The Acoustic Correlates of word-level stress and focus-related prominence in Mankiyali

Jonathan Charles Paramore
jcparamo@ucsc.edu
University of California, Santa Cruz

Aurangzeb
zeb296@gmail.com
Independent Researcher

Abstract

This paper investigates the acoustic correlates of word-level stress and phrase-level focus-related prominence in Mankiyali, a highly endangered Indo-Aryan language spoken in Northwest Pakistan that utilizes a weight-sensitive stress system. Of the acoustic properties measured (duration, f0, intensity, spectral tilt, and vowel quality), duration was the only feature found to robustly and consistently correlate with word-level stress across syllable types. In contrast, phrase-level focus-related prominence corresponded to an amplification of all five acoustic features measured. Given that vowel duration serves a vital role in preserving lexical contrast in Mankiyali, these findings present difficulties for a strong version of the Functional Load Hypothesis, which claims that acoustic properties bearing a high functional load in a language will not be used to mark prominence. In addition, results support an analysis of Mankiyali's stress system as having five distinct levels of weight, a pattern which is extremely rare, if not unattested, elsewhere in the world's languages.

1 Introduction

Mankiyali is an endangered and understudied language situated in the Khyber Pakhtunkhwa Province of Northwest Pakistan spoken by about 500 people. Impressionistic descriptions state that the language utilizes a complex weight-sensitive stress system (Paramore, 2021). Many studies have covered the subject of stress in the world's languages, and the use of quantitative measures to determine the acoustic correlates of stress – which most notably began with Fry's (1955) examination of acoustic correlates of word-level stress in English – has increased significantly in recent years. However, recent work suggests that many phonological descriptions of word-level stress that we do possess – including Fry's study of English – may unintentionally confound descriptions of prominence at the phrase level with prominence at the word level (Gordon, 2011; Gordon, 2014; Gordon & Roettger, 2021). The present study intentionally teases apart word-level and phrase-level prominence in its experimental design, thereby providing much-needed insights into how the manifestation of prominence at these two separate prosodic levels might differ in Mankiyali. Additionally, this paper provides quantitative evidence supporting the analysis of a 5-tier weight-sensitive stress scale in Mankiyali, which, if correct, has important implications for theories of syllable weight.

Beyond enhancing our understanding of the nature of stress in Northwest Indo-Aryan languages and providing quantitative evidence of the stress system, Mankiyali provides fertile ground for the investigation of other current issues surrounding the acoustic analysis of stress. To begin with, Mankiyali possesses a large number of phonemic vowels, along with phonemic distinctions in length. Since duration is perhaps the most persistent acoustic correlate of stress in many of the world's languages (Van Heuven & Turk, 2021), the question arises as to whether a language like Mankiyali can call upon duration to indicate stress if the language's phonological system already utilizes the acoustic feature for other means. There is no consensus in the literature on this issue. While Berinstein (1979), Hayes (1995), Gordon & Applebaum (2010), Vogel et al. (2016), and others assert that the use of acoustic properties as contrastive features in the lexicon precludes their use as reliable phonetic markers of stress since their use would obscure lexical contrasts – a phenomenon known as the Functional Load Hypothesis (FLH) – Van Heuven & Turk (2021) conclude that empirical evidence from Swedish (Heldner & Strangert, 2001) suggests otherwise. Similarly, Lunden et al. (2017) did not find any evidence to back up the claims of the FLH in relation to duration in their survey of 140 languages. Nakai et. al. (2009, 2012) take a middle ground regarding the FLH, presenting data on utterance-final lengthening in Northern Finnish to show that languages with phonemic vowel length contrasts enforce a "durational ceiling effect" that limits the magnitude of non-phonemic lengthening that can occur. Considering the disparate views on this topic, exploring the status of duration as a reliable acoustic correlate in Mankiyali brings more data and thus further clarity to the debate.

Another topic that has recently received significant attention regarding stress correlates is the role of f0 as a reliable acoustic correlate of word-level stress. Setting aside lexical tone and pitch-accent languages – both of which utilize f0 as a primary correlate to word-level phenomena in different ways - most studies that have examined the acoustic correlates of stress at the word level over the past several decades have included an analysis of f0, and it is often considered one of the primary acoustic correlates to word-level stress (Gordon, 2004 (Chickasaw); Gordon & Applebaum, 2010 (Turkish Karbadian); Garellek & White, 2015 (Tongan); a.o.). However, as mentioned above, several recent studies argue that it is likely that many, if not most, acoustic analyses (including the three studies just cited) do not disentangle word-level stress from phrase-level prominence effects, and f0 is often a primary acoustic correlate of phrase-level intonational events rather than of word-level stress (Ortega-Llebaria & Prieto, 2011; Bruggeman et al., 2020; Gordon & Roettger, 2021; Van Heuven & Turk, 2021). If this is true and the acoustic correlates of these two levels of prominence are often conflated, more studies that control for this conflation are needed.

One final question this paper aims to address is whether any acoustic properties that might correlate with stress, as well as the magnitude of the effect of stress on those properties, differ across syllable types. Yakup & Sereno (2016), in their acoustic study of Uyghur stress, found that stressed/unstressed {CV} syllables showed a higher contrast in duration but a lower contrast in f0 compared to stressed/unstressed {CVC} syllables. This finding indicates that, at least in Uyghur, syllable type may influence the patterning of acoustic correlates of stress between {CV} and {CVC} syllables. With that said, while the Uyghur target words were placed in the non-final position of a sentence frame in this study, they were likely marked with contrastive (phrasal) prominence correlates. As such, the claim that the effect sizes of word-level stress correlates changes across syllable types in Uyghur was potentially confounded with phrasal prominence. In this study, we hope to examine the potential effect of syllable type on phonetic correlates without the confound of focus-related prominence by comparing the magnitude of the change between stressed and unstressed syllables in non-focused positions across all five syllable types relevant to the Mankiyali stress criterion: {CV, CVC, CVCC, CVː, CVːC}. Mankiyali is an apt language in which to explore the impact of syllable type on acoustic correlates, as it is a weight-sensitive stress system reported to have at least a ternary distinction in syllable weight, {CVː(C)} > {CVC(C)} > {CV}, and possibly a five-way distinction.

Notably, the relationship of {CVːC} to {CVː} in terms of weight, as well as the status of {CVCC} in relation to {CVC}, are not clarified in the sole impressionistic description of the Mankiyali stress system (Paramore, 2021). One of the goals of this paper is to explore, using quantitative data, whether these two syllable weight distinctions are warranted. Specifically, this study assumes that if a syllable is phonologically heavier, its attraction of stress will be reflected in greater enhancement of the language-specific phonetic correlates associated with stress (e.g., increased duration, greater intensity, pitch, etc.). Under this assumption, we test whether {CVːC} patterns as heavier than {CVː} by examining whether the presence of a {CVːC} syllable in the same word as {CVː} results in phonetic evidence for stress shifting from {CVː}.[1] For instance, in a word like [ˈ*bẽː.ˌɣĩː*] 'rooster.M.PL' in which both syllables are {CVː}, stress impressionistically falls on the penultimate syllable by default. However, if the final syllable is superheavy {CVːC} – as in [ˌ*bẽː.ˈɣĩːz*] 'rooster.M.PL-DAT' – stress will shift to the final syllable if {CVːC} outweighs {CVː}. By measuring and comparing the acoustic properties of the penultimate syllables in [ˈ*bẽː.ˌɣĩː*] vs. [ˌ*bẽː.ˈɣĩːz*], we can examine whether the acoustic changes correlating with stress changes in known weight relationships are also present in these words. If the penultimate syllable in [ˈ*bẽː.ˌɣĩː*] is found to exhibit amplified phonetic properties compared to the penultimate syllable in [ˌ*bẽː.ˈɣĩːz*], and the acoustic changes are similar to other weight relationships in the language, this provides good evidence that {CVːC} does indeed outweigh {CVː} for stress. More broadly, if heavier syllables consistently show greater acoustic enhancement under stress, then comparing stressed and unstressed syllables across syllable types offers a direct quantitative test of the hypothesized weight hierarchy in Mankiyali.

The structure of the paper proceeds as follows. In Section 2, we provide a general background for the quantitative study. Sections 2.1 and 2.2 outline previous cross-linguistic work on both word-level stress and phrase-level focus-related prominence, and Section 2.3 supplies an overview of previous acoustic studies of word-level stress and focus-related prominence in Indo-Aryan languages. Section 2.4 gives a brief sketch of the relevant aspects of Mankiyali phonology. A detailed outline of the methodology for the study is given in Section 3, including a description of participants, speech materials, elicitation procedure, and the acoustic and statistical methods used for data analysis. The experiment results are discussed in Section 4, and Section 5 considers the implications of these results. Concluding remarks are given in Section 6.

2 Background

2.1 Previous research on word-level stress and phrase-level focus-related prominence

The usual acoustic suspects that correlate with stress at the word level include some combination of increased duration, intensity, spectral tilt, spectral expansion/vowel quality, and, to a lesser extent, f0 (Van Heuven, 2018). The primary acoustic correlate of phrase-level, focus-related

[1] We were unable to find near minimal pairs to quantitatively examine the weight relationship between {CVCC} and {CVC} syllables. Thus, any potential weight distinctions between these two syllable types can only be infered from the phonological weight patterns of other syllable types in the language. See section 5.5 for further discussion.

prominence, on the other hand, is generally thought to be a change in f0, which is more directly related to intonational pitch accents associated with phrasally stressed syllables (Van Heuven & Turk, 2021). This change can take the form of a rising f0, a falling f0, or a specific f0 contour. Phrase-level focus-related prominence often falls on the primary stressed syllable of the focused word in a phrase (see Gordon (2014) for a discussion of languages that do not fit this pattern). The term "focus" is used ubiquitously in the literature to refer to a pragmatic concept in which alternatives relevant for the interpretation of a linguistic expression are introduced into a discourse (e.g., Rooth, 1992; Kügler & Calhoun, 2021). Oftentimes, focused constituents are highlighted at the phrase level with some type of prosodic marking, though other methods such as word order are sometimes employed as additional focus markers in languages like Urdu (Jabeen, 2022). Types of focus include the introduction of new information into a conversation, contrastive focus that highlights a word as contrary to its alternatives, emphasis on a constituent as corresponding to the wh-phrase in a wh-question, and a type of broad focus that applies to large portions of or even whole sentences (Halliday, 1967b, pp. 204-207; Ladd, 1980; Féry, 2013). As will be evident in Section 3, based on our experimental design, the relevant types of focus for this paper include an individual lexical item acting as some combination of new and/or contrastive information, which is typically described as contrastive, narrow focus. *Narrow focus* refers to a technique used by a speaker to "pragmatically single out" information that is important or new in a discourse (Sbranna et. al., 2023). Conversely, *broad focus* does not emphasize an individual portion of an utterance but rather the entire utterance as a unit.

