difficult to find out what materials exist in the collection and whether it is

appropriate for the user's purposes or not.

One final issue I will mention with respect to digital language archives is that there is very rarely an option to download an entire archival collection as once as a bulk download. Usually, after identifying which files are most likely to be useful, they will need to be downloaded one at a time from the archive website. While this does not prevent the use of archival collections, it is an initial time-consuming step that one might not factor into plans for working with these materials.

Accessing and using archival collections

Just as each archive is structured differently, each individual language collection may be structured differently as well. Standards and recommendations for the organization of archival collections have cropped up over the years, but none of these have become widespread enough to standardize language collections across the board. Babinski et al. (2022) conducted a survey of 20 archival collections to investigate the structure of these materials and found wide variation in all aspects of the organization of archival deposits.

Sometimes, archives collapse existing file structures when collections are deposited. This can make navigating the collection overall less streamlined, and issues arise especially when not all corresponding files (e.g. audio WAV file, video of the same speech event, transcription of the speech) have the same file name. Corresponding metadata files can often be hard to find in the collection, i.e. the archive website does not highlight or tag this file and it is not always clearly labeled as 'metadata' or something similar. A compounding issue here is that metadata files sometimes refer to older versions of files that were named differently, which can cause confusion for the user as well as for any potential scripts one might want to run on the files to extract transcriptions or the like.

Constructing and depositing one's language materials into an archive is not a standardized process by any means, and depositors often have very little guidance available as to the

most usable way to organize their files. Babinski et al. (2022) provide some recommendations for those looking to archive their materials in order to make it clear what practices are helpful and which can cause confusion or complete lack of usability from the perspective of the archive user.

Archival collections will differ along multiple dimensions. They will vary in the time of recording, from the earliest recordings of the 1960s and earlier to present day. The impact of this fact is that recording qualities of a given collection may not be up to current standards, and the equipment used will likely not be the modern gold standard. Detailed information about the equipment used is not always available, but this can impact certain phonetic measurements one may want to study. A second way language collections will differ is in the environment in which the recording took place. Archival materials of this type were usually collected in a field setting, which means recordings may have been made outside or otherwise in locations with substantial background noise. There may be multiple speakers at different distances from the recording device, or a group of speakers having a conversation. Finally, archival collections will differ as to the types of language that are recorded. Some will have full narratives and conversations, while others may have primarily traditional elicitation of isolated words and sentences. Most will have both, and many archives have the capability to tag the type of recording in their web interface so that each type will be easily searchable.

8.3 Post-processing techniques

All of the data acquisition methods discussed in the previous section— remote recording, archival materials, and field-collected audio— all pose their own issues in terms of background noise and linguistic variability. While variation and noise in audio recordings is all but inevitable, a variety of post-hoc digital signal processing methods can help to mitigate the negative effects of it. This section outlines some of the major tools and techniques that

can be used to help clean data and speed up processing time, especially when working with audio data collected outside of lab settings.

When working with more than one archival deposit at once for comparative/typological purposes, these processing methods are important for making the recordings as comparable as possible. They may have been collected using different equipment, and certainly were collected in different places, at different times, in different settings. Eliminating the extra variation introduced by all of these factors is crucial for having comparable audio corpora for each language.

8.3.1 Forced alignment

Automatic speech-to-text alignment can greatly reduce processing time and can be used even with small amounts of data using a pretrained model from another language. The resulting alignments will need manual adjustment, but overall time to align is cut down substantially by first using the automatic algorithm.

Many forced alignment algorithms require large amounts of data to create a language-specific training model. When conducting research on endangered or under-researched languages, this amount of aligned audio is often non-existent and would require a massive time investment on the part of the researcher(s) in order to do by hand. For this reason, a growing body of work has investigated the usability and accuracy of forced aligners in endangered and under-researched language work. Using a forced alignment algorithm trained on a model that differs from the target language results in higher rates of error in the resulting alignment, but with manual correction it can offer a way of greatly reducing segmentation time.

DiCanio et al. (2015) compares the performance of the P2FA and HM-Align alignment models on Yoloxchitl Mixtec language data. The data consisted of elicited word lists that were constructed to collect words of varying lengths. They found that HM-Align made

fewer errors in alignment than P2FA, and that certain types of segments had higher error rates than others. The authors attribute the differences in performance primarily to the fact that HM-Align uses an allophonic English phone set, allowing for a greater phonetic specificity than P2FA, which uses a context-free phonemic English set of phones. When using English-trained models on non-English language data, the availability of a wider set of phones for transcription can only increase the accuracy of forced alignment results.

Another study investigating the feasibility of forced alignment for aligning endangered language recordings is Johnson, Di Paolo & Bell (2018), who look at the performance of the Prosodylab Aligner on Tongan field data. The recordings used in this test were word lists in Tongan, although the authors note that they planned to run forced alignment on connected speech in the future. Recordings were made in a field setting with all the requisite background noise that that environment entails. The authors considered both the raw recorded audio ('dirty' files) as well as audio that had been cleaned of ambient noise ('clean' files). Results were fairly accurate for both types of files, as long as the aligner was trained only on cleaned data. Furthermore, when compared to two different humans' manual alignments of a subset of the data, the Prosodylab Aligner's results did not differ from manual alignment any more than one human's alignment might differ from another. The authors conclude that using forced alignment in this way, even with manual corrections post-alignment, is a viable time-saving option for those looking to align their field recordings.

Babinski et al. (2019) considered the performance of forced alignment specifically for Australian languages, using around 45 minutes of running Yidiny speech as the test data. The data were collected in a field setting and transcribed at the utterance level. The aligners compared in this study were P2FA, DARLA, and MFA. Of particular interest here were the various potential approximations in transcription using the ARPABET transcription system. Because ARPABET was created to transcribe English phones, some Yidiny sounds were not available and needed to be approximated in some way. Specifically, alternatives were con-

sidered for the transcription of stops, which need to be specified for voicing in ARPABET despite having no voicing contrast in Yidiny; the palatal nasal, which was transcribed alternately as N, Y, and N+Y; and the trill, tap, and retroflex rhotic, which are separate phonemes in Yidiny but have very different distributions in English. The optimal transcriptions for these sets of phones were: stops transcribed as voiceless P T K; the palatal nasal transcribed as N; and the rhotics as R (trill), D (tap), and R (retroflex). These are the standards used in transcribing language data in this dissertation. In some cases, this results in apparent neutralization of phonemic contrasts that do exist in the language; however, these are all recoverable neutralizations that were judged acceptable for the purposes of this study focusing mainly on vowel acoustics. Single sounds represented with two ARPABET characters, such as N+Y for the palatal nasal, can be combined after forced alignment in Praat, while in the case where it is transcribed as N reference needs to be made to the orthographic word, which is retained in the output TextGrids. Those looking to do this kind of work should be aware of potential neutralizations and determine the best way to transcribe in ARPABET in a way that retains the most crucial contrasts for the object of study.

The P2FA and MFA aligners performed similarly to one another in terms of prosodic alignment, vowel measurements, and consonant durations, and were fairly accurate to the gold standard manual alignments. The unconstrained version of DARLA was used, and this model performed significantly worse than the other aligners tested. There is no particular reason why we should expect English-trained models such as these to perform well on non-English data, as that is not what they were created to do. However, it is useful for endangered language work to know that some of these algorithms have fairly high accuracy that can substantially cut down on processing time.

Forced alignment models can save the researcher time in segmentation of audio even with manual correction and can reduce the variability introduced by having multiple human segmenters, as all of the mistakes that the automatic alignment will make will at least be

internally consistent with the particular algorithm. This can help ease the burden of the 'transcription bottleneck,' where the time to align materials collected can be many times longer than the time it takes to record them. The processing of data for this dissertation is a case study in the potential to save time in creating segmented transcripts, as the time it would take to segment sixteen language collections by hand would likely greatly exceed the time of an average PhD program. More details about how forced alignment is used in this dissertation can be found in §3.1.

8.3.2 Data normalization

Even if you are able to collect novel recordings in the field yourself, there may be residual background noise as well as structured sources of uncontrolled variation in your data. Normalization of data is always important, but it is even more so when working with more-highly variable data that were collected in field conditions, remotely, or from archives. Here, two types of normalization will be discussed, to deal with the two types of data noise. To deal with background noise and variability in recording conditions, considerations when normalizing intensity are discussed. Then, vowel space normalization methods are discussed; this type of normalization is used to control for linguistic factors that naturally affect vowel formants but that may not be the object of an investigation into vowel space variability.

Perhaps the most relevant acoustic measure that will be affected by levels of background noise is intensity. Intensity can be affected by recording device used, distance of the speaker from the microphone, levels of white noise or other interference in the background, as well as linguistically variable factors such as individual speaker, speech style, and random variation in the loudness of speech. Because many of these factors may differ across different instances of recording— e.g. the speaker may be a different distance from the microphone on a different day, or after taking a break— it is best to use locally relative

normalization methods. Normalizing intensity relative to the following syllable (as was done in this dissertation, cf. §3.2.1), or to the word or utterance average, is ideal to account for the local variability in this measure.

Another important type of data normalization is vowel space normalization, which is used to isolate the aspects of vowel space variability that is relevant for your particular research purposes; for this reason, the appropriate method to use will differ depending on the aspects of vowel articulation you are interested in. In this dissertation, I used Johnson (2020)'s ΔF method, which is a good method for cross-linguistic consistency in normalization (cf. §3.2.1). However, the needs of a specific project will vary; there are methods that will be more appropriate for isolating sociolinguistic variation within one language, for example, or for doing more general language documentation work and normalizing across speakers.

8.3.3 Controlling for noise with statistics

Certain statistical models can mitigate the effects of recording variability to an extent in addition to the data normalization methods already discussed. Here I will mention specifically regression models, which have been used extensively in this dissertation, especially in Chapters 4 and 5. Regression models allow for the inclusion of random intercepts and slopes, as well as additional fixed factors that can help to tease out your linguistic factor(s) of interest from other sources of natural language variation.

Let us take Regression Model A (the duration model, from (3.2) in §3.2.2) as an example:

(8.1) **Regression Model A:** vowel duration

```
lmer(vowel.duration ~ (1|word) + (1|seg.identity)
+ (1|speaker) + phonemic.length + word.finality + stress
```

By using the continuous variable, in this case vowel duration, as the dependent variable, the model is able to be tailored to include factors that have to do with the dependent variable, without including factors that are not hypothesized to be relevant for this specific acoustic measure. Here, we can include phonemic vowel length as a fixed factor, as well as word-finality, in addition to the fixed factor of research interest, stress. And indeed, in Chapter 4 we found that word finality is a significant predictor of vowel duration in many languages, and phonemic vowel length is important for languages that make such a distinction. These effects are found in the model results separately from the effects found for the stress variable, meaning that the effects found for stress do not include the confound that stress is never final and final vowels are often quite long.

Statistical methods such as regression modeling can help to isolate one's variables of interest when working with natural speech data that do not isolate these variables already, as we might expect from a targeted experiment. In addition, these models allow for the discovery of other effects related to the dependent variable, that may be tangential to the research question but could reveal interesting interactions and effects that would not be revealed using a highly controlled experimental setup.

8.4 Conclusion

There are many ways of conducting good acoustic phonetic research, and recent innovations in remote communication and computational processing of language materials have made these methods more accessible than ever. It is important to keep in mind, however, that all recording settings and methods have their pros and cons, so documentation of data collection and processing methods is crucial.

Another result of recent technological innovations is that data collection does not always need to take place in person with a dedicated recorder. Remote recording methods are a good option for those who have existing community relationships, and who are lim-

ited in terms of funding, time, or health restrictions and concerns about travel. In addition, language archives are a great resource for working with languages that are endangered or are no longer spoken, as the materials have already been collected and can usually be accessed via download from the Internet. Per the discussion above about the idiosyncrasies of accessing and using archival materials, those who do collect their own audio recordings in the field should be aware of potential user experiences with their collections while they are organizing them for depositing. One's own field recordings can be paid forward for future researchers, especially in cases where the language in question is already severely endangered, since your materials may be some of the few resources remaining for the language in the near future.

When working with natural speech data, proper data cleaning and processing methods are crucial, as have been discussed in this chapter. However, this should not preclude the use of these materials for phonetic study; natural speech materials have been prioritized in sociolinguistic data collection methods for a long time (cf. Beckman (1997)) and are becoming more highly valued in phonetics and phonology more generally (Cohn & Renwick 2021). There is clear value to the ability to see a phonetic or phonological phenomenon existing 'in the wild' of natural speech, to more fully understand how it is realized in everyday language use and how it interacts with other phenomena in the language. Experimental methods can more clearly isolate a phenomenon of research interest, and measure its effects separately from other aspects of language, but this is only one way to understand the totality of a linguistic process. Spending more time with natural speech data can help us to understand how all of the many linguistic processes occur simultaneously with one another to create all the nuances that languages encode.

Chapter 9

Conclusion

This dissertation has presented a study of the structured variation present in the acoustic correlates to lexical stress, using spontaneous speech data from sixteen Indigenous languages of Australia. The correlates to stress in each language were identified using regression modeling in Chapter 4, and qualitative comparison of the observed cross-linguistic variation was discussed. The phonetic cues to stress were also found to vary across speakers of the same language in some cases, as presented in Chapter 5, resulting in a more nuanced view of the language-wide effects. The structure of this variation was investigated more systematically in Chapter 6, which presented results of two phylogenetic tools, AMOVA and F_{ST}, to determine that there is indeed significant variation in the cues to stress both across languages and across speakers within the same language and to create a network model of language relationships that were identified using only this acoustic data. These results provide evidence that the acoustic correlates of lexical stress can be considered to undergo regular sound change and can be studied as such.

Lexical stress is not the only prosodic phenomenon that I propose to undergo regular sound change; the same principle should generalize across prosodic levels into phrasal and sentential prosody as well. However, the Australian languages in this study were

particularly well-suited to an investigation of lexical stress, because most Australian languages have consistent initial stress placement (Goedemans 2010, Fletcher & Butcher 2014). Phrasal prosody in these languages varies, however, so hypotheses about what these phrasal contours are was necessary. Chapter 7 presents an automated categorization of phrasal contours using the methods put forth by Kaland (2021) to establish initial hypotheses about phrasal categories in these languages. Automatic prosodic analyses such as this are very useful tools for fieldworkers hoping to provide some prosodic description of a language, which is sorely needed in language documentation work (Whalen, DiCanio & Dockum 2020, Macaulay 2021). Prosody has yet to be integrated into the models of lexical stress variation and change presented earlier in this thesis; this is an avenue of future research.

Chapter 8 provided some practical discussion concerning the acquisition and processing of archival language materials for language research and documentation. The tools and skills involved in accessing and processing materials available in language archives are vitally important and will only increase in demand as archives grow and more and more languages cease to be spoken. This dissertation project provides a case study in ways to work with this type of data to answer novel research questions.

What follows is the final chapter of this dissertation. The main claims of this study are presented in light of the results in §9.1 along with a summary table. There is a brief discussion of endangered language phonetics in §9.2, and §9.3 discusses implications of this work and concludes the dissertation.

9.1 Revisiting the Claims

Below are the claims put forth in §1.3 of this dissertation. The scope of these questions easily extends beyond what has been done here, but each claim has been addressed to the extent possible in a work of this size.

Claim (9.1) concerns the linguistic heritability of the phonetic cues to stress. The main points of evidence for this claim are within-language consistency across speakers (9.1a), variation across languages and shared cues in closely related languages (9.1b), and the existence of significant cross-linguistic variation in the population structure analysis (9.1c). Point (9.1a) was addressed in Chapter 5, when the speakers of a language are consistent in their use of an acoustic cue. Point (9.1b) was investigated in the cross-linguistic results of Chapter 4, and in the F_{ST} values and resulting NeighborNet diagram in Chapter 6. While further study will be able to quantify the amount of shared stress cues among related languages versus unrelated languages, similarities among related languages were observed, along with similarities between areally related languages in some cases as well. Finally, point (9.1c) was addressed in the population structure analysis in Chapter 6 using Analysis of Molecular Variance (AMOVA). Cross-linguistic variation in the data was found to be significant separately from within-language and within-speaker variation.

(9.1) The phonetic factors that cue linguistic prominence are linguistically heritable, meaning that they vary and change in similar ways to phonemes. They vary in a structured way within a language and will remain relatively stable until a change occurs, in a way that is analogous to phonological change. For this claim to be true, the following must hold:

a. Prominence cues are consistent across speakers of the same language. Speakers converge on the same cues to prominence and use these to the exclusion of other potential cues.

b. Closely related languages will be more likely to share cues to prominence than languages that are more distantly related. This should be the most clearly observed when languages have the same pattern of stress assignment, as changes in stress position will increase the likelihood of cue changes.

c. Study of population structure will show that significant variation exists across languages, separately from any within-language or within-speaker variation.

Claim (9.2) concerns within-language sociolinguistic variation in the acoustic correlates of stress. The main focus of Chapter 5 was inter-speaker, within-language variation, which addresses point (9.2a) here. The presence of interspeaker variation for many of the acoustic cues investigated implies that this variation is socially conditioned in some way (point 9.2b). While this point is a logical extension of point (9.2a), it is not an answerable question using the data set in this dissertation, because direct speaker consultation and observation of social structure was not possible. In some cases this is a fruitful area of future research, but in other cases there are no longer speakers of a language and further study is not possible. Finally, point (9.2c) refers to population structure, which was addressed in Chapter 6. The results of the AMOVA modeling found significant variation across speakers of the same language, providing support for this point.

(9.2) Just like all other linguistically variable phenomena, the phonetic cues to stress can also vary along sociolinguistic lines within a language. In order for this to hold, the following must be true:

a. Some cues to stress within a language may only be cues for some speakers, and not for others. Similarly, speakers may vary in their use of these cues based on the social situation.

b. This variation falls along defineable social lines, such as gender identity, dialect, social status, register, etc. (This cannot be studied well with these data)

c. Study of population structure will show that significant variation exists across speakers within a language, separately from any cross-linguistic or within-speaker variation.

Claim (9.3) is addressed in investigations of cross-linguistic and inter-speaker varia-

tion. Almost all of the languages in the study (summarized again in Table 9.1) have multiple acoustic factors that correlate with lexical stress. Likewise, in all languages that have more than one speaker, at least one of these factors showed interspeaker variation. These results show that stress is often marked by multiple cues, and not all of these cues are only doing the work of marking the stress contrast. Some factors that show interspeaker variation are additionally conveying some information about the speaker, be it age, gender, or social status. This sociolinguistic variation could potentially be an example of change in progress, where one group has taken up the change and another has not yet. It could also be an example of stable sociolinguistic variation that is not moving in one particular direction.

(9.3) Different prominence cues may co-occur with one another, marking the same type of prominence with multiple acoustic factors. However, the presence of multiple cues may make each individual factor more unstable in the system, as the crucial contrast would not be lost with the loss of one cue. Some cues may hold for all speakers in the language, while others may be sociolinguistically variable.

Language	V. Dur.	Onset Dur.	Post-T. Dur.	Inten.	F0 Max.	F0 Rng.	Vowel
Bardi	+	+		∼	∼	+	+
Burarra	∼	∼					
Gunnartpa	∼	∼					
Gija	∼	∼∼∼			+	+	+
Dalabon		∼					
Gunwinggu			+	+	∼		
Kunbarlang		+			∼		
Kayardild	+		+ ∼∼	∼	+	+	+
Malak Malak	+	+ ∼		+	+		
MurrinhPatha*			+				
Ngan'gi	+	∼∼		+	∼		
Wanyjirra*		+ + +			+		
Warlpiri*	+	+	++		+		+
Warnman*	+						
Yannhangu*	+	+ + +			+		+
Yidiny	+	+	+	+	+		

Table 9.1: Summary of regression model results from Chapters 4 and 5; + indicates an overall effect of stress in the language; ∼ indicates an effect of stress among some (but not all) speakers in the language. Languages grouped by historical affiliation. Those languages with only one speaker are marked with an asterisk (*).