Importantly, narrow focus is usually highlighted by a specific pitch accent corresponding to a change in f0 or an f0 contour, the characteristics of which are usually language-dependent (Jackendoff, 1972). Some languages, including certain varieties of Italian, use a falling f0 to indicate different types of narrow focus (Sbranna et. al., 2023). Other languages (e.g. Hungarian) rely on a specific f0 contour to highlight the narrowly focused constituent (Vogel et. al., 2016, p.149). While no comprehensive study of the Mankiyali intonation system has been conducted to date, several other Indo-Aryan languages mark focus with a rising f0. For instance, in Hindi/Urdu, Harnsberger (1994) shows that focused constituents exhibit a marked rise in f0. Similarly, Jabeen (2017) found that preverbal nouns with narrow focus in Urdu are optionally marked with a rising f0 across the length of the word. Additionally, Lahiri & Fitzpatrick-Cole (1999) showed that constituents with narrow focus in Bengali are consistently realized with a rise in f0.

2.2 Separating word-level stress and phrase-level focus-related prominence

To disentangle word-level stress from focus-related prominence, one must study word-level stress on words in a position without phrasal accent. Target words that appear in a non-focused position of a sentence ideally provide us with a word that is only marked by word-level stress, without the changes in f0 often associated with focus-related prominence described above. Moreover, it is important to place these non-focused words away from phrasal boundaries, which often carry their own variations in pitch that could influence a measurement of word-level stress correlates (Gordon, 2014).

In contrast, target words receiving narrow focus in a phrase are associated with both word-level stress *and* focus-related prominence. In this study, we can compare the acoustic properties of a syllable in a non-final, non-focused position with the same syllable in a non-final, *focused* position to determine what acoustic properties only intensify when a word receives narrow focus; these properties will likely be our acoustic correlates of focus-related prominence. With that said, because not much is known about phrase-level intonation in Mankiyali, it is possible that other confounding effects may still be present, such as f0 downtrend, phrase-level compression effects, or other types of intonational boundary tones. While we attempt to rule these potential confounds out, a thorough examination of phrase-level intonation in Mankiyali is still needed to confirm our results for focus-related prominence.

2.3 Acoustic analyses of word-level stress and focus-related prominence in Indo-Aryan

To our knowledge, a quantitative analysis of stress in Mankiyali, or any phonetic analysis of the language, has never been undertaken. The only literature on the phonology of Mankiyali is the initial impressionistic description of the language's sound system in Paramore (2021). Moreover, quantitative analyses of stress systems for other languages in Pakistan also remain scant. As Ohala (1991) notes, quantitative research on stress in Indo-Aryan languages – which are spoken as first languages by 80% of Pakistanis (Cardona & Jain, 2003) – remains in its infancy compared to phonetic research in other regions of the world, and Abbasi et al. (2018) suggest that Ohala's evaluation still holds true today. Though approximately 80 indigenous languages are spoken in Pakistan, most studies on the acoustic correlates of word-level stress in the region are confined to Hindi-Urdu (Ohala, 1986; Hussain, 1997; Dyrud, 2001; Nair et al., 2001; Mumtaz et al., 2020). The limited number of non-Hindi-Urdu studies include Balti (Caplow, 2016), Gujarati (Shih, 2018;

Bowers, 2019), Punjabi (Kiranpreet, 2016), Sindhi (Abbasi & Hussain, 2015), and Wakhi (Ivanov & Silanteva, 2020). Two of these languages do not belong to the Indo-Aryan family but are spoken in Northern Pakistan (Wakhi is Iranian, and Balti is Tibeto-Burman).

Unfortunately, outside of Dyrud's (2001) phonetic analysis of Hindi-Urdu, all the above-mentioned studies seemingly confound word-level stress with various types of phrasal prominence by analyzing the acoustics of target words either in the focus position of a carrier sentence or in isolation. Furthermore, though Hindi-Urdu boasts a weight-sensitive stress system, Dyrud's study of stress in the language focuses on a small number of minimal pairs in which stress is lexically contrastive rather than weight-sensitive. This means that out of approximately 80 languages currently in use in Pakistan, perhaps only a single study exists that has successfully analyzes the acoustic correlates of word-level stress in a language in the region, and that study focuses on stress as a lexically contrastive linguistic feature rather than analyzing stress in a weight-sensitive system. Suffice it to say that, at best, the reliable data available to us on the acoustic correlates of weight-sensitive stress in Indo-Aryan – and on the languages of Pakistan in particular – is severely impoverished.

Concerning phrase-level prominence, while nothing has been published on Mankiyali phrasal prosody to date, several studies on other Indo-Aryan languages have been conducted examining focus-related prominence. As mentioned above, Harnsberger (1994) studied focus marking on nouns in Hindi/Urdu and concluded that narrow focus is correlated with a LH rise in f0 on the focused constituent. Other work on focus marking in Hindi/Urdu found somewhat similar results but with more variation in the realization of the accent associated with narrow focus. That is, each argument in simple SOV declarative sentences in Hindi/Urdu - similar in shape to the carrier sentences used in the present study on Mankiyali - have been found to act as Minor intonational Phrases (MiP) with a low boundary tone marking the left edge and high boundary tone marking the right edge (Jabeen, 2019; Jabeen & Delais-Roussarie, 2019). Jabeen & Delais-Roussarie (2020) conducted another study examining narrow focus in Urdu and found that words with narrow focus are marked with a rising f0 that is distinct from these MiP boundary markers, though Jabeen (2017) reports that this narrow focus marker is optional in some cases. Additionally, Butt & King (1996b) found that locating a word with narrow focus in the preverbal position in Urdu produces more consistent f0 peaks than when placed post-verbally, indicating that the position of a focused constituent is connected to its acoustic correlates to some degree. In sum, focus-related prominence in Hindi/Urdu generally correlates with a rising f0.

Work on Bengali (Khan, 2008, 2014) describes an inventory of various pitch excursions at phrase boundaries and a significant rise in f0 on several different kinds of focused elements. In Tamil (Keane, 2007), a Dravidian language, accentual (AP) and intonational (IP) phrases are produced with edge-aligned tonal excursions and an extra AP rising f0 contour on words with narrow focus. To sum up, work on Indo-Aryan and South Asian prosody highlights the consistent presence of phrase-level intonational activity and various degrees of rising f0 on constituents in narrow focus.

As an SOV language with similar flexibility of word movement to languages like Urdu, it is possible that words in pre-verbal position exhibit a more consistent revelation of the acoustic correlates of focus-related prominence in Mankiyali. As such, our stimuli are all elicited in the preverbal position. Additionally, as f0 changes have been shown to act as the primary correlate of narrow focus in other Indo-Aryan languages, we hypothesize that this will be the case in Mankiyali as well and place the target words sentence-medially to avoid potential f0 activity associated with different phrasal boundaries that could be present in the language. However, it should be noted that sentence-medial placement does not guarantee the absence of phrase boundaries, particularly in a language like Mankiyali whose intonational structure remains unstudied. Thus, while sentence-medial position reduces the likelihood of boundary-related f0 movement, it cannot eliminate it entirely.

2.3 Mankiyali phonology

A treatment of Mankiyali phonology can be found in Paramore (2021). What follows is a summary of the relevant portions of that work. Mankiyali has a 15-vowel system, as depicted in table 1. Vowels are phonemically distinguished in terms of their length, height, backness, tenseness, and nasality. Crucially, vowel length is a distinctive factor for four places of articulation in the language (/ʌ/ is analyzed as the short counterpart to /ɑː/). Due to the robust utilization of duration in the vocalic inventory, the strict version of the Functional Load Hypothesis mentioned above would predict that the use of duration as an acoustic correlate of stress will be avoided.

Table 1: Mankiyali Vocalic Inventory

		Front			Central			Back		
		short	long	long nasal	short	long	long nasal	short	long	long nasal
High	tense	i	iː	ĩː				u	uː	ũː
	lax	ɪ	ɪː	ɪ̃ː						
Mid	tense		eː						oː	õː
	lax				ʌ					
Low	tense								ɑː	ɑ̃ː
	lax									

Five syllable types in Mankiyali are relevant for determining stress placement: {CV, CVC, CVCC, CVː, CVːC}. Stress in Mankiyali, based on native-speaker judgments, is weight-sensitive and unbounded in that the heaviest syllable in a word attracts primary stress regardless of its position in the word. Moreover, Mankiyali possesses at least a ternary distinction in syllable weight, as demonstrated by the data in (1).

(1) {CVː(C)} > {CVC(C)} > {CV} (Paramore, 2021, p.44-47)

a. Default penultimate stress
   i. [ˌkʌ.mʌ.ˈkʌ.l-ʌ]
      stupid-M.SG
   ii. [ˈzɑːn.ˌdɑːk]
      child.F.SG
b. {CVC(C)} > {CV}
   i. [ˌzo.rʌ.ˈʋʌr]
      forceful person.M/F.SG
   ii. [ˈbʌŋg.su.ˌʋ-ʌ]
      buckle-M.SG
c. {CVː(C)(C)} > {CVC(C)}
   i. [muk.ˈl-eː]
      open-IMP
   ii. [sʌŋg-.ˈtoːb]
      friend-NMLZ
d. Tie for heaviest syllable
   i. [ˌkɑː.ˈleː.z-ʌ]
      liver-M.SG
   ii. [ˈlʌkʰ.ser]
      many

When syllable weight is neutral in Mankiyali – where all syllables in a word are of the same weight – primary stress falls on the penultimate syllable (1a). As is evident from the scale, syllables with long vowels, {CVː(C)}, are the heaviest syllable type (1c), and {CVC(C)} syllables represent the middle level of the hierarchy (1b). Syllables with a short vowel and no coda occupy the lightest position on the scale and only receive primary stress if no other syllable type is present in a word. If there is a tie for the heaviest syllable in a word between two or more syllables, primary stress falls on the rightmost non-final occurrence (1d). Secondary stress does occur in the language and falls on all CV:(C) syllables that do not receive primary stress (1aii, di). Short vowel syllables only receive secondary stress to avoid a stress lapse, two adjacent unstressed syllables (1ai, bi-ii).

Paramore (2021) notes that the status of {CVCC} in relation to {CVC} and the status of {CVːC} in relation to {CVː} are uncertain in terms of weight distinctions for stress. Most speakers provide impressionistic judgments that primary stress falls on the penultimate syllable in words with two {CVː} syllables like [ˈkɑː.ˌɣɑː] 'crows', and they also judge that primary stress shifts to the word-final syllable when the final syllable is {CVːC} in words like [ˌkɑː.ˈɣɑːz] 'paper'. However, a sizeable minority of speakers provide inconsistent, variable judgments as to whether stress shifts to the word-final position in words of this shape. Regarding the distinction between {CVCC} and {CVC}, Paramore (2021, p.44) cites a single example demonstrating that {CVCC} draws stress from a penultimate {CVC}: [kʌɽ.ˈsʌŋg] 'a huge heap'. However, this example is a compound word (cf. Urdu [gɑɽ] 'jumble' and [sʌŋg] 'together'), and stress placement in compounds often manifests distinct patterns that deviate from a language's standard stress system.