9.2 Endangered language phonetics

One of the major goals of this dissertation is to demonstrate the practical implementation of corpus phonetic methods on endangered language materials sourced from language archives. It accomplishes this goal in two ways. First, the process of acquiring and processing archival materials before the main analyses is documented in detail, both in the presentation of data and methods in Chapters 2-3 and in the specific discussion of archival phonetic methods in Chapter 8. Second, this dissertation presents a case study in the use of archival materials to address novel research questions in phonetics. The audio that was recorded in these languages was not created with this type of study in mind, but through forced alignment and data normalization meaningful analysis can be done to learn more about variation and change in prosodic phenomena.

Work on prosody and stress in endangered languages is not as common as other types of phonetic documentation (Whalen, DiCanio & Dockum 2020, Macaulay 2021). This dissertation presents a broad typological study of lexical stress in sixteen Australian languages, considering both cross-linguistic and inter-speaker variation in the phonetic correlates of stress, as well as an automated identification of prosodic contours in these languages. The methods used for these analyses was kept maximally automated in order to provide practical tools for conducting such typological studies in any group of languages with available narrative speech data. The wider use of largely automated processing and analysis methods can serve to increase the number of languages with some prosodic description.

9.3 Implications

This dissertation has presented a typological study of lexical stress and prosody in sixteen Indigenous languages of Australia. Through this investigation, variation has been observed both across languages, and across speakers within languages. This is not just a qualitative

difference in stress correlates, but has been quantified using phylogenetic methods in Chapter 6. In fact, these methods have revealed that the phonetic cues to stress can be shared strongly between historically related languages, as well as between geographically close languages.

These results support the claim that the phonetic markers of a prosodic phenomenon such as lexical stress varies in structured ways that indicate these markers vary and change in a principled way, and thus can be studied similarly to linguistic studies of segmental change. Despite the fact that the acoustic correlates to stress— phonetic factors such as duration, intensity, f0— all serve to mark the same type of phonological event, in this case initial lexical stress, the phonetic variation in this marking is still structured in the way that a phonologized phonetic factor such as phonemic stop voicing might be. The results presented in this dissertation indicate that the phonetic correlates of stress are shared among related languages, although changes can apparently be diachronic or contact-based, similarly to many types of segmental change. Change can also occur within subpopulations of a language, creating variation along sociolinguistic lines. Further work is required to understand the level of awareness speakers have about this sort of speaker variation, but it does exist at least in some cases.

These results are in keeping with the findings of e.g. Kakadelis (2018), who found structured variation in voice onset time among languages that do not make a phonemic voice onset time distinction. Such observations speak to the nature of the language faculty and the cognitive organization of language, even below the abstract level of the phoneme, and to our theories of phonetic change and the phonetic precursors to phonological change. Hypothesizing that variation is structured and change is regular at this sub-phonemic level is in keeping with exemplar-theoretic models of language change, which hold that fine-grained phonetic changes are what drive the higher-level phonemic changes in language (Pierrehumbert 2001, Wedel 2006, Cohn & Renwick 2021).

This work is the beginning of what I hope to be a larger research program looking at variation and change in the phonetic correlates of stress and other types of prosody. While these historical and areal relationships have been detected in this set of languages, where stress is always placed on the initial syllable of the word, the question arises of what these phonetic factors look like when stress placement is variable across a group of languages, or when one language has variable stress assignment itself. Perhaps a change in the correlates of stress can result in a reanalysis of stress placement altogether, changing from a weight insensitive to a weight sensitive system, for example. On the other hand, a change in stress placement patterns may result in a reanalysis of the phonetic cues to stress, or perhaps variation and change in stress placement versus phonetic stress marking are entirely unrelated. While these questions could not be addressed here because of the stability in stress placement, what has been presented in this dissertation builds a foundation on which more complex patterns and interactions can be studied.

Appendix A

Archival Collections

This appendix includes information about the archival deposits sourced for audio materials for this project, as discussed in Chapter 2.

Table A.1: Archival collections used for this project.

Language	Collector	Archive	Speakers	Vowel Tokens	Code & Link
Bardi	Claire Bowern	AIATSIS	3	8,949	Bowern_C05
Burarra	Margaret Carew	PARADISEC	5	1,857	MLC1
Dalabon	Maia Ponsonnet	ELAR	5	7,721	0071-IGS0125
Gija	Frances Kofod	ELAR	5	11,009	0098-MDP0190
Gunnartpa	Margaret Carew	ELAR	3	634	0276-SG0161
Gunwinggu	Aung Si	PARADISEC	3	689	SI1
Kayardild	Erich Round	ELAR	5	17,953	0029-FTG0025; 0029-IGS0039
Kunbarlang	Isabel O'Keefe, Ruth Singer	ELAR	2	1,656	0384-SG0324
MalakMalak	Dorothea Hoffmann	ELAR	2	2,235	0166-IPF0189
MurrinhPatha	Danielle Barth, John Mansfield	PARADISEC	1	823	SoCog-mwf01
Ngan'gi	Nicholas Reid	PARADISEC	9	7,200	NR01
Wanyjirra	Chikako Senge	ELAR	1	2,121	0125-IGS0079
Warlpiri	David Nash	N/A	1	1,948	N/A
Warnman	Nicholas Thieberger	PARADISEC	1	168	NT10
Yan-nhangu	Claire Bowern	ELAR	1	16,919	Bowern_C05
Yidiny	R.M.W. Dixon	AIATSIS	2	5,340	Dixon

Appendix B

ARPABET transcriptions

This appendix includes all of the grapheme-to-ARPABET conversions used when preparing transcripts for forced alignment, as discussed in §3.1.4.

IPA	Orthography	ARPABET	English equivalent
/i/	i	IH1	[ɪ]
/iː/	ii; i:	IY1	[i]
/e/	e	EH1	[ɛ]
/eː/	ee; e:	EY1	[eɪ]
/a/	a	AH1	[ʌ]
/aː/	aa; a:	AA1	[ɑ]
/o/	o	AO1	[ɔ]
/oː/	oo; o:	OW1	[oʊ]
/u/	u	UH1	[ʊ]
/uː/	uu; u:	UW1	[u]
/p/∼/b/	p; b	P	[p]
/t̻/∼/d̻/	th; dh	T	[t]
/t/∼/d/	t; d	T	[t]
/ʈ/∼/ɖ/	rt; rd	R T	[ɻt]
/c/∼/tʃ/	j; dy	CH	[tʃ]
/k/∼/g/	k; g	K	[k]
/m/	m	M	[m]
/n̻/	nh	N	[n]
/n/	n	N	[n]
/ɳ/	rn	R N	[ɻn]
/ɲ/	ny	N	[n]
/ŋ/	ng	NG	[ŋ]
/w/	w	W	[w]
/j/	y	Y	[j]
/l/	l	L	[l]
/ʎ/	ly	L	[l]
/ɭ/	rl	R L	[ɻl]
/ɻ/	r	R	[ɻ]
/ɾ/	rr	D	[d]

Table B.1: ARPABET transcription conventions

Appendix C

Vowel plots

The vowel plots in Chapter 4 were very small and hard to read in order to get a sense of the cross-linguistic variation. Here (Figures C.1-C.16) are those same plots in a larger format. Polygons represent vowel spaces for each speaker, and labels represent each speaker's mean vowel quality.

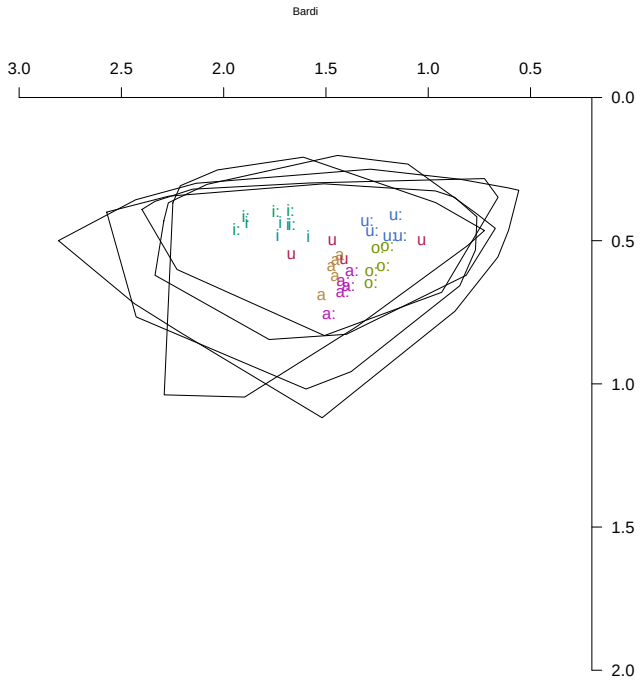

Figure C.1: Bardi vowel space.

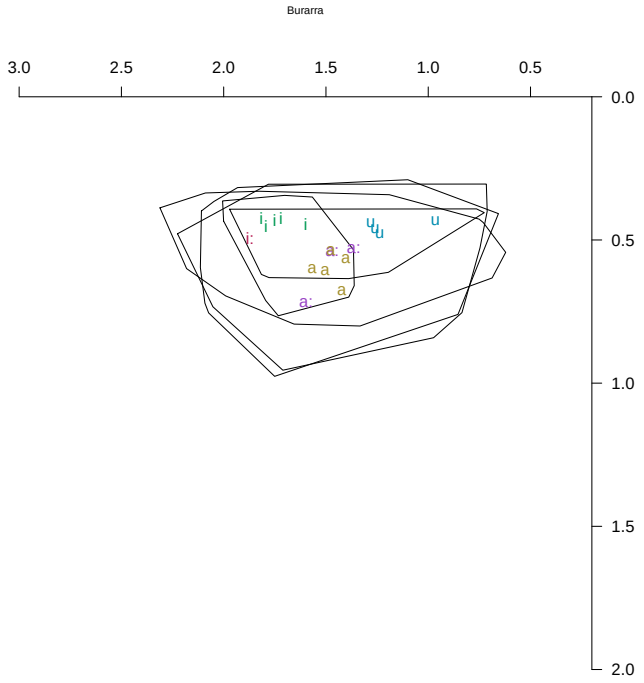

Figure C.2: Burarra vowel space.