Our own intuition is that word-final syllables consistently attract primary stress to the word-final position in words like [ˌkɑː.ˈɣɑːz], meaning that {CVːC} is treated as heavier than {CVː} for stress. Furthermore, in most languages in which {CVːC} outweighs {CVː}, {CVCC} (when attested) is also found to outweigh {CVC} (e.g., Kelkar' (1968) examination of Hindi). Based on this, we adopt the strongest possible version of Mankiyali's weight-sensitive stress system, hypothesizing five distinct weight levels, as shown in (2). To test this hypothesize, we measure and compare the acoustic properties of penultimate syllables when they bear stress versus when stress is attracted to a heavier final syllable. For example, in [ˈmʌ.zʌ], both syllables are {CV}, and the penult is unambiguously stressed. In [mʌ.ˈzʌɽ], on the other hand, we hypothesize that the final {CVC} syllable attracts stress from the penult {CV}. If correct, the penult {CV} in [ˈmʌ.zʌ] should show greater acoustic enhancement than the same syllable in [mʌ.ˈzʌɽ], where it is unstressed. Systematically comparing such minimal pairs across syllable types provides a quantitative means of evaluating the proposed five-level syllable weight hierarchy.

(2) {CVːC} > {CVː} > {CVCC} > {CVC} > {CV}

As discussed in Section 5.4, we argue that the quantitative results of this study support our hypothesis that Mankiyali's stress system utilizes five discrete levels of syllable weight.

2.4 Predictions

Our experiment aims to examine two aspects of the acoustic correlates of word-level stress. First, we look for acoustic contrasts between stressed and unstressed syllables of the same syllable type. Specifically, we investigate the contrasts of all five syllable types by comparing the acoustic properties of these syllables in a presumably primary stress position to the acoustic properties of the same syllable type in a presumably non-primary stress position. These contrasts should elucidate whether primary-stressed syllables exhibit a significant strengthening of specific acoustic correlates compared to their unstressed counterparts. Our expectation, based on the large body of literature devoted to this topic, some of which were enumerated above, is that primary-stressed syllables will exhibit some combination of increased duration, greater intensity, a flatter spectral tilt, and an expanded vowel space in comparison to unstressed syllables in the same environment. Furthermore, we expect no significant difference in f0 for syllables with word-level stress in a non-focused position of the phrase, as increased f0 is more likely a correlate of narrow focus rather than word-level stress.

Second, we compare the most prominent acoustic correlates of the five different syllable types analyzed in the study to determine the extent to which syllable types differ in their use of specific acoustic correlates. Our prediction is that, just as the use of acoustic correlates used to signify other phonological phenomena will not mitigate their use as acoustic correlates of stress, so too will the acoustic correlates of stress remain unaffected by changes in syllable type. Our reasoning is that for stress to be reliably perceived, its acoustic correlates will remain relatively stable across different contexts within a language. This stability ensures that listeners can consistently detect and interpret stress patterns, regardless of variations in syllable type. Thus, we predict that the only interaction between the acoustic correlates of stress and syllable type will be the magnitude of absolute durational differences between stressed/unstressed short vowels ({CV, CVC, CVCC}) and stressed/unstressed long vowels ({CVː, CVːC}). The reason for this predicted interaction is that long vowels require a greater absolute change in duration between stressed and unstressed syllables compared to short vowels for the change to be reliably heard by listeners (Stevens, 2000; Lunden, 2013). As such, we predict that long-vowel syllables will exhibit greater durational changes as a function of stress compared to short-vowel syllables.

That said, native speaker judgments indicate that all long vowels ({CV:(C)}) receive secondary stress if they do not bear primary stress. Because secondary stress adds a degree of prominence, the durational contrast between primary- and secondary-stressed long vowels may be attenuated. In contrast, short vowels are either fully stressed or entirely unstressed. As a result, comparisons involving long vowels reflect contrasts between primary and secondary stress, while comparisons involving short vowels reflect the more distinct contrast between stressed and unstressed syllables. This asymmetry may obscure the expected durational enhancement in long-vowel syllables—which is precisely what we observe in the results.

Another facet of this study pertains to prominence related to contrastive, narrow focus. We compare the acoustic correlates of stressed syllables in narrow focus to stressed non-focused syllables, where the syllable should be devoid of focus-related prominence as well as pitch excursions commonly present at phrasal boundaries. By so doing, we can ascertain the acoustic correlates of focus-related prominence. Given that narrow focus has consistently been found to correlate with a marked change in f0 in Indo-Aryan languages, we anticipate that Mankiyali focus-related prominence will correlate with a significant change in f0.

3 Methodology

3.1 Participants

Thirty male native speakers of Mankiyali participated in the study. 27 participants live in Danna or Dameka, two remote mountain villages in Mansehra District, Khyber Pakhtunkhwa, Pakistan, located about 1 kilometer from each other. Three of the participants live in larger cities for work most of the year but grew up speaking Mankiyali and continue to speak with family regularly. All Mankiyali speakers are at least trilingual in Hindko, the predominant language of the region, and Urdu, the national language of Pakistan. Many speak Pashto and/or English as well. The age of every participant ($\mu = 32$, sd = 7.78) was recorded at the time of elicitation.

3.2 Speech materials

As laid out in section 2.3, the default stress position in Mankiyali falls on the penultimate syllable. Thus, we targeted the penultimate syllable of (mostly) disyllabic words for our acoustic analysis. However, we targeted the final syllable for our analysis of {CV:C} syllables since finding the necessary tokens in which to analyze this syllable type with both primary and secondary stress in the penultimate position is impossible. While this approach may be confounded by word-final lengthening effects, no other approach for analyzing {CVːC} is available. This falls out from the fact that we hypothesize {CVːC} to be the heaviest available syllable type in Mankiyali's stress system, so there is presumably never a case when {CVːC} occurs in the penultimate position in its unstressed form. Consequently, to derive both primary-stressed and secondary-stressed versions of {CVːC} in the same environment, we analyze it in word-final position in both cases.

The complete list of tokens is provided in Table 2. Disyllabic/Trisyllabic words with the target syllable in the penultimate position were grouped into near minimal pairs, one of which presumes the target syllable receives primary stress, the other of which presumes the target syllable either receives no stress or secondary stress (depending on syllable type). For languages with contrastive lexical stress, finding identical words in which to compare stressed and unstressed iterations of a syllable is relatively simple. For a weight-sensitive stress language like Mankiyali, on the other hand, primary stress always falls on the heaviest syllable, so finding exact minimal pairs in which to compare the acoustic correlates is not possible. As a result, we settle for near minimal pairs. This approach still enables us to evaluate the acoustic properties of each syllable type, both in stressed and unstressed conditions. By utilizing highly similar segmental environments in our near minimal pairs, we control for other factors, such as coarticulatory effects and inherent differences in acoustic properties across vowel qualities, that could impact the target syllable's acoustic features. Controlling for the neighboring consonants and target vowel qualities in this way enables us to examine whether the PRIMARY VS. SECONDARY/UNSTRESSED factor produces measurably significant acoustic differences.

Table 2: Tokens consist of 25 sets of near minimal pairs (5 sets for each syllable type) differing in stress position. The target syllable for each token is bolded.[2]

CV		**CVC**		**CVː**		**CVCC**		**CVːC**	
ˈ**kʰʌ**.bʌ	**kʰʌ**.ˈbɑːɽ	ˈ**gʌɽ**.ku	**gʌɽ**.ˈkuː	ˈ**bẽː**.ˌɣĩː	ˌ**bẽː**.ˈɣĩːz	ˈ**sʌŋg**.jo	**sʌŋg**.ˈtoːb	lʌŋ.ˈ**gɑːr**	ˈɖɑːŋ.ˌ**gɑːr**
ˈ**mʌ**.zʌ	**mʌ**.ˈzʌɽ	ˈ**mus**.ki	**mus**.ˈkiː	ˈ**peː**.ˌkiː	ˌ**peː**.ˈkiːz	ˈ**gʌnd**.jo	**gʌnd**.ˈgiː	kʌ.ˈ**dɑːr**	ˈkɑːr.ˌ**dɑːr**
ˈ**kʌ**.t͡sʌ	**kʌ**.ˈt͡sʌɽ	ˈ**bel**.ti	**bel**.ˈtiːz	ˈ**kɑː**.ˌɣɑː	ˌ**kɑː**.ˈɣɑːz	ˈ**mist**.ri.ˌjo	**mist**.ˈriː	go.ˈ**dɑːɽ**	ˈdʒoːn.ˌ**dɑːr**
ˈ**jʌ**.kʌ	**jʌ**.ˈkʌɽ	ˈ**kut**.re	**kut**.ˈreːz	ˈ**ɖeː**.ˌkiː	ˌ**ɖeː**.ˈkiːz	ˈ**ist**.ri.ˌjo	**ist**.ˈriː	mʌz.ˈ**dɑːr**	ˈdɑːɣ.ˌ**dɑːr**
ˈ**t͡su**.ki	**t͡su**.ˈkiːɳ	ˈ**pʌtʰ**.re	**pʌtʰ**.ˈreːz	ˈ**t͡ʃeː**.ˌbiː	ˌ**t͡ʃeː**.ˈbiːz	ˈ**d͡ʒʌnt**.ri.ˌjo	**d͡ʒʌnt**.ˈriː	bʌz.ˈ**vɑːɳ**	ˈɑːz.ˌ**vɑːɳ**

[2] Complete glosses of all tokens are given in the appendix. A threeway stop contrast exists in Mankiyali between voiceless aspirated, plain voiceless, and voiced stops. Every phonologically aspirated stop in the token set is indicated with the IPA aspiration diacritic: <ʰ>.

In attempting to analyze the acoustic correlates of a light {CV} syllable, the first word in the near-minimal pair takes the shape {ˈCV.CV}, with stress falling on the penultimate {CV}. To change the position of stress in the second word, we rely on a word with a heavier second syllable, e.g., {CV.ˈCVC}, to draw stress away from the default penultimate position. In both words, the acoustic properties of the penultimate {CV} are analyzed for comparison. The analysis of {CV}, {CVC}, {CVCC}, and {CVː} syllable types was approached in this way. For {CVːC}, we targeted the word-final syllable for the reasons discussed above.

To reduce the possibility of confounding the phonetic correlates of word-level stress with other phonetic correlates, tokens were embedded in a carrier sentence, which made up part of a three-sentence mini-monologue, given in (3). The carrier sentence was identical across tokens. A similar method used to elicit tokens in the non-focus position was first used by Bouchhioua's (2008) analysis of stress in disyllabic tokens in Tunisian Arabic and then subsequently by Almbark et al. (2014) and Bruggeman et al. (2020), and each study reports success in terms of deriving the intended changes in focus. Each mini-monologue was elicited in the order provided in (3) as a whole unit, and participants were instructed to read the units as naturally as possible, pausing briefly between mini-monologues. The second author, a native speaker of Mankiyali, presided over the elicitations and prompted participants to repeat a mini-monologue if speech was unnatural. Since Mankiyali does not have an official writing system, the mini-monologues were presented to participants in the Urdu script, which is often used informally by the Mankiyali speech community.

(3) Three-sentence "Mini Monologues" used in the experiment

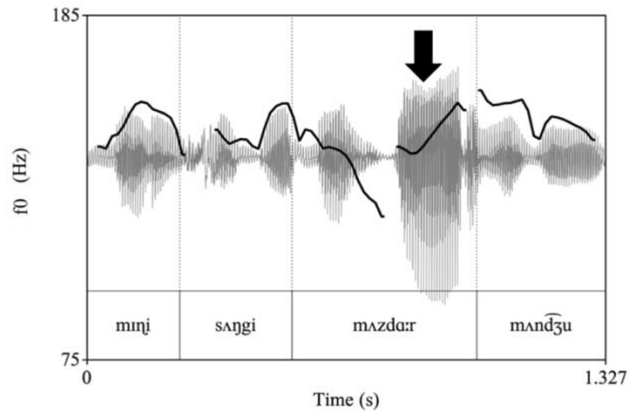
Sentence 1	*Sentence 2*	*Sentence 3*
mɪɳi sʌŋgi __ mʌnd͡ʒu	mɪɳi sʌŋgi du ʋʌɾʌ __ mʌnd͡ʒu	mɪɳi sʌŋgi t͡soːr ʋʌɾʌ __ mʌnd͡ʒu
My friend __ said	My friend two times __ said	My friend four times __ said
"My friend said __"	"My friend said __ two times"	"My friend said __ four times"

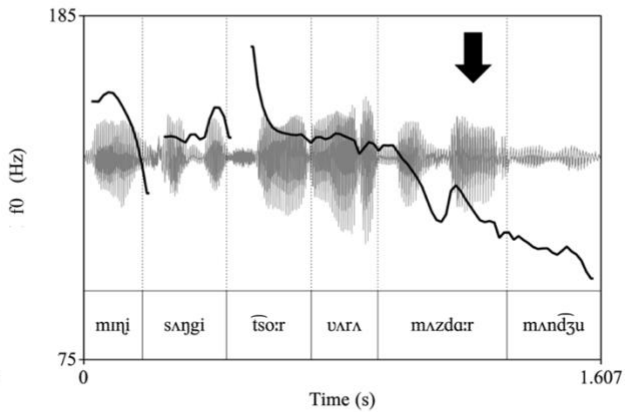
Embedding the tokens in this construction enables us to isolate the specific acoustic correlates we hope to analyze because the context procured via the mini-monologues disentangles focus-related phrasal prominence from word-level stress. In sentence 1 in (3), the token is new information and thus takes narrow focus in the sentence. We used the tokens in this sentence to analyze focus-related prominence. Sentence 2 sets up the context for the token to appear in non-focused positions in sentence 3; the tokens in sentence 2 were not used in the analysis. The focus in sentence 3 is likely on [*t͡so:r*] because of the pragmatic context provided by sentence 2. Thus, tokens are non-focused in sentence 3; this is the token in which we analyzed word-level stress.

To confirm that focus indeed falls on [t͡so:r] in sentence 3, rather than on the stressed syllable of the target token, we fit a linear mixed-effects model to predict the mean f0 over the middle 60% of the vowel of all target tokens for a single speaker (AZ1). To rule out the possibility that sentence-level f0 differences between the two sentences could account for the observed effect, we included the mean f0 of [sʌŋgi] as a covariate in the model. The presumed focus status of the target token (focused vs. non-focused) was included as a fixed effect, and a random intercept was added for target token identity: token_MeanF0 ~ sʌŋgi_MeanF0 + Focus + (1 | target_token). The model shows that, even after controlling for f0 variability across sentences, target tokens in the non-focused condition (sentence 3) had significantly lower mean f0 than those in the focused condition (sentence 1). Specifically, target f0 was lower by approximately 16 Hz: Est. (non-focused) = - 15.81 Hz, 95% CI = [-18.97, -12.65], t = -9.81, p < .001*** (The full model is provided in Appendix A). In sum, the model corroborates the expected change in focus-related prominence.

Figure 1 shows sentence pitch tracks for the token [mʌzdɑ:r] embedded in sentences 1 and 3, where narrow focus falls on the stressed syllable of [mʌzdɑ:r] in sentence 1 but on [t͡so:r] 'four' in sentence 3. A visual comparison of the the two waveforms shows that the vowel of the stressed syllable of [mʌzdɑ:r] in sentence 1 is realized with a rising f0, increased intensity, and a longer duration than the same syllable in sentence 3. This rising f0 indicates that similar to other Indo-Aryan languages mentioned above, Mankiyali signifies new/contrastive information with a rise in f0 on the stressed syllable compared to neighboring non-focused syllables. It is also interesting to note that the relative f0 level of both sentences is similar (except for [mʌnd͡ʒu] in sentence 3), suggesting that, at least for AZ1, f0 downtrend is not a major factor.

Figure 1. f0 tracks overlaid on the waveforms of sentence 1 (left) and sentence 3 (right) for the token [mʌzˈda:r] spoken in the second session by AZ1. In sentence 1, [mʌzˈda:r] is new information and marked with focus-related prominence, and in sentence 3, [t͡so:r] 'four' is marked with focus, and [mʌzˈda:r] is not. Breaks in f0 tracks correspond to obstruents in the signal that did not produce a reliable f0. The black arrow points to the vowel of the word-final stressed syllable in [mʌzˈda:r].





As mentioned above, an added benefit of embedding the token in the middle of a simple declarative carrier sentence is that we avoid the interference of pitch excursions that often occur at phrasal boundaries (Gordon, 2014). However, a reviewer notes that sentence 1 has three fewer syllables than sentence 3. Given this, it is possible that the durational effects reported to arise from differences in focus-related prominence in section 4.2.4 may instead result from polysyllabic shortening within intonational phrases that results from a compression of syllables to fit the prosodic structure. To examine this potential confound, we conducted a Welch's Two Sample t-test comparing the vowel duration of the penultimate syllable of [sʌŋ.gi] in sentence one and sentence three for one speaker (MY0). If polysyllabic shortening was in effect, we should observe the penultimate syllable of [sʌŋ.gi] in sentence three to be significantly shorter than that in sentence one because of the extra syllables. Nevertheless, results of the t-test show no significant difference between the vowels in these two conditions: $t(178.84) = -.84$, $p = .40$, $n = 189$. As such, assuming data from MY0 is representative of the entire dataset, we conclude that polysyllabic shortening within the intonational phrase is not responible for any measurable changes in duration between focused and non-focused words in these sentences.

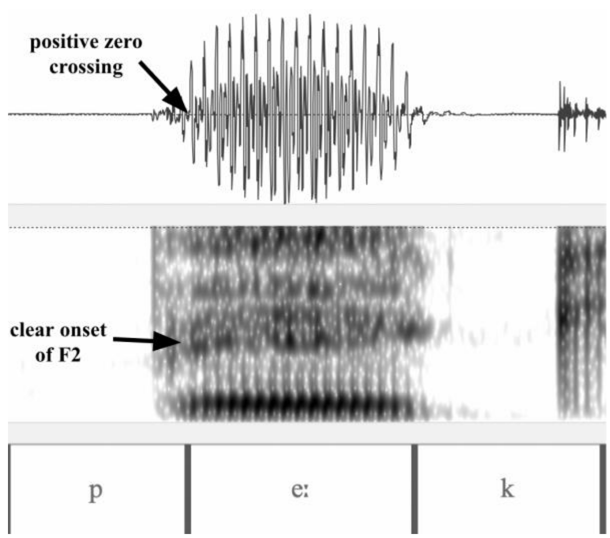
3.3 Elicitation procedure

Elicitations took place in the summer of 2022 in a quiet home in the Mankiyali-speaking villages of Danna and Dameka. A head-worn Audio-Technica BP894X Cardioid Condenser Microphone connected to a Zoom H5 Handy Recorder was used to record the productions. For all participants, the first author positioned the microphone approximately 2 cm from the right corner of the mouth to minimize measurement perturbations from plosive bursts and head movements. Additionally, input gains on the Zoom H5 were calibrated to -12dB for each speaker's normal voice amplitude before the recording began. Participants were recorded individually and given oral instructions for the task in Mankiyali by the second author, a native Mankiyali speaker. Participants were then shown a document on a laptop containing the mini-monologues discussed in the previous section and were subsequently asked to take five minutes to read through the sentences to familiarize themselves with the stimuli before the session began. Any questions about the content were answered in Mankiyali. Once the session began, participants were told to read the mini-monologues out loud at a normal pace. The order in which the mini-monologues appeared was randomized but fixed across speakers. Each speaker was recorded reading the entire wordlist twice with 2-4 days in between sessions. The entire set of audio files collected for this experiment is archived with the Computational Resource for South Asian Languages (CoRSAL), an online repository for under-resourced languages of South Asia (CoRSAL, 2023).

3.4 Acoustic analysis

The segmentation process proceeded as follows. The vowel of each target syllable was labeled in Praat TextGrids (Boersma & Weenink, 2015) by the first author. The acoustic intervals associated with vowels were determined according to the first zero crossing of a clear change in periodicity and amplitude at the onset and offset of the vowel on a waveform. For tokens lacking an abrupt change in amplitude in the waveform, a change in the finer structure of the wave cycles was used as a boundary cue instead. When neither a change in amplitude nor distinctions in wave cycles were clear, the spectrogram was consulted to determine the location of the onset and offset of a clear F2. An example of the segmentation criteria is shown in Figure 2 for the vowel /eː/.

Figure 2. Representative image of vowel segmentation in Praat for /peːkiːz/ extracted from Sentence 1 of the mini-monologue from the second session of speaker MS1.



The duration of target vowels was calculated using the vocalic intervals with the boundaries stipulated above. The intensity for each vowel was calculated in two ways. First, the average intensity over the middle 60% of each vowel was measured to avoid transition effects for surrounding consonants at vowel edges. This was computed using the built in Praat intensity calculator, which computes the total intensity accumulated over the specified timeframe, from t_1 to t_2, and divides by that timeframe to determine the mean: $\frac{1}{(t_2-t_1)}\int_{t_1}^{t_2} x(t)dt$. In addition, the peak intensity over this interval was extracted. Results from average and peak intensity were similar, so we only report the statistical results of peak intensity.

A static measure of f0 was taken in a similar fashion, calculating the mean f0 across the middle 60% of the vowel to avoid measurement effects from segments flanking the vowel. Moreover, to include a dynamic analysis of f0, a series of f0 values were extracted at ten equally spaced time points across the length of each vowel, which permits an analysis of how f0 might change over the course of the vowel.

To examine whether stress vowels are produced with more vocal effort, the spectral tilt of each vowel was also measured, following the method proposed by van Heuven & Turk (2021, p.164) of analyzing the slope of a regression line fit through the long-term average spectrum (LTAS) of a vowel. After the vowel interval was annotated, The LTAS was computed with a bandwidth of 100 Hz; a regression line was then fit to the LTAS with a frequency range of 50 to 4,000 Hz. The spectral tilt corresponds to the slope of the regression line and is reported in dB/Hz.