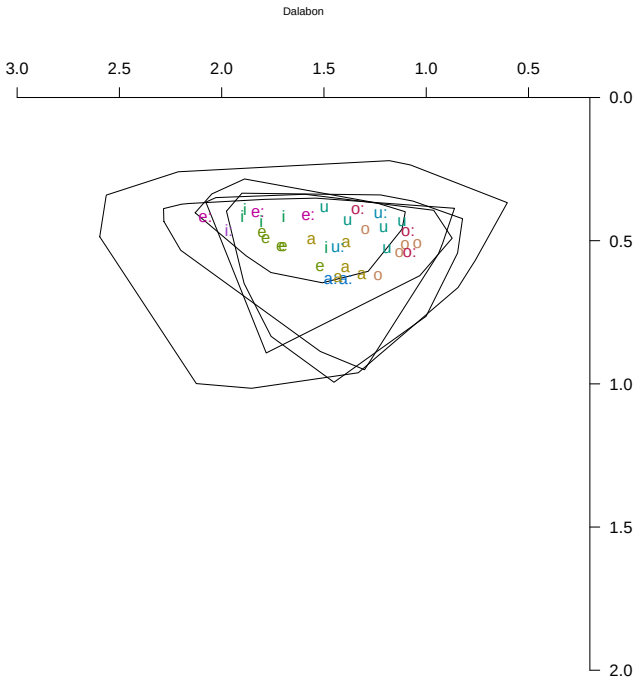

Figure C.3: Dalabon vowel space.

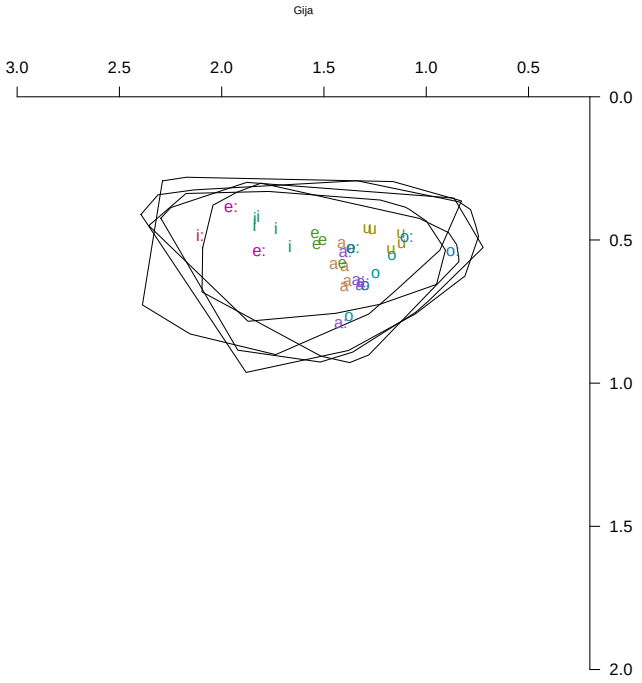

Figure C.4: Gija vowel space.

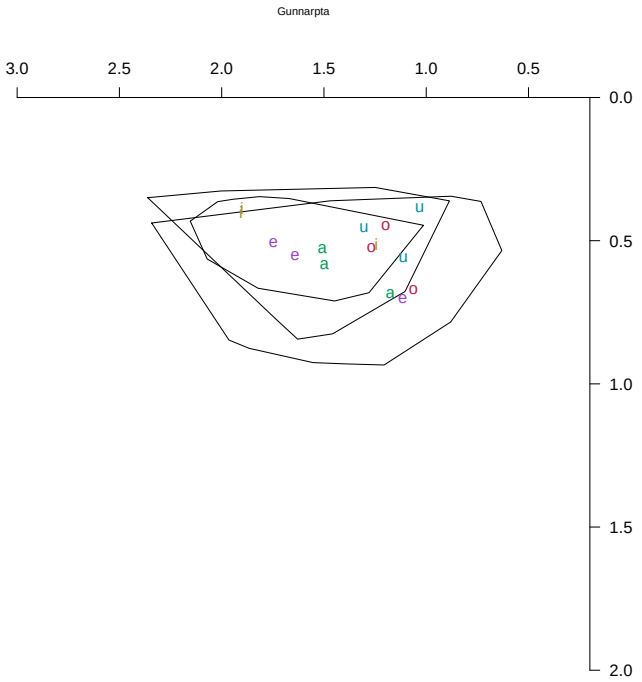

Figure C.5: Gunnartpa vowel space.

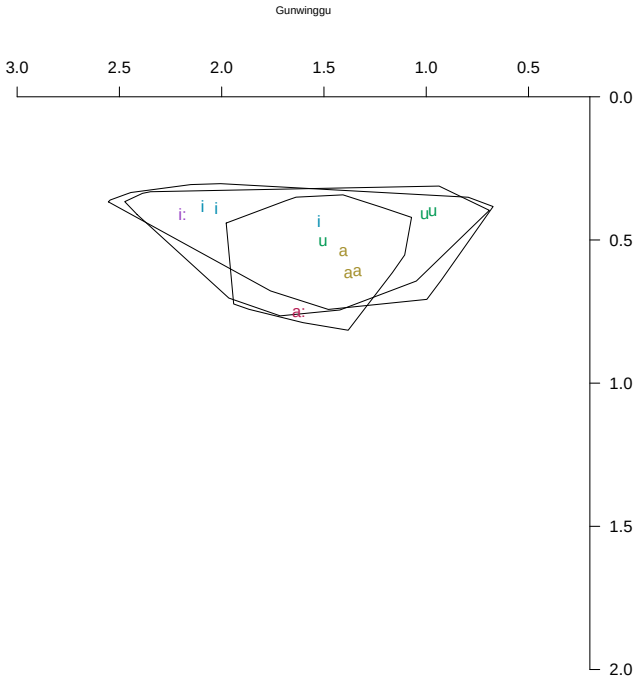

Figure C.6: Gunwinggu vowel space.

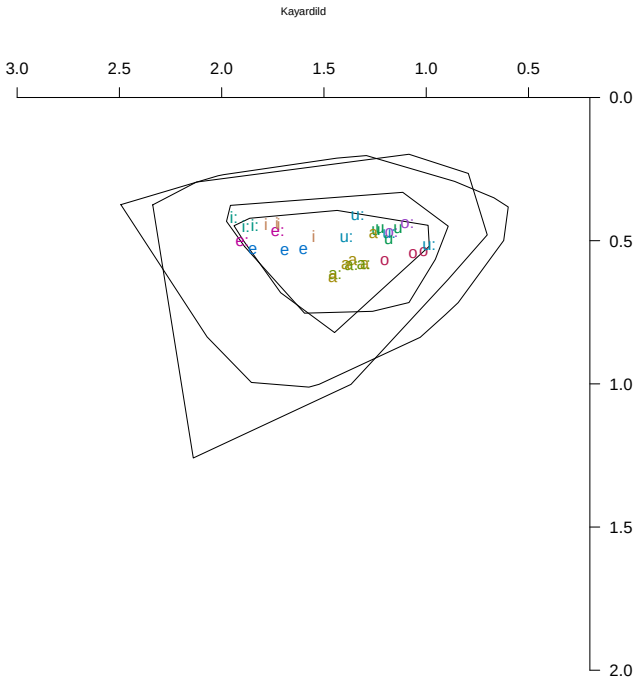

Figure C.7: Kayardild vowel space.

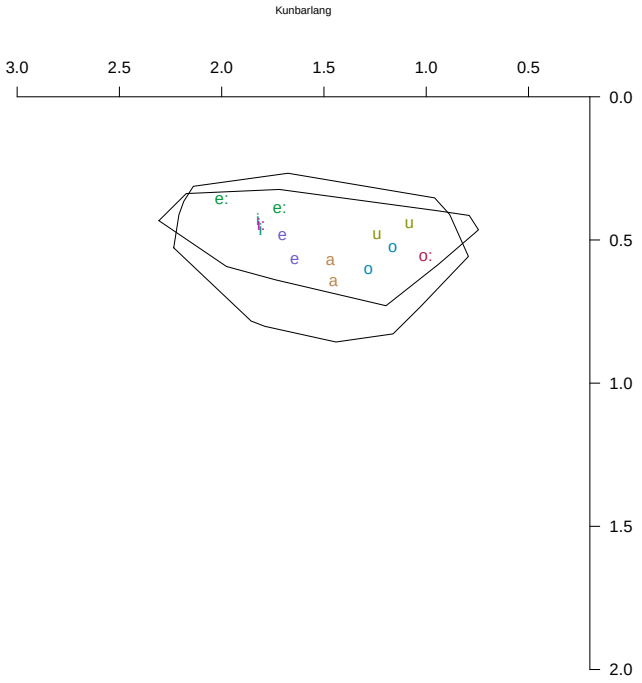

Figure C.8: Kunbarlang vowel space.

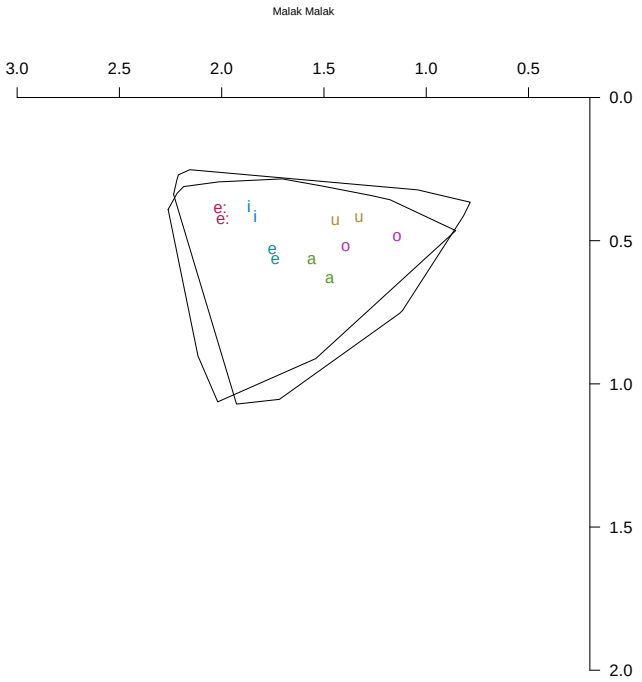

Figure C.9: Malak Malak vowel space.