Vowel space expansion refers to the potential for the quality of a vowel to expand toward the periphery or reduce toward a more neutral schwa-like position, depending on whether a vowel is stressed or unstressed. The relative peripherality of each vowel was calculated as the Euclidean distance of the target vowel from the hypothesized middle of each speaker's vowel space. To mitigate inter-speaker variation, F1 and F2 formant values were first normalized by taking the ratio of the first two formants compared to F3 (F1/F3 and F2/F3) before calculating the Euclidean distance of a vowel from each speaker's central point ("neutral schwa") in this F3-normalized plane. As discussed in Monahan & Idsardi (2010), using these formant ratio algorithms instead of raw formant values has been shown to remove variation across speakers. Because the vowel space changes from speaker to speaker, the neutral schwa was calculated individually for all 30 participants. This was done in the following way. First, the mean F1/F3 and F2/F3 values of /i/,

/u/, and /a:/ were calculated for each participant using the tokens they produced during the experiment. These three vowels were selected because they provide an estimate of the outer bounds of the vowel space in three separate directions away from the center of the vowel space. After the formant ratio values were calculated for each of these three vowels, the F1/F3 and F2/F3 center of gravity between them was pinpointed: F1/F3ə = (μF1/F3i + μF1/F3u + μF1/F3ɑː)/3 and F2/F3ə = (μF2/F3i + μF2/F3u + μF2/F3ɑː)/3. Once the formant ratio values of the neutral schwa were determined for each speaker, the Euclidean distance of every target vowel token from each speaker's neutral schwa was then computed in the F3-normalized vowel space on a speaker-by-speaker basis. Comparing the Euclidean distance of stressed and unstressed vowels allows us to explore potential differences in the vowel space between the two stress conditions.

3.5 Statistical analysis

Statistical analysis was carried out using linear mixed-effects models in R (R Core Team, 2016) with the *lme4* package (Bates et al., 2023) and *lmerTest* (Kuznetsova et al., 2020). For each of the five acoustic properties examined in this study, a series of models was run. Full model specifications are outlined in detail in the corresponding results sections. In general, each analysis begins with a complex lmer model including main effects and interactions, followed by a series of pairwise comparisons extracted using the *emmeans()* function in R (Lenth, 2024). Bonferroni corrections to p-values were applied based on the number of comparisons made using the *adjust = 'bonferroni'* option in *emmeans*. Given the relatively small effect sizes found in this study, the Bonferroni correction was chosen to provide a conservative adjustment and reduce the risk of Type I error in multiple comparisons.

Additionally, it is well known that speech rate, f0, and intensity baselines vary widely across speakers. To avoid unintended effects of speaker variation in the statistical models, duration, intensity, spectral tilt, and f0 values were converted to speaker-specific z-scores. To remove outliers, values greater than 2.5 units from the mean for both z-scored variables and Euclidean Distance were removed before the complex lmer models were run. The number of tokens trimmed in this process are provided in table 3.

Table 3: number of outliers removed for each z-scored model.

Measures	unstressed vs. stressed	non-focused vs. focused
Duration	0 (0.00%)	7 (0.27%)

Intensity	58 (2.13%)	32 (1.26%)
Spectral Tilt	11(0.4%)	14 (0.55%)
f0	7 (0.26%)	73 (2.87%)
Euclidean Distance	66 (2.43%)	55 (2.16%)

Because the magnitude of the effect of stress and focus-related prominence on the different acoustic properties are crucial for understanding the results and ensuing discussion in this paper, we chose to report pairwise comparisons in raw measurements rather than z-scored values. Using raw values makes interpreting the effect sizes more straightforward.

Finally, while vowel quality is known to impact duration, our experimental tokens lack a balance in vowel quality across syllable type conditions, which arose from difficulty in identifying near-minimal pairs with varying vowel qualities. If included as a random effect, the disparity in vowel quality across syllable type conditions might cause vowel quality to erroneously soak up some of the variation accounted for by the effect of syllable type. Thus, it was excluded as a predictor in the models. Random effects for both SPEAKER and TARGET SYLLABLE were included where appropriate. Including SPEAKER as a random effect helps the model to account for inherent differences in f0, speaking rate, etc., across participants, and including random slopes and intercepts for TARGET SYLLABLE (e.g., /kʰʌ/, /mus/, /kɑː/, etc.) ensures inherent variability across near-minimal word pairs does not deleteriously alter the estimated coefficients.

All visual plots are presented using normalized values throughout the paper to avoid different speaker baselines obscuring the visuals.

4 Results

4.1 Word-level stress correlates

This section presents results for the impact of word-level stress on the acoustic properties of the signal. Crucially, as discussed in section 3, all tokens included in the analysis of word-level stress correlates are presumably devoid of phrase-level prosodic effects since the tokens are neither focused nor utterance-final. As such, this section compares word-level primary-stressed syllables to word-level secondary-stressed/unstressed syllables, both embedded in the third sentence of the mini-monologues described above.

4.1.1 Duration

With 30 speakers producing each of the 50 tokens twice, the total possible tokens from the dataset was N = 3,000 (1,500 stressed and unstressed). 173 of these 3,000 tokens were discarded due to poor recording quality, reading errors, or unnatural pauses. In addition, tokens from the near-minimal pair, [belti]~[beltiːz], were removed entirely because the [e] in the initial syllable of both words was consistently produced as a long vowel rather than a short vowel by all speakers, thereby impacting the predicted stress location. After removing these items, the resulting token set included an actual total of N = 2,719 words used in the analysis. Furthermore, possibly because Mankiyali has no official writing system, participants sometimes misread a token from the "stressed" category as a token from the "unstressed" category (e.g. [ˈmus.ki] was sometimes read with a long vowel in word-final position: [mus.ˈkiː]). Importantly, long vowels are approximately double the length of short vowels in Mankiyali, so spotting these errors is straightforward. Therefore, rather than discarding these tokens, they were recategorized based on their actual productions under the appropriate stress condition; 95 tokens were recategorized in this way. Altogether, a total of 1,251 tokens with primary stress on the target syllable and 1,468 tokens lacking primary stress on the target syllable were included. Table 4 details the exact token quantities broken up by stress condition for each of the five syllable types used in the duration analysis.

Table 4: Number of tokens for each syllable type used in the analysis of word-level duration.[3]

	Primary Stress	**Secondary/No Stress**	**Total per Syllable Type**
CV	291	306	597
CVC	181	242	423
CVCC	226	322	548
CVː	269	292	561
CVːC	284	306	590
Total	1,251	1,468	2,719

Table 5 summarizes the outcome of the effects-coded *lmer* model testing the effects of the different levels of STRESS, SYLLABLE TYPE, and their interaction on duration. As is evident, the effect of STRESS on duration is significant, and, out of the interaction of stress with each of the five syllable types, only {CVːC} reached significance. This means that, in general, syllables exhibit

[3] An anonymous reviewer points out that the token numbers are not balanced across conditions. Unlike ANOVAs, the linear mixed effects models used in this paper are specifically equipped to handle this degree of inbalance between groups considering the relatively large dataset being used (Baayen et al, 2008; Speelman et al, 2018). However, to address this concern, we added statistical tests and visualizations for durational differences between unstressed and stressed syllables for individual near-minimal pairs in appendix B.

significant changes in duration between the stressed and unstressed conditions that is similar across syllables types. That is, the use of duration as an acoustic correlate of stress in Mankiyali is relatively constant across the five syllable types. The significant interaction between stress and the {CVːC} syllable type on duration is likely an artifact of the experimental design. That is, because stressed and unstressed {CVːC} syllables were compared, by necessity, in word-final position, word-final lengthening effects presumably increased the mean duration of the vowel itself in both stressed and unstressed tokens. Additionally, comparisons in the final position likely result in a greater absolute effect size because a greater proportional difference in absolute duration must be achieved between two vowels in the word-final position for the difference to be perceptible (Hogoboom, 2013).

Table 5: z-scored *lmer* model summary showing the effects of STRESS and SYLLABLE TYPE on vowel duration: Duration~STRESS*SYLLABLE TYPE + (1 + STRESS|speaker) + (1|target.syllable).

Factors	Estimate (z-scored over ms)	95% CI	t-value	p-value
Intercept (grand μ)	-.25	[-0.36, -0.15]	-4.59	< .001 ***
stressed	**.14**	**[0.13, 0.15]**	**22.03**	**< .001 *****
CV	**-.67**	**[-0.87, -0.47]**	**-6.46**	**< .001 *****
CVC	**-.64**	**[-0.86, -0.42]**	**-5.68**	**< .001 *****
CVCC	**-.56**	**[-0.77, -0.36]**	**-5.43**	**< .001 *****
CVː	**.72**	**[0.52, 0.93]**	**6.99**	**< .001 *****
CVːC	**1.15**	**[0.95, 1.35]**	**11.09**	**< .001 *****
stressed:CV	-.01	[-0.03, 0.01]	-.70	.48
stressed:CVC	-.01	[-0.04, 0.01]	-.89	.38
stressed:CVCC	-.01	[-0.04, 0.01]	-1.01	.31
stressed:CVː	-.01	[-0.03, 0.02]	-.56	.58
stressed:CVːC	**0.04**	**[0.02, 0.06]**	**3.32**	**< .001 *****

Considering these word-final lengthening effects, it is safe to conclude that our study found no evidence that syllable type alters the use of duration as an acoustic correlate of stress. Furthermore, while we expected to see an interaction effect of syllable type and stress that distinguished long vowels from short vowels for duration, only {CVːC}, and not {CVː}, was found to exhibit a significant difference in durational effect size. This is likely because syllables with long vowels always receive some degree of stress compared to short vowel syllables, which are either stressed or unstressed in the tokens used for this experiment.

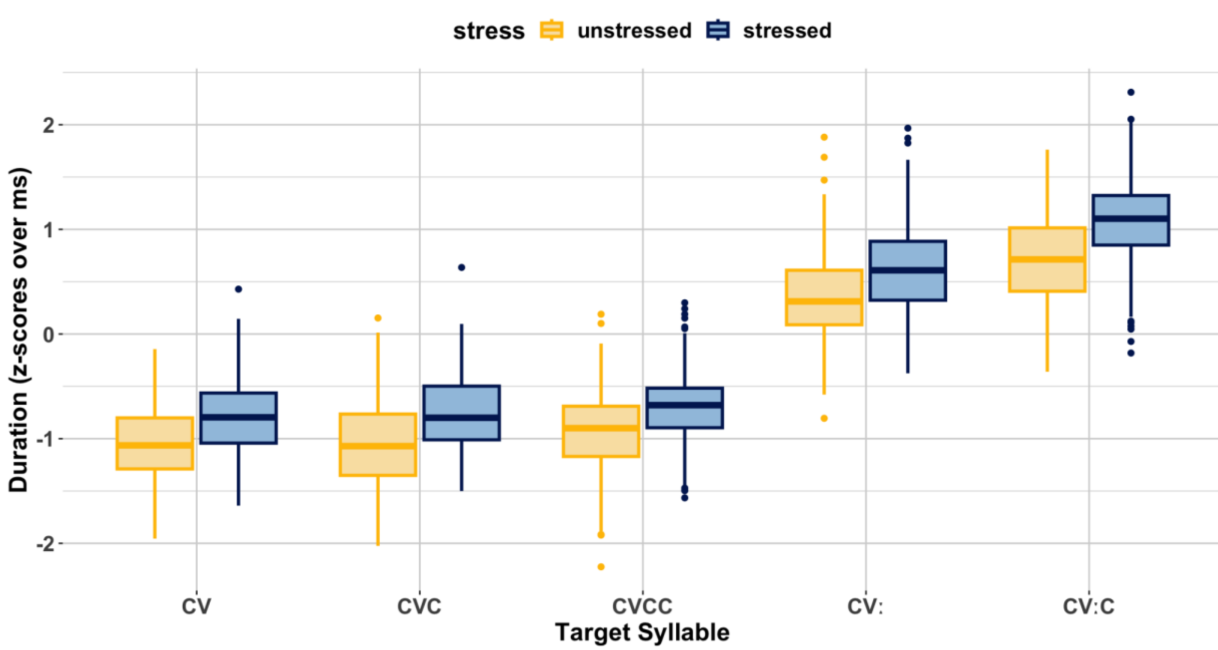
The interesting portion of the results concerns the effect of stress on vowel duration. As the pairwise comparisons in Table 6 show, the presence of stress on the five different syllable types caused a mean vowel length increase between 10-15 ms (10-19%), with t- and p-values indicating a significant difference between the two stress conditions for all syllable types.

Table 6: Pairwise comparisons showing the effect of STRESS on vowel duration for each of the five syllable types (CV, CVC, CVCC, CV:, CV:C) in the experiment: emmeans(lmer, ~ STRESS | SYLLABLE TYPE).

syllable type	unstressed est. (ms)	95% CI	stress effect (ms)	stress effect 95% CI	t-value	p-value
CV	**56**	**[46, 67]**	**11 (19%)**	**[9, 13]**	**9.11**	**< .001 *****
CVC	**57**	**[46, 69]**	**11 (18%)**	**[8, 13]**	**7.52**	**< .001 *****
CVCC	**60**	**[50, 71]**	**10 (17%)**	**[8, 13]**	**8.18**	**< .001 *****
CV:	**112**	**[102, 122]**	**11 (10%)**	**[8, 13]**	**8.67**	**< .001 *****
CV:C	**128**	**[117, 138]**	**15 (11%)**	**[12, 17]**	**12.13**	**< .001 *****

Figure 3 displays boxplots of tokens categorized by stress and syllable type. As shown in the plot and the above model summaries, stressed syllables are consistently longer than their unstressed counterparts, and the effect size remains relatively stable across all syllable types.

Figure 3: *Boxplots grouped by syllable type depicting z-scored vowel duration differences between word-level stressed and unstressed syllables.*



4.1.2 Intensity

The number of tokens used in the *lmer* model for intensity equaled the number used in the duration model. As shown in Table 7, STRESS had a small but statistically significant effect on peak intensity. Interactions between STRESS and SYLLABLE TYPE were generally not significant, indicating that the effect of stress on intensity was relatively stable across different syllables. An exception was stressed CVC, which showed a significant but modest increase in peak intensity.

Table 7: z-scored *lmer* model showing correlation of word-level stress and syllable type with intensity. PeakIntensity~STRESS*SYLLABLE TYPE + (1 + STRESS|speaker) + (1|target.syllable).

Factors	Estimate (z-scored over dB)	95% CI	t-value	p-value
Intercept (grand μ)	-.42	[-0.62, -0.21]	-3.94	< .001 ***
stressed	**.04**	**[0.02, 0.07]**	**3.26**	**= .001 ****
CV	-.06	[-0.46, 0.34]	-.30	.77
CVC	-.28	[-0.71, 0.15]	-1.26	.22
CVCC	-.22	[-0.61, 0.18]	-1.08	.29
CV:	.25	[-0.15, 0.64]	1.22	.24
CV:C	.34	[-0.1, 0.78]	1.5	.15
stressed:CV	-.01	[-0.05, 0.04]	-.31	.76
stressed:CVC	**.06**	**[0.01, 0.12]**	**2.19**	**.03 ***
stressed:CVCC	0	[-0.05, 0.05]	-.13	.90
stressed:CV:	-.04	[-0.09, 0.01]	-1.55	.12
stressed:CV:C	-.03	[-0.08, 0.02]	-1.07	.28

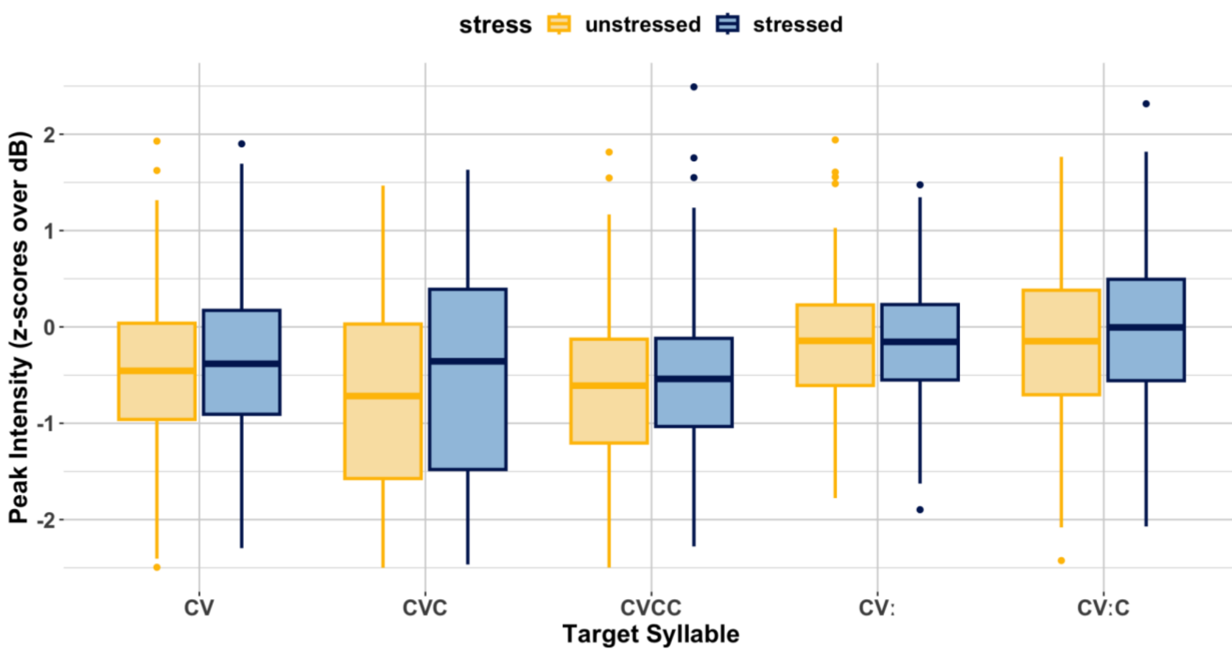
Pairwise comparisons by syllable type of estimated marginal means (EMMs) of peak intensity, provided in Table 8, show that no difference in intensity arose as a factor of stress for any of the syllable types except CVC. Though the interaction of STRESS and SYLLABLE TYPE for CVC syllables does exhibit a significant effect, the estimated effect size is negligible, at approximately 1.4 decibels, is barely perceptable using *wideband* noise in highly controlled environments (Stevens, 2000; Moore, 2013), let alone for speech with ambient noise. Given this and the lack of significance for the other four syllable types, we conclude that intensity is not a reliable correlate of word-level stress in Mankiyali for any syllable type.

Table 8: Pairwise comparisons showing the correlation of STRESS and peak vowel intensity for each of the five syllable types (CV, CVC, CVCC, CV:, CV:C) in the experiment: emmeans(lmer, ~ STRESS | SYLLABLE TYPE).

syllable type	unstressed est. (dB)	95% CI	stress effect (dB)	stress effect 95% CI	t-value	p-value
CV	68.1	[65.5, 70.8]	0.4 (0.6%)	[-0.1, 1]	1.7	.09
CVC	**66.3**	**[63.4, 69.3]**	**1.4 (2.1%)**	**[0.8, 2]**	**4.48**	**< .001 *****
CVCC	67.6	[65, 70.2]	0.3 (0.5%)	[-0.2, 0.9]	1.12	.26
CV:	69.9	[67.3, 72.5]	0.1 (0.1%)	[-0.4, 0.6]	.3	.77
CV:C	70.3	[67.7, 72.9]	0.2 (0.2%)	[-0.4, 0.7]	.61	.54

The boxplots in Figure 4 provide a visual confirmation that stressed and unstressed syllables tended to have no measurable effect on intensity values across syllable types in this study.

Figure 4: *Boxplots grouped by syllable type depicting vowel intensity differences between word-level stressed and unstressed syllables.*



4.1.4 Spectral Tilt

In many languages, stressed syllables have been found to exhibit a notable increase in intensity in the higher frequency ranges compared to lower frequencies, and this change in spectral tilt is more readily perceived than aggregate spectral changes in overall intensity (Sluijter & van Heuven, 1996). This asymmetric increase in intensity across the spectrum due to stress causes the spectral slope of stressed syllables to be less steep than that of unstressed syllables. Consequently, fitting a regression line to the intensity values of a vowel's spectrum and measuring the slope of that line (i.e., the spectral tilt) is an important potential correlate to word-level stress.

The full token set (N = 2,719) was used in the analysis of spectral tilt. The lmer output summary in Table 9 shows the effects of STRESS and SYLLABLE TYPE on the spectral tilt measurements. As is evident, the presence of stress was found to significantly impact spectral tilt by flattening the slope, indicating that higher frequencies increase in intensity compared to lower frequencies in stressed syllables. Additionally, the effect of stress on spectral tilt was significantly stronger in CVCC syllables compared to other syllable types.

Table 9: z-scored *lmer* model summary showing the effects of STRESS and SYLLABLE TYPE on spectral tilt: SpectralTilt~STRESS*SYLLABLE TYPE + (1 + STRESS|speaker) + (1|target.syllable).