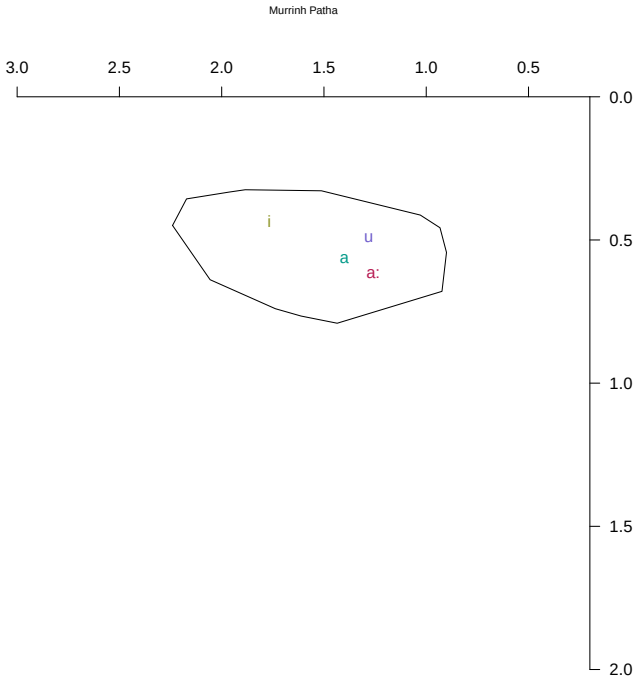

Figure C.10: Murrinh Patha vowel space.

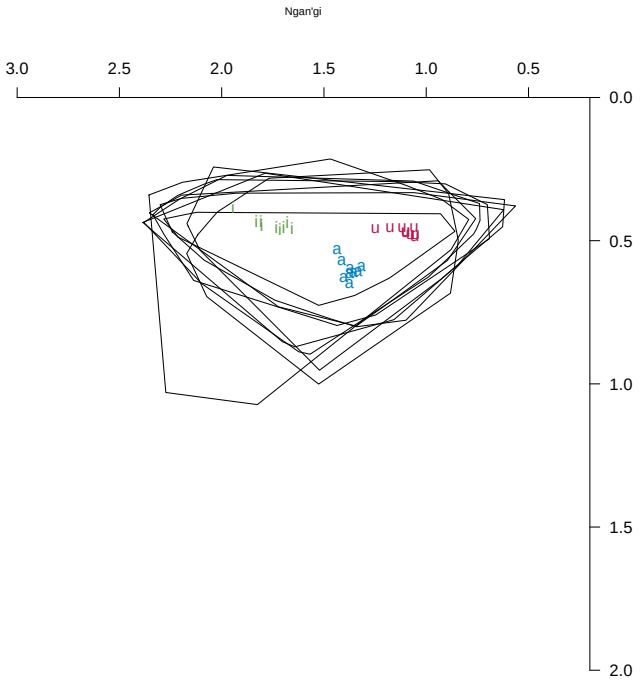

Figure C.11: Ngan’gi vowel space.

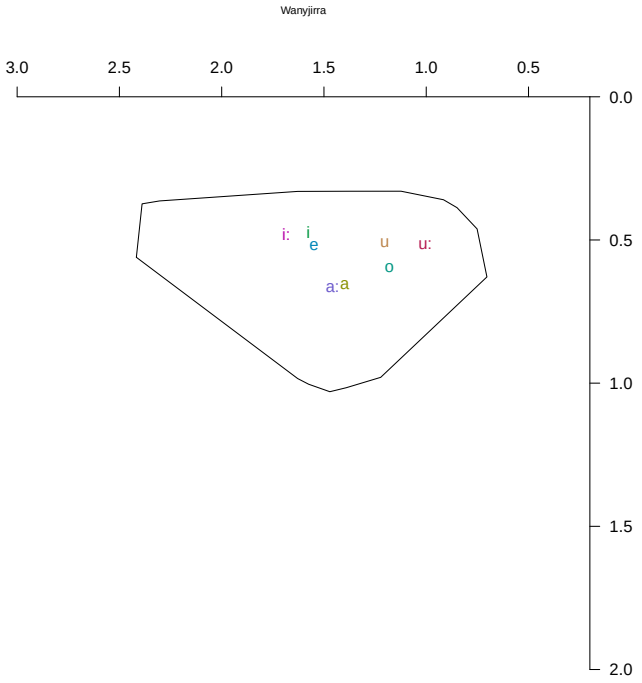

Figure C.12: Wanyjirra vowel space.

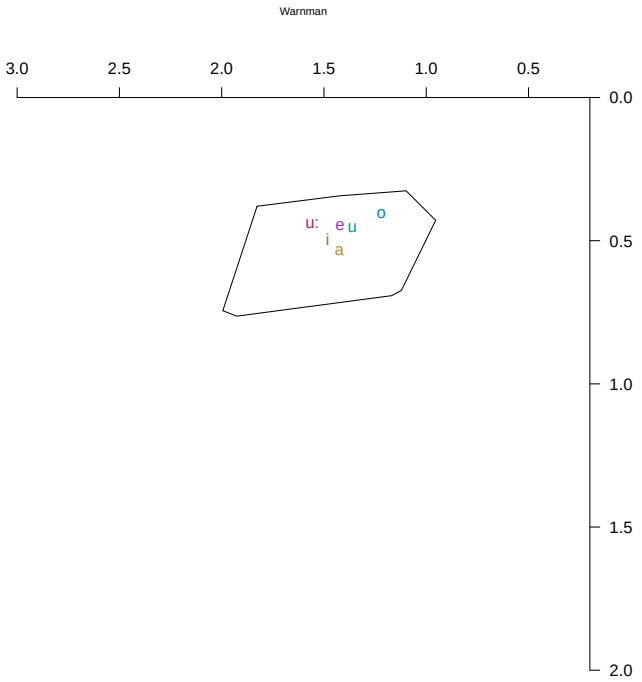

Figure C.13: Warnman vowel space.

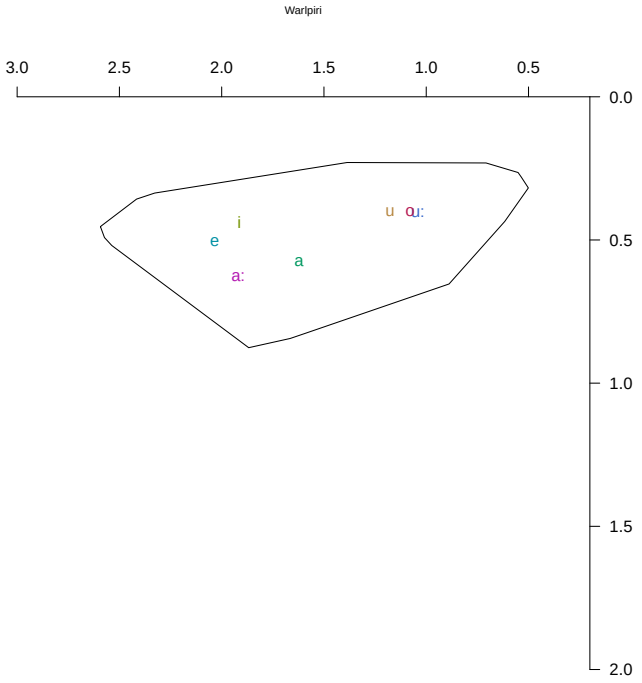

Figure C.14: Warlpiri vowel space.

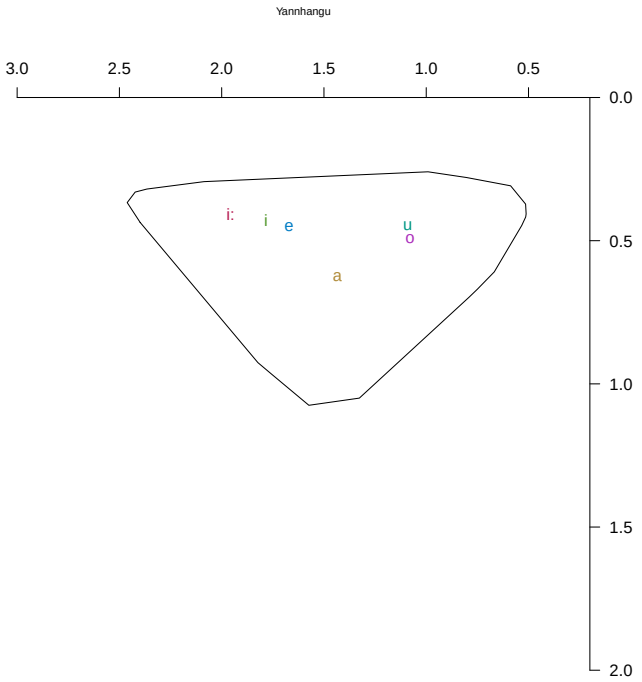

Figure C.15: Yannhangu vowel space.

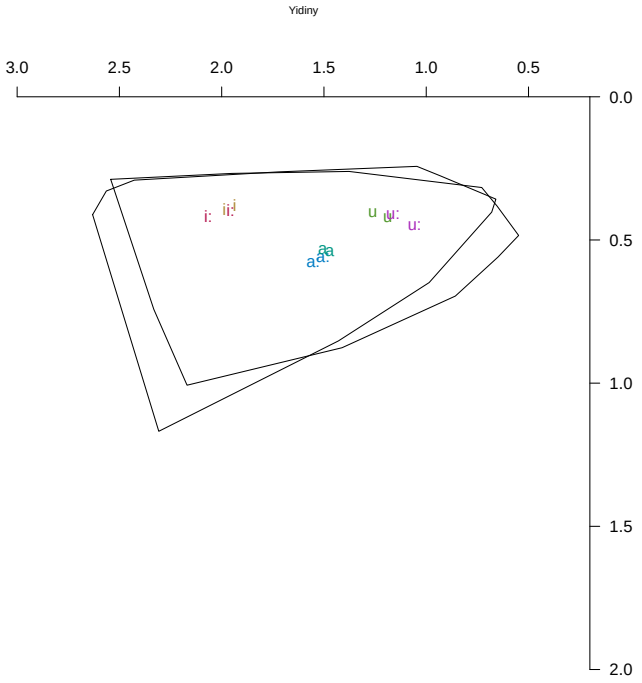

Figure C.16: Yidiny vowel space.