Factors	Estimate (z-scored over dB/Hz)	95% CI	t-value	p-value
Intercept (grand μ)	-.16	[-0.39, 0.07]	-1.37	.19
stressed	**.04**	**[0.02, 0.06]**	**3.46**	**< .001 *****
CV	-.46	[-0.9, -0.02]	-2.04	.06
CVC	**-.54**	**[-1.02, -0.06]**	**-2.22**	**.04 ***
CVCC	**.49**	**[0.05, 0.93]**	**2.18**	**.04 ***
CVː	**.98**	**[0.54, 1.42]**	**4.38**	**< .001 *****
CVːC	**-.48**	**[-0.93, -0.03]**	**-2.1**	**.05 ***
stressed:CV	-.01	[-0.05, 0.03]	-.46	.64
stressed:CVC	-.01	[-0.06, 0.04]	-.43	.67
stressed:CVCC	**.05**	**[0, 0.09]**	**2.16**	**.03 ***
stressed:CVː	.01	[-0.04, 0.05]	.25	.80
stressed:CVːC	-.03	[-0.07, 0.01]	-1.53	.13

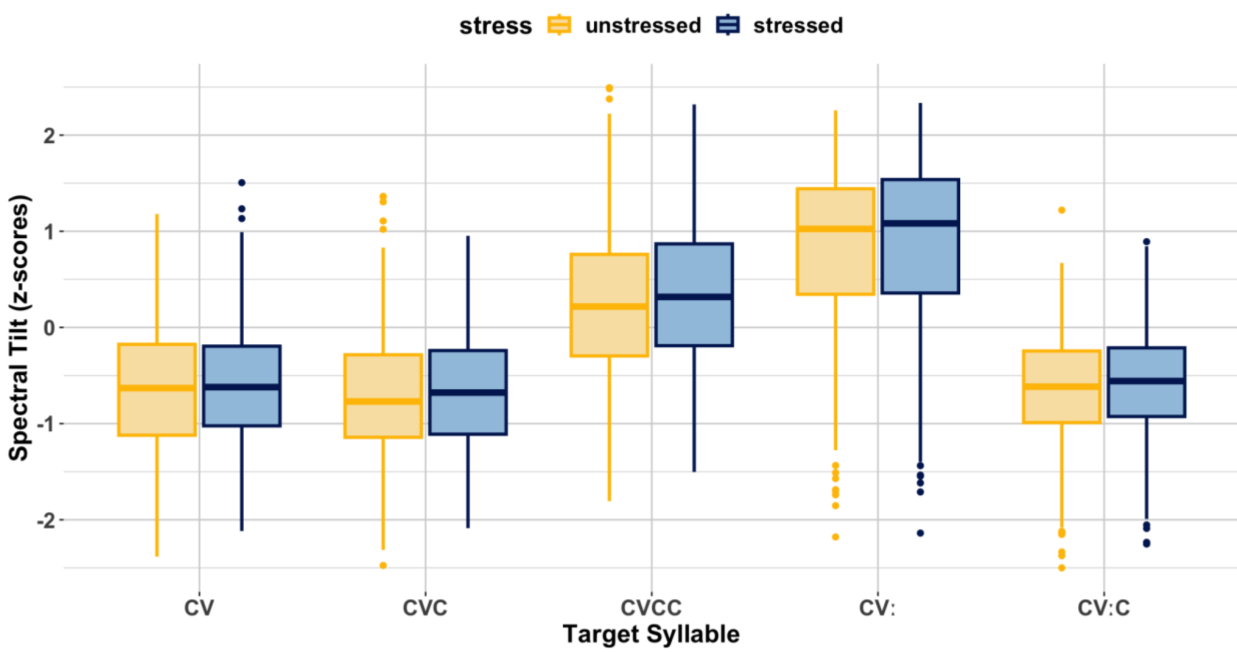
Despite the main effect of stress found in the complex model, pairwise comparisons in Table 10 and the corresponding boxplots in Figure 5 highlight that a significant change in spectral tilt correlating with stress was only observed for CVCC syllables.

Table 10: Pairwise comparisons showing the correlation of STRESS and spectral tilt for each of the five syllable types (CV, CVC, CVCC, CVː, CVːC) in the experiment: emmeans(lmer, ~ STRESS | SYLLABLE TYPE).

syllable type	unstressed est. (db/Hz)	95% CI	stress effect (db/Hz)	stress effect 95% CI	t-value	p-value
CV	-.009	[-.01, -.007]	.0001 (1.7%)	[-.0001, .0004]	1.29	.2
CVC	-.009	[-.01, -.007]	.0001 (1.4%)	[-.0001, .0004]	.91	.37

CVCC	**-.007**	**[-.008, -.005]**	**.0004 (6.3%)**	**[.0002, .0006]**	**3.33**	**< .001 *****
CV:	-.005	[-.007, -.004]	.0002 (4.4%)	[0, .0005]	1.91	.06
CV:C	-.009	[-.01, -.007]	0 (0%)	[-.0002, .0002]	.02	.98

Figure 5: *Boxplots grouped by syllable type depicting spectral tilt differences between word-level stressed and unstressed syllables.*

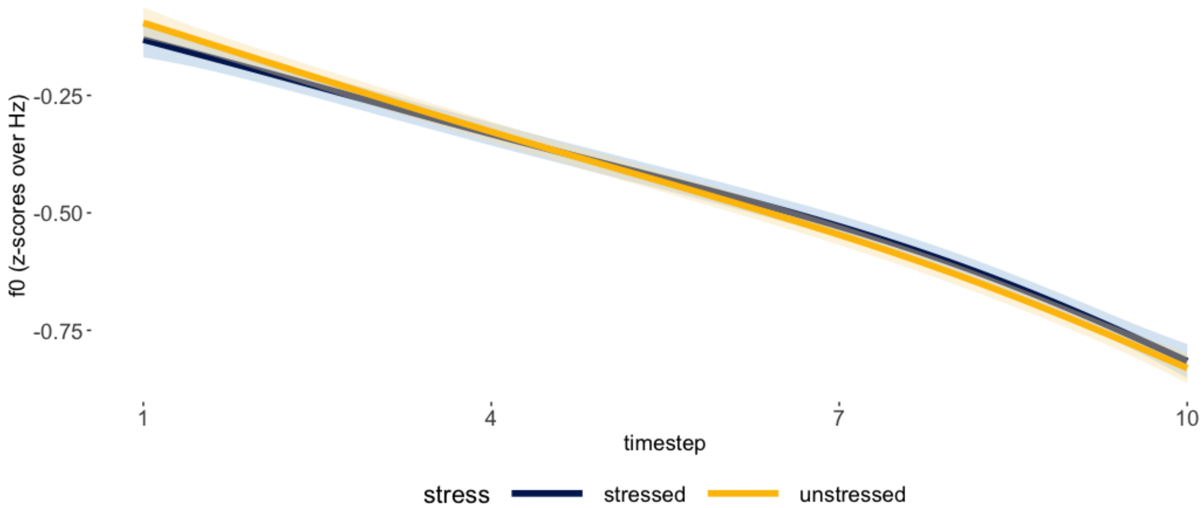


To sum up, while the full *lmer* model revealed a significant main effect of stress on spectral tilt, follow-up pairwise comparisons using estimated marginal means indicated that this effect was limited to CVCC syllables. This suggests that although spectral tilt may weakly correlate with stress in Mankiyali, the effect is restricted and inconsistent across syllable types, limiting its reliability as a robust correlate to word-level stress.

4.1.4 f0

Figure 6 displays two GAM (Generalized Additive Model) curves plotting f0 values separated by stress condition. After the duration of each vowel token in the experiment was calculated, we extracted ten f0 measurements equally spaced across each vowel's duration. The curves are fit to these ten f0 measurements, with each time point averaged across all speakers and vowel tokens. As is evident from the plot, the two curves are almost identical in the change in f0 over time and their f0 values at each timestep.

Figure 6: GAM curves separated into word-level stressed (dark blue) and unstressed (yellow) conditions showing mean vowel f0 across 10 normalized timesteps for all non-focused conditions.



The lmer model output in Table 11 summarizes the interaction between stress and timestep and shows no significant interaction between these two factors. Thus, it is left out of the following models, and a static measure of mean f0 over the middle 60% of the vowel is used instead to simplify the statistical interpretations.

Table 11: z-scored *lmer* model summary showing the effects of STRESS and STEP on f0: f0~STRESS*STEP + (1 + STRESS|speaker) + (1|target.syllable).

Factors	Estimate (z-scored over Hz)	95% CI	t-value	p-value
Intercept (grand μ)	-.42	[-0.56, -0.28]	-5.83	< .001 ***
stressed	.01	[-0.02, 0.03]	.57	.57
stressed:step1	-.01	[-0.04, 0.02]	-.69	.49
stressed:step2	0	[-0.03, 0.02]	-.23	.82
stressed:step3	-.01	[-0.03, 0.02]	-.74	.46
stressed:step4	-.01	[-0.04, 0.01]	-.83	.41
stressed:step5	0	[-0.03, 0.02]	-.04	.97
stressed:step6	0	[-0.02, 0.03]	.37	.71
stressed:step7	.01	[-0.02, 0.03]	.40	.69
stressed:step8	.01	[-0.01, 0.04]	.83	.41
stressed:step9	.01	[-0.02, 0.03]	.76	.45
stressed:step10	.01	[-0.02, 0.03]	.76	.45

The total number of tokens used for the *lmer* model analyzing mean f0 was 55 less (N = 2,662) than for models analyzing other acoustic properties because these tokens were either too short or too creaky to find a reliable mean f0 across the middle 60% of the vowel. Notably, almost 70% of these tokens (37) that lacked a measurable f0 were in unstressed conditions. Table 12 summarizes the effects of STRESS and SYLLABLE TYPE on mean f0. The interaction of stress and syllable type was significant for CVC, CVCC, and CV:C, with the effect of stress on f0 increasing for CVC and slightly decreasing for CVCC and CV:C.

Table 12: z-scored *lmer* model summary showing the effects of STRESS and SYLLABLE TYPE on Meanf0. Meanf0~STRESS*SYLLABLE TYPE + (1 + STRESS|speaker) + (1|target.syllable).

Factors	Estimate (z-scored over Hz)	95% CI	t-value	p-value
Intercept (grand μ)	-.45	[-0.58, -0.33]	-7.04	< .001 ***
stressed	.02	[-0.01, 0.04]	1.37	.17
CV	.08	[-0.15, 0.3]	.69	.5
CVC	**.33**	**[0.09, 0.58]**	**2.66**	**.02 ***
CVCC	.03	[-0.19, 0.26]	.29	.78
CV:	-.10	[-0.33, 0.13]	-.87	.39
CV:C	**-.36**	**[-0.59, -0.12]**	**-3**	**.007 ****
stressed:CV	0	[-0.05, 0.05]	.13	.90
stressed:CVC	**.14**	**[0.09, 0.2]**	**5**	**< .001 *****
stressed:CVCC	**-.09**	**[-0.14, -0.04]**	**-3.61**	**< .001 *****
stressed:CV:	0	[-0.05, 0.05]	.05	.96
stressed:CV:C	**-.06**	**[-0.11, -0.01]**	**-2.25**	**.03 ***

As indicated by the pairwise comparisons in Table 13, however, only CVC showed a significant stress effect. Thus, although f0 may weakly correlate with stress in CVC syllables specifically, it does not emerge as a consistent correlate of word-level stress in Mankiyali.