Bibliography

Abramson, Arthur S. & D. H. Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics* 63. 75–86. https://doi.org/10.1016/j.wocn.2017.05.002. (9 March, 2022).

Alpher, Barry, Nicholas Evans, Mark Harvey, et al. 2003. Proto-Gunwinyguan verb suffixes. In *The non-Pama-Nyungan languages of northern Australia: Comparative studies of the continent's most linguistically complex region*. Pacific Linguistics.

Anand, Pranav, Sandra Chung & Matthew Wagers. 2011. Widening the net: Challenges for gathering linguistic data in the digital age. *NSF SBE 2020. Rebuilding the mosaic: Future research in the social, behavioral and economic sciences at the National Science Foundation in the next decade*.

Atkinson, Quentin D. & Russell D. Gray. 2005. Curious Parallels and Curious Connections—Phylogenetic Thinking in Biology and Historical Linguistics. *Systematic Biology* 54(4). 513–526. https://doi.org/10.1080/10635150590950317. (28 December, 2021).

Babinski, Sarah. 2021a. *Forced Alignment and the Archive*. Talk. online.

Babinski, Sarah. 2021b. Intrinsic F0 and Sound Change: Evidence from Australian Languages. In *Proceedings of the Annual Meeting on Phonology*, vol. 9.

Babinski, Sarah. 2022. *Best practices in the collection and analysis of "noisy" audio in phonetics*. Talk. Washington, D.C.

Babinski, Sarah & Claire Bowern. 2021. *Contemporary Digital Linguistics and the Archive: An Urgent Review*. Talk. online.

Babinski, Sarah, Rikker Dockum, J. Hunter Craft, Anelisa Fergus, Dolly Goldenberg & Claire Bowern. 2019. A Robin Hood approach to forced alignment: English-trained algorithms and their use on Australian languages. *Proceedings of the Linguistic Society of America* 4(1). 3. https://doi.org/10.3765/plsa.v4i1.4468. (29 April, 2020).

Babinski, Sarah, Jeremiah Jewell, Kassandra Haakman, Juhyae Kim, Amelia Lake & Claire Bowern. 2022. *How usable are digital collections for endangered languages? A review*. Talk. Washington, D.C.

Baker, Brett. 2014. Word Structure in Australian languages. In H. Koch & Rachel Nordlinger (eds.), *The Languages and Linguistics of Australia* (The World of Linguistics), 76. De Gruyter.

Barth, Danielle. 2009. *Social Cognition Project*. Digital collection managed by PARADISEC SoCog.

Beckman, Mary E. 1997. A typology of spontaneous speech. *Computing prosody*. 7–26.

Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The Original ToBI System and the Evolution of the ToBI Framework. In Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 9–54. Oxford: Oxford University Press.

Ben Hamed, Mahé. 2015. Phylo-linguistics: Enacting Darwin's Linguistic Image. In Thomas Heams, Philippe Huneman, Guillaume Lecointre & Marc Silberstein (eds.), *Handbook of Evolutionary Thinking in the Sciences*, 825–852. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9014-7_39. (28 December, 2021).

Berinstein, Ava E. 1979. A cross-linguistic study on the perception and production of stress. *UCLA Working Papers in Phonetics* 47.

Bishop, Judith. 2003. *Aspects of intonation and prosody in Bininj Gun-wok: an autosegmental-metrical analysis*. University of Melbourne PhD Dissertation.

Blankenship, Barbara. 2002. The timing of nonmodal phonation in vowels. *Journal of Phonetics* 30(2). 163–191.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Boersma, Paul & David Weenink. 2018. *Praat: doing phonetics by computer. Version 6.0.43*. www.praat.org.

Bouckaert, Remco R., Claire Bowern & Quentin D. Atkinson. 2018. The origin and expansion of Pama–Nyungan languages across Australia. *Nature Ecology & Evolution* 2(4). 741–749. https://doi.org/10.1038/s41559-018-0489-3. (30 January, 2019).

Bowern, Claire. 2007. *Yan-nhaŋu Language Documentation 1*. http://hdl.handle.net/2196/00-0000-0000-0001-5178-5.

Bowern, Claire. 2011. *How many languages were spoken in Australia?* Publication Title: Anggarrgoon: Australian languages on the web Type: Wordpress. https://anggarrgoon.wordpress.com/2011/12/.

Bowern, Claire. 2012. *A Grammar of Bardi* (Mouton Grammar Library 57). De Gruyter Mouton.

Bowern, Claire. 2016. Chirila: Contemporary and Historical Resources for the Indigenous Languages of Australia. *Language Documentation and Conservation* 10.

Bowern, Claire. 2018a. Computational Phylogenetics. *Annual Review of Linguistics* 4(1). 281–296. https://doi.org/10.1146/annurev-linguistics-011516-034142. (26 December, 2021).

Bowern, Claire. 2018b. *Language, culture, and Australian exceptionalism*. Type: Invited talk. University of Arizona.

Bowern, Claire. 2020. *Tangkic and Pama-Nyungan: Sister or Subgroup?* Talk. Online. https://campuspress.yale.edu/clairebowern/australian-linguistic-society-talk/.

Bowern, Claire, Barry Alpher & Erich Round. 2013. *Yidiny stress, length, and truncation reconsidered*. Poster. UConn.

Bowern, Claire & Quentin Atkinson. 2012. Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* 88(4). 817–845. https://doi.org/10.1353/lan.2012.0081. (29 May, 2019).

Bowern, Claire & Bentley James. 2005. Yan-nhaŋu revitalization: Aims and accomplishments. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, vol. 41, 61–70. Issue: 2. Chicago Linguistic Society.

Bowern, Claire, Joyce McDonough & Katherine Kelliher. 2012. Bardi. *Journal of the International Phonetic Association* 42(3). 333–351. https://doi.org/10.1017/S0025100312000217. (9 June, 2021).

Carew, Margaret. 1993. *Gun-nartpa and Burarra audio recordings from Gochan Jiny-jirra and Maningrida*. Digital collection managed by PARADISEC MLC1.

de Carvalho, Fernando O. 2010. Vowel acoustics in Pirahã. *Revista de Estudos da Linguagem* 18(1).

Cohn, Abigail C. & Margaret E. L. Renwick. 2021. Embracing multidimensionality in phonological analysis. *The Linguistic Review* 38(1). 101–139. https://doi.org/10.1515/tlr-2021-2060. (29 October, 2021).

Cole, Jennifer & Stefanie Shattuck-Hufnagel. 2016. New methods for prosodic transcription: Capturing variability as a source of information. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1). Publisher: Ubiquity Press.

Coombs, Afton L. 2013. High tone processes in Ibibio. In *Proceedings of the Meeting on Acoustics*, vol. 19.

Coto-Solano, Rolando, Sally Akevai Nicholas, Brittany Hoback & Gregorio Tiburcio Cano. 2022. *Data Management in Untrained Forced Alignment for Phonetic Research: Examples from Costa Rica, Mexico, the Cook Islands and Vanuatu*. Poster. Washington, D.C. https://lsa.slidespiel.com/event/lsa-2022-01-06/poster/0a96204d-9e58-416b-9cd1-5498e5448ef5.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. Pearson Education.

Darwin, Charles. 1871. The descent of man, and selection in relation to sex. Publisher: John Murray.

de Dear, Caroline, Francesco Possemato & Joe Blythe. 2020. Gija (East Kimberley, Western Australia)–language snapshot. *Language Documentation and Description* 17. 134–141.

Dicanio, Christian. 2017. *Praat scripts*. http://www.acsu.buffalo.edu/%20cdicanio/scripts.html.

DiCanio, Christian, Hosung Nam, Jonathan D Amith, Rey Castillo García & Douglas H Whalen. 2015. Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec. *Journal of Phonetics* 48. Publisher: Elsevier, 45–59.

DiCanio, Christian & D H Whalen. 2015. The interaction of vowel length and speech style in an Arapaho speech corpus. In *Proceedings of the 18th Congress of Phonetic Sciences*, 5. Glasgow, UK: University of Glasgow.

Dixon, R.M.W. 1977a. *A Grammar of Yidiny*. Cambridge: Cambridge University Press.

Dixon, R.M.W. 1977b. Some phonological rules in Yidiny. *Linguistic Inquiry* 8. 1–34.

Dixon, R.M.W. 2001. The Austrlian linguistic area. In Alexandra Y. Aikhenvald & R.M.W. Dixon (eds.), 64–104.

Dixon, R.M.W. 2002. *Australian Languages: their nature and development*. Vol. 1. Cambridge: Cambridge University Press.

Dockum, Rikker. 2017. Computational modeling of tone in language documentation: citation tones vs. running speech in Chindwin Khamti. *Proceedings of the 43rd Annual Meeting of the Berkeley Linguistics Society*. Publisher: Zenodo. https://doi.org/10.5281/ZENODO.2575294. (12 April, 2021).

Eberhard, David M., Gary F. Simons & Charles D. Fenning (eds.). 2021. *Ethnologue: Languages of the World*. 24th edn. Dallas, Texas: SIL International. http://www.ethnologue.com.yale.idm.oclc.org.

Ecology Disrupted. 2021. *Ecology Disupted: Genetic Distance Values*. American Museum of Natural History.

ELAN. 2018. Njimegen: Max Planck Institute for Psycholinguistics.

Esposito, Christina M. 2010. The effects of linguistic experience on the perception of phonation. *Journal of Phonetics* 38(2). 306–316. https://doi.org/10.1016/j.wocn.2010.02.002. (17 May, 2021).

Evanini, Keelan, Stephen Isard & Mark Liberman. 2009. Automatic formant extraction for sociolinguistic analysis of large corpora. In *Tenth Annual Conference of the International Speech Communication Association*.

Evans, Nicholas. 2003a. *Bininj Gun-wok: a pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Canberra: Australian National University.

Evans, Nicholas (ed.). 2003b. *The Non-Pama-Nyungan Languages of Northern Australia: comparative studies of the continent's most linguistically complex region* (Studies in Language Change). Pacific Linguistics.

Excoffier, L, P E Smouse & J M Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131(2). 479–491. https://doi.org/10.1093/genetics/131.2.479. (26 December, 2021).

Excoffier, Laurent, G. Laval & S. Schneider. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1. 47–50.

Fabricius, Anne H., Dominic Watt & Daniel Ezra Johnson. 2009. A comparison of three speaker-intrinsic vowel formant frequency normalization algorithms for sociophonetics. *Language Variation and Change* 21(3). 413–435. https://doi.org/10.1017/S0954394509990160. (31 August, 2021).

Fletcher, Janet & Andrew R. Butcher. 2003. Local and global influences on vowel formants in three Australian languages. English. In *15th ICPhS Barcelona*, 905–908. OCLC: 781074463. Barcelona: Universitat Autónoma de Barcelona.

Fletcher, Janet & Andrew R. Butcher. 2014. Sound patterns of Australian languages. In Rachel Nordlinger & H. Koch (eds.), *The Languages and Linguistics of Australia: A Comprehensive Guide* (The World of Linguistics). De Gruyter.

Fletcher, Janet & Nicholas Evans. 2002. An acoustic phonetic analysis of intonational prominence in two Australian languages. *Journal of the International Phonetic Association* 32(2). 123–140. https://doi.org/10.1017/S0025100302001019. (7 June, 2019).

Fletcher, Janet, Nicholas Evans & Erich Round. 2002. Left-edge tonal events in Kayardild (Australian)-a typological perspective. In *Speech Prosody 2002, International Conference*.

Fletcher, Janet, Hywel Stoakes, Deborah Loakes & Ruth Singer. 2015. Accentual prominence and consonant lengthening and strengthening in Mawng. *ICPhS*. https://doi.org/10.1080/07268602.2015.1023169. (27 October, 2021).

Freeman, Valerie & Paul De Decker. 2021. Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *The Journal of the Acoustical Society of*

America 149(2). 1211–1223. https://doi.org/10.1121/10.0003529. (19 December, 2021).

Fry, Dennis B. 1958. Experiments in the perception of stress. *Language and speech* 1(2). 126–152.

Garellek, Marc. 2019. *The phonetics of voice*. Routledge Handbooks Online. https://doi.org/10.4324/9780429056253-5. (9 March, 2022).

Garellek, Marc & Patricia Keating. 2011. The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association* 41(2). 185–205. https://doi.org/10.1017/S0025100311000193. (17 May, 2021).

Garrett, Andrew & Keith Johnson. 2013. Phonetic bias in sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization*. Oxford: Oxford University Press.

Gasser, Emily & Claire Bowern. 2014. Revisiting Phonological Generalizations in Australian Languages. In, 11.

Glasgow, K. 1981. Burarra phonemes. In Bruce Waters (ed.), *Australian phonologies: collected papers*, vol. 15 (Working papers of SIL-AAB A), 63–89.

Goedemans, Rob. 2010. An overview of word stress in Australian Aboriginal languages. In Harry van der Hulst, Rob Goedemans & Ellen van Zanten (eds.), *A Survey of Word Accentual Patterns in the Languages of the World*. Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110198966.1.55. (23 March, 2021).

Gordon, Matthew. 2011. Stress systems. In John Goldsmith, Jason Riggle & Alan CL Yu (eds.), *The New Handbook of Phonology*, 141–163. Wiley-Blackwell. (14 April, 2017).

Gordon, Matthew. 2015. Consonant-tone interactions: A phonetic study of four indigenous languages of the Americas. In Heriberto Avelino, Matt Coler & Leo Wetzels (eds.), *The Phonetics and Phonology of Laryngeal Features in Native American Languages*. Leiden/Boston: Brill.

Gordon, Matthew & Timo Roettger. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3(1).

Gorman, Kyle, Jonathan Howell & Michael Wagner. 2011. Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics* 39(3). 192–193.

Grabowski, Emily & Laura McPherson. 2019. DAPPr: A (semi-)automated tool for pitch annotation. In *Proceedings of the 19th International Congress of the Phonetic Sciences*, 5.

Gray, RD & FM Jordan. 2000. Language trees support the express-train sequence of Austronesian expansion, 2000. *Nature* 405. 1052.

Gray, Russell D. & Quentin D. Atkinson. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426(6965). Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6965 Primary_atype: Research Publisher: Nature Publishing Group, 435–439. https://doi.org/10.1038/nature02029. (28 December, 2021).

Green, Ian. 1995. The Daly language family: a reassessment. Australian National University, Canberra.

Green, Rebecca. 1987. *A sketch grammar of Burarra*. Canberra: Australian National University Honors Thesis.

Gussenhoven, Carlos. 2004. The phonology of tone and intonation. Publisher: Cambridge University Press.

Hall, Nancy, Andie Niederecker, Elica Sue & Irene Orellana. 2019. Annotating Archival Recordings of Hocank (Winnebago). *Proceedings of the Annual Meetings on Phonology* 7. https://doi.org/10.3765/amp.v7i0.4489. (8 March, 2022).

Hamming, Richard W. 1950. Error detecting and error correcting codes. *The Bell system technical journal* 29(2). Publisher: Nokia Bell Labs, 147–160.

Harvey, Mark et al. 2003. An initial reconstruction of Proto Gunwinyguan phonology. In *The Non-Pama-Nyungan languages of northern Australia: Comparative studies of the continents most linguistically complex region*. Pacific Linguistics.

Hayes, Bruce. 1985. Iambic and trochaic rhythm in stress rules. *Proceedings of the XIth Annual Meeting of the Berkeley Linguistics Society*. 429–466.

Hayes, Bruce. 1995. *Metrical Stress theory*. University of Chicago Press.

Himmelmann, Nikolaus P. 2008. Prosody in language documentation. In *Essentials of language documentation*, 163–182. De Gruyter Mouton.

Hoffmann, Dorothea. 2015. *Documenting MalakMalak, an endangered language of Northern Australia*. http://hdl.handle.net/2196/00-0000-0000-000F-4832-4.

Holden, Clare Janaki & Ruth Mace. 2003. Spread of cattle led to the loss of matrilineal descent in Africa: a coevolutionary analysis. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270(1532). Publisher: Royal Society, 2425–2433. https://doi.org/10.1098/rspb.2003.2535. (28 December, 2021).

Hyman, Larry. 1977. On the nature of linguistic stress. In Larry M. Hyman (ed.), *Studies in stress and accent*, 37–82. Los Angeles: Department of Linguistics, University of Southern California.

Jepson, Kathleen, Janet Fletcher & Hywel Stoakes. 2019. Prosodically Conditioned Consonant Duration in Djambarrpuyŋu. *Language and Speech*. 002383091982660. https://doi.org/10.1177/0023830919826607. (23 March, 2021).

Jepson, Kathleen Margaret. 2019. *Prosody, prominence and segments in Djambarrpuyŋu* PhD Thesis.

Johnson, Keith. 2020. The ΔF method of vocal tract length normalization for vowels. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1). 10. https://doi.org/10.5334/labphon.196. (31 August, 2021).

Johnson, Lisa M, Marianna Di Paolo & Adrian Bell. 2018. Forced alignment for understudied language varieties: Testing Prosodylab-Aligner with Tongan data. Publisher: University of Hawaii Press.

Kager, Rene. 1999. *Optimality Theory* (Cambridge Textbooks in Linguistics). Cambridge University Press.

Kakadelis, Stephanie M. 2018. *Phonetic Properties of Oral Stops in Three Languages with No Voicing Distinction*. CUNY PhD.

Kaland, Constantijn. 2021. Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association*. 1–30. https://doi.org/10.1017/S0025100321000049. (19 April, 2021).

Keating, Patricia, Christina M. Esposito, Marc Garellek, Sameer ud Dowla Khan & Jianjing Kuang. 2010. Phonation Contrasts Across Languages. *UCLA Working Papers in Phonetics* (108). 188–202.

Kisler, Thomas, Florian Schiel & Han Sloetjes. 2012. Signal processing via web services: the use case WebMAUS. In *Digital Humanities Conference 2012*.

Kitchen, Andrew, Christopher Ehret, Shiferaw Assefa & Connie J. Mulligan. 2009. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proceedings of the Royal Society B: Biological Sciences* 276(1668). Publisher: Royal Society, 2703–2710. https://doi.org/10.1098/rspb.2009.0408. (28 December, 2021).

Kofod, Frances. 2013. *The painter's eye, the painter's voice: language, art and landscape in the Gija world*. http://hdl.handle.net/2196/00-0000-0000-0001-D032-0.

Kuznetsova, Alexandra, Per B Brockhoff & Rune HB Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of statistical software* 82(1). 1–26.

Ladd, D. R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Lake, Amelia, Juhyae Kim, Kassandra Haakman, Jeremiah Jewell, Sarah Babinski & Claire Bowern. 2022. *Accessibility, discoverability, and functionality of digital language archives*. Talk. Washington, D.C.

Lawyer, Lewis C. 2015. Patwin Phonemics, Phonetics, and Phonotactics. *International Journal of American Linguistics* 81(2). 221–260. https://doi.org/10.1086/680310. (17 May, 2021).

Lehiste, Ilse. 1970. *Suprasegmentals*. Publisher: Massachusetts Inst. of Technology P. Cambridge, MA: The MIT Press.

Lennes, Mietta. 2018. *Praat script*. http://phonetics.linguistics.ucla.edu/facilities/acoustic/praat.html.

Lewontin, R. C. 1972. The Apportionment of Human Diversity. In Theodosius Dobzhansky, Max K. Hecht & William C. Steere (eds.), *Evolutionary Biology*, 381–398. New York, NY: Springer US. https://doi.org/10.1007/978-1-4684-9063-3_14. (26 December, 2021).

Lobanov, B. M. 2005. Classification of Russian Vowels Spoken by Different Speakers. *The Journal of the Acoustical Society of America* 49(2B). Publisher: Acoustical Society of AmericaASA, 606. https://doi.org/10.1121/1.1912396. (31 August, 2021).

Lunden, Anya, Jessica Campbell, Mark Hutchens & Nick Kalivoda. 2017. Vowel-length contrasts and phonetic cues to stress: an investigation of their relation. *Phonology* 34(3). 565–580. https://doi.org/10.1017/S0952675717000288. (4 March, 2022).

Macaulay, Ben. 2021. The Race to Document Endangered Languages, Now That We Have the Technology. *Gizmodo*. https://gizmodo.com/the-race-to-document-endangered-languages-now-that-we-1847883858.

Macklin-Cordes, Jayden L., Claire Bowern & Erich R. Round. 2021. Phylogenetic signal in phonotactics. *Diachronica* 38(2). Publisher: John Benjamins, 210–258. https://doi.org/10.1075/dia.20004.mac. (28 December, 2021).

Mansfield, John. 2019. *Murrinhpatha morphology and phonology*. Vol. 653. Walter de Gruyter GmbH & Co KG.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2017. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In *Interspeech 2017*, 498–502. ISCA. https://doi.org/10.21437/Interspeech.2017-1386. (24 April, 2020).

McCarthy, John J. & Alan Prince. 1993. Generalized alignment. In *Yearbook of morphology 1993*, 79–153. Springer. (9 May, 2017).

McCloy, Daniel. 2012. *Semi-Auto Pitch Extractor*.

Michalakis, Yannis & Laurent Excoffier. 1996. A Generic Estimation of Population Subdivision Using Distances Between Alleles With Special Reference for Microsatellite Loci. *Genetics* 142(3). 1061–1064. https://doi.org/10.1093/genetics/142.3.1061. (26 December, 2021).

Nash, David George. 1980. *Topics in Warlpiri grammar.* Massachusetts Institute of Technology PhD Thesis.

Nearey, Terrance Michael. 1978. *Phonetic Feature Systems for Vowels*. Bloomington, Indiana: Indiana University Linguistics Club dissertation.

Nespor, Marina & Irene Vogel. 1986. *Prosodic Phonology* (Studies in Generative Grammar). Dordrecht: Foris Publications.

Nordlinger, Rachel. 2017. The languages of the Daly region (Northern Australia). Publisher: Oxford University Press.

O'Grady, Geoffrey N., C.F. Voegelin & F.M. Voegelin. 1966. Languages of the world: Indo-Pacific fascicle 6. *Anthropological Linguistics* 8(2). 1–199.

O'Keeffe, Isabel, C. Coleman, Ruth Singer, Linda Barwick, J. Mardbinda & T. Wilton. 2017. *Documentation of Kunbarlang*. http://hdl.handle.net/2196/00-0000-0000-000F-BF4E-0.

Ochshorn, RM & Max Hawkins. 2017. Gentle forced aligner. *github. com/lowerquality/gentle*.

Ohala, John J. 1993. The phonetics of sound change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*, 237–278. London: Longman.

Pentland, Christina. 2004. Stress in Warlpiri: Stress domains and word-level prosody.

Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English Intonation*. Cambridge, MA: MIT PhD Dissertation.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In J.L. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–158. John Benjamins.

Prince, Alan. 1990. Quantitative consequences of rhythmic organization. *Cls* 26(2). 355–398. (2 May, 2017).

R Core Development Team. 2020. *R: A language and environment for statistical computing*. Version 4.0.0. www.r-project.org.

Reddy, Sravana & James Stanford. 2015. A web application for automated dialect analysis. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 71–75.

Reid, Nicholas J. 1990. *Ngan-gityemerri: A Language of the Daly River Region, Northern territory of Australia*. Australian National University PhD Dissertation.

Reid, Nicholas J. 2013. *Talking Ngan'gi*. Digital collection managed by PARADISEC NR01. University of New England.

Rosenberg, Noah A., Jonathan K. Pritchard, James L. Weber, Howard M. Cann, Kenneth K. Kidd, Lev A. Zhivotovsky & Marcus W. Feldman. 2002. Genetic Structure of Human Populations. *Science* 298(5602). 2381–2385. https://doi.org/10.1126/science.1078311. (26 December, 2021).

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini & Jiahong Yuan. 2011. FAVE (forced alignment and vowel extraction) program suite. *URL http://fave. ling. upenn. edu*.

Ross, Belinda Britt. 2011. *Prosody and grammar in Dalabon and Kayardild* PhD Thesis.

Ross, Bella, Janet Fletcher & Rachel Nordlinger. 2016. The alignment of prosody and clausal structure in Dalabon. *Australian Journal of Linguistics* 36(1). Publisher: Taylor & Francis, 52–78.

Ross, Robert M., Simon J. Greenhill & Quentin D. Atkinson. 2013. Population structure and cultural geography of a folktale in Europe. *Proceedings of the Royal Society B: Biological Sciences* 280(1756). 20123065. https://doi.org/10.1098/rspb.2012.3065. (28 October, 2021).

Round, Erich. 2015. *Documentation of Kayardild*. http://hdl.handle.net/2196/00-0000-0000-0001-39A4-C.

Rzeszutek, Tom, Patrick E. Savage & Steven Brown. 2012. The structure of cross-cultural musical diversity. *Proceedings of the Royal Society B: Biological Sciences* 279(1733). 1606–1612. https://doi.org/10.1098/rspb.2011.1750. (28 October, 2021).

Sanker, Chelsea, Sarah Babinski, Roslyn Burns, Marisha Evans, Juhyae Kim, Slater Smith, Natalie Weber & Claire Bowern. 2021. (Don't) try this at home! The effects of recording devices and software on phonetic analysis. *Language* 97(4). lingbuzz/005748.

Savelyev, Alexander & Martine Robbeets. 2020. Bayesian phylolinguistics infers the internal structure and the time-depth of the Turkic language family. *Journal of Language Evolution* 5(1). 39–53. https://doi.org/10.1093/jole/lzz010. (28 December, 2021).

Schliep, Klaus Peter. 2011. Phangorn: phylogenetic analysis in R. *Bioinformatics* 27(4). Publisher: Oxford University Press, 592–593.

Senge, Chikako. 2016. A Grammar of Wanyjirra, a language of Northern Australia. 757.

Si, Aung. 2014. *Kune*. Digital collection managed by PARADISEC SI1. University of Melbourne.

Simard, Candide. 2010. *The prosodic contours of Jaminjung, a Northern Australian language*. University of Manchester PhD Dissertation.

Simard, Candide & Eva Schultze-Berndt. 2011. Documentary linguistics and prosodic evidence for the syntax of spoken language. In *Documenting Endangered Languages*, 151–176. De Gruyter Mouton.

Slatkin, Montgomery. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139(1). Publisher: Oxford University Press, 457–462.

Sokal, Robert R. & Charles D. Michener. 1958. A statistical method for evaluating systematic relationships. *The University of Kansas Science Bulletin* 38. 1409–1438.

Street, Chester. 1987. *An introduction to the language and culture of the Murrinh-Patha*. Darwin: Summer Institute of Linguistics.

Street, Chester. 2012. Murrinhpatha to English dictionary. Wadeye, Australia.

Tabain, Marija. 2016. Aspects of Arrernte prosody. *Journal of Phonetics* 59. Publisher: Elsevier, 1–22.

Tang, Kevin & Ryan Bennett. 2018. Contextual predictability influences word and morpheme duration in a morphologically complex language (Kaqchikel Mayan). *The Journal of the Acoustical Society of America* 144(2). 997–1017. https://doi.org/10.1121/1.5046095. (17 May, 2021).

Taylor, Peter & Joy Taylor. 1971. A tentative statement of Kitja phonology. In *Papers on the languages of Australian Aboriginals*, 100–109. AIAS.

Tuttle, Siri. 2003. Archival phonetics: Tone and stress in Tanana Athabaskan. *Anthropological Linguistics* 45. 316–336.

Van Heuven, Vincent J. 2018. Acoustic Correlates and Perceptual Cues of Word and Sentence Stress. In Rob Goedemans, Jeffrey Heinz & Harry Van der Hulst (eds.), *The Study of Word Stress and Accent: Theories, Methods, and Data*. Cambridge University Press.

Vogel, Irene, Angeliki Athanasopoulou & Nadya Pincus. 2016. Prominence, contrast, and the functional load hypothesis: An acoustic investigation. *Dimensions of phonological stress*. Publisher: Cambridge University Press Cambridge, England), 123–167.

Walsh, Michael J. 1976. *The Murinypata Language of North-West Australia*. Canberra: Australian National University PhD Dissertation.

Wedel, Andrew B. 2006. Exemplar models, evolution and language change. *The Linguistic Review* 23(3). https://doi.org/10.1515/TLR.2006.010. (23 February, 2017).

Whalen, D. H., Christian DiCanio & Rikker Dockum. 2020. Phonetic documentation in three collections: Topics and evolution. *Journal of the International Phonetic Association*. 1–27. https://doi.org/10.1017/S0025100320000079. (9 March, 2022).

Whalen, Douglas H. & Joyce M. McDonough. 2019. Under-researched languages: Phonetic results from language archives. In William F. Katz & Peter F. Assmann (eds.), *The Routledge Handbook of Phonetics*, 51–71. London/New York: Routledge.

Xu, Yi. 2011. Speech prosody: A methodological review.

Xu, Yi. 2019. Prosody, tone, and intonation. In *The Routledge handbook of phonetics*, 314–356. Routledge.

Yu, Alan C. L. 2010. Perceptual Compensation Is Correlated with Individuals' "Autistic" Traits: Implications for Models of Sound Change. *PLoS ONE* 5(8). e11950. https://doi.org/10.1371/journal.pone.0011950. (6 June, 2019).

Yu, Alan C.L. 2008. The phonetics of quantity alternation in Washo. *Journal of Phonetics* 36(3). 508–520. https://doi.org/10.1016/j.wocn.2007.10.004. (17 May, 2021).

Yu, Alan CL, Julian Grove, Martina Martinovic & Morgan Sonderegger. 2011. Effects of working memory capacity and "autistic" traits on phonotactic effects in speech perception. In *Proceedings of the International Congress of the Phonetic Sciences XVII,*

Hong Kong: International Congress of the Phonetic Sciences, 2236–2239. (16 February, 2017).

Zhang, Cong, Kathleen Jepson, Georg Lohfink & Amalia Arvaniti. 2021. Comparing acoustic analyses of speech data collected remotely. *LingBuzz preprint*. lingbuzz/005790.

Zhang, Jingwei. 2018. A Comparison of Tone Normalization Methods for Language Variation Research. *Information and Computation*. 9.