Table 13: Pairwise comparisons showing correlation of STRESS and mean f0 for each of the five syllable types (CV, CVC, CVCC, CV:, CV:C) in the experiment.

syllable type	unstressed est. (Hz)	95% CI	stress effect (Hz)	stress effect 95% CI	t-value	p-value
CV	124.2	[116.4, 132.1]	0.4 (0%)	[-1, 1.8]	.58	.56
CVC	**125.6**	**[117.7, 133.6]**	**3.3 (3%)**	**[1.7, 5]**	**3.91**	**< .001 *****
CVCC	125.1	[117.3, 132.9]	-1.3 (-1%)	[-2.8, 0.1]	-1.77	.08
CV:	122.5	[114.6, 130.3]	0.5 (0%)	[-1, 1.9]	.62	.53
CV:C	120.4	[112.6, 128.2]	-0.9 (-1%)	[-2.3, 0.4]	-1.35	.18

4.1.5 Vowel Space Expansion

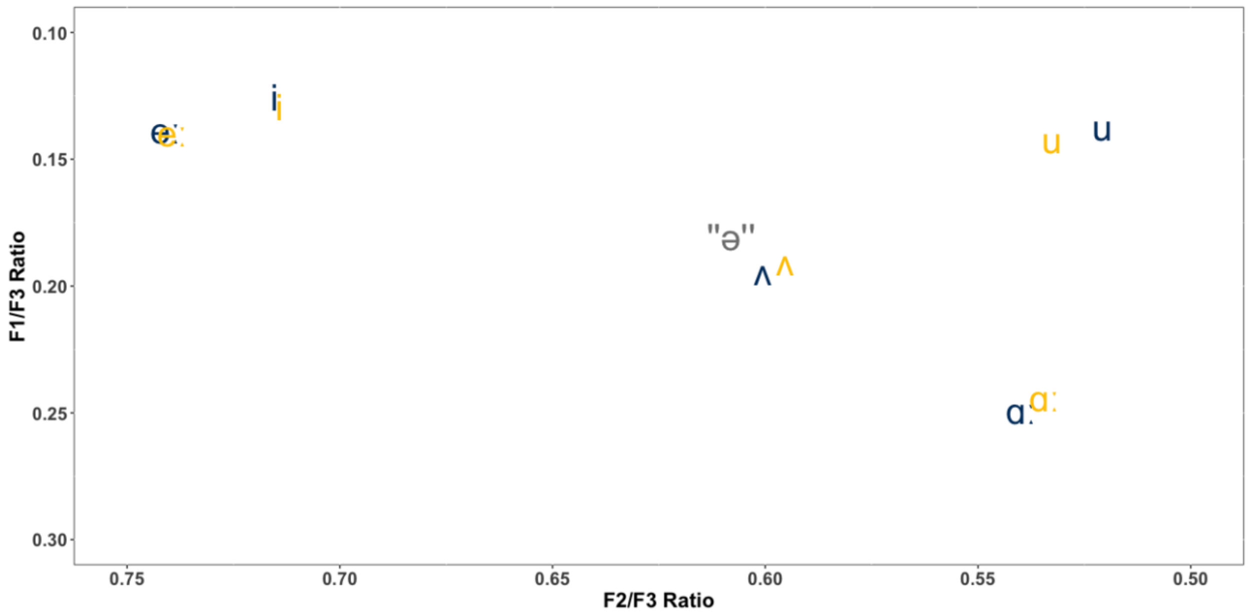
To quantify the potential effect of stress on vowel quality, we calculated the Euclidean distance of stressed vowel tokens from the hypothetical center of each speaker's vowel space and compared it to the Euclidean distance of unstressed vowel tokens from this center point. As mentioned above, the Euclidean distance measurements were taken in relation to the F1/F3-F2/F3 plane.

A summary of the pairwise comparisons of estimated marginal means between unstressed and stressed vowels is given in Table 14. Syllable type had no effect on the results, so the findings are reported for each vowel collapsed across syllable types. Stress was not found to significantly impact vowel quality for any of the five vowels. This is visually corroborated by examining the vowel space in Figure 7, which shows the word-level stressed vowels (dark blue) falling almost on top of the word-level unstressed vowels (yellow) in terms of their distance from the neutral schwa in the F1/F3-F2/F3 space. Altogether, the results examining the correlation between word-level stress and vowel quality indicate no relationship between the two variables.

Table 14: Pairwise comparisons showing the correlation between stress and Euclidean Distance from a neutral schwa for each vowel quality in the study. Euclidean Distance is calculated in an F1/F3-F2/F3 space, so estimates are not in Hertz but with respect to the first two formants as ratios of F3. EucDist~STRESS + (1|speaker) + (1|target.syllable).

vowel	unstressed est. (EucDist)	95% CI	stress effect (EucDist)	stress effect 95% CI	t-value	p-value
ɑː	.11	[.09, .12]	.003 (3%)	[-.003, .01]	.95	.34
eː	.14	[.12, .16]	.003 (2%)	[-.005, .011]	.78	.43
u	.11	[.09, .13]	.005 (4%)	[-.005, .014]	.93	.35
ʌ	.09	[.08, .1]	0 (0%)	[-.006, .005]	-.18	.86
i	.13	[.1, .15]	-.001 (-1%)	[-.013, .011]	-.18	.85

Figure 7: F1/F3-F2/F3 vowel space showing the mean F1/F3 and F2/F3 ratio values of all 30 speakers for stressed (dark blue) and unstressed (yellow) word-level tokens with the mean hypothetical neutral schwa (grey) across all speakers.



4.1.6 Summary

This section explored the effect of word-level stress on five acoustic properties commonly associated with stress cross-linguistically: duration, intensity, spectral tilt, f0, and vowel quality. No evidence was found indicating that intensity, spectral tilt, f0, or vowel quality covary with stress consistently across syllable types. In contrast, duration emerged as the only robust and consistent correlate: primary stressed vowels were approximately 10-11 ms longer (10-19%) than their secondary/unstressed counterparts for {CV}, {CVC}, {CVCC}, and {CV:} syllables and approximately 15 ms longer (11%) for {CV:C} syllables. Although the magnitude of this difference increased for {CV:C} syllables, the overall durational effect remained relatively stable when considering word-final lengthening. In short, our findings suggest that duration is the most reliable acoustic correlate of word-level stress in Mankiyali across syllable types.

Weak but statistically significant effects of stress on f0 and spectral tilt were observed for isolated syllable types. Specifically, pairwise comparisons showed a small but significant increase in f0 (3.4 Hz) for stressed CVC syllables, and a significantly flatter spectral tilt for stressed CVCC syllables. Thus, while it is possible that different syllable types in Mankiyali utilize distinct acoustic correlates to signal stress, the evidence for this is weak.

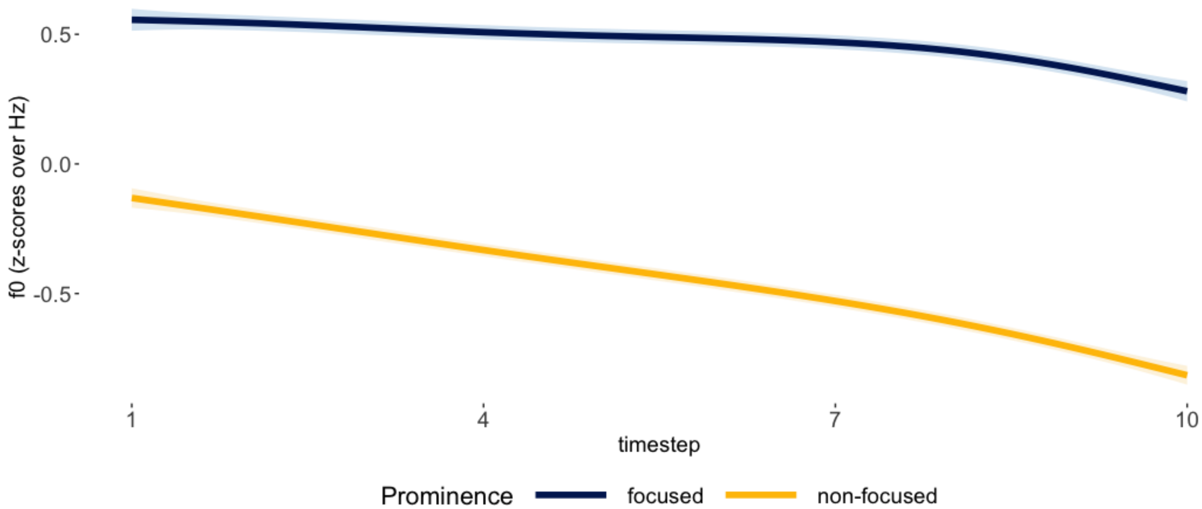
4.2 Focus-related prominence correlates

This section examines the correlates to focus-related prominence by comparing focused tokens (tokens with word-level stress produced in sentence 1 of the mini-monologue) to tokens without focus (tokens with word-level stress produced in sentence 3 of the mini-monologue). Importantly, both groups of tokens are marked with word-level stress, so other than differences in sentence syllable count, the primary difference between the two tokens is the presence/absence of narrow focus. This means that we can attribute any acoustic differences between the two groups to the effects of focus-related prominence. None of the interaction effects for FOCUS*SYLLABLE TYPE were significant for any of the acoustic properties measured except for duration, so results in this section are reported with the SYLLABLE TYPE predictor dropped from all models except for the lmer model with duration as the independent variable.

4.2.1 f0

Similar to the GAM curves in Figure 6 in section 4.1.4 for word-level stressed vs. unstressed tokens, Figure 8 below shows GAM curves fit to 10 normalized timesteps at which f0 was measured for every stressed vowel in both the focused and non-focused conditions. Whereas the curves comparing word-level stressed and unstressed tokens in section 4.1.4 were essentially on top of each other (indicating that f0 is not a correlate of word-level stress in Mankiyali), the curves here demonstrate that f0 is substantially higher in the focused condition compared to the unfocused condition across the length of the vowel.

Figure 8: GAM curves separated into focused (dark blue) and non-focused (yellow) conditions showing mean vowel f0 across 10 normalized timesteps for all tokens.



As shown by the summary of the effects for FOCUS and STEP in Table 15, the first four timesteps exhibit significant *negative* effect sizes, but the final four timesteps exhibit significant *positive* effects. These significant interactions indicate that the effect of focus-related prominence is more pronounced later in the vowel.

Table 15: z-scored *lmer* model summary showing the effects of FOCUS and STEP on f0: f0~FOCUS*STEP + (1 + FOCUS|speaker) + (1|target.syllable).

Factors	Estimate (z-scored over Hz)	95% CI	t-value	p-value
Intercept (grand μ)	.11	[-0.08, 0.3]	1.11	.28
Focused	**.51**	**[0.44, 0.57]**	**15.36**	**< .001 *******
Focused:step1	**-.05**	**[-0.09, -0.02]**	**-3.33**	**< .001 *******
Focused:step2	**-.07**	**[-0.1, -0.04]**	**-4.29**	**< .001 *******
Focused:step3	**-.06**	**[-0.09, -0.03]**	**-3.65**	**< .001 *******
Focused:step4	**-.04**	**[-0.07, -0.01]**	**-2.34**	**.02 ***
Focused:step5	-.02	[-0.05, 0.01]	-1.24	.22
Focused:step6	0	[-0.03, 0.03]	.01	.99
Focused:step7	**.03**	**[0, 0.06]**	**2.01**	**.04 ***
Focused:step8	**.06**	**[0.03, 0.09]**	**3.71**	**< .001 *******
Focused:step9	**.07**	**[0.04, 0.1]**	**4.57**	**< .001 *******
Focused:step10	**.07**	**[0.04, 0.11]**	**4.76**	**< .001 *******

Because the effect of FOCUS on f0 changes across the timespan of the vowel, it is important to analyze the differences in effect sizes at different points. To do this, we categorized the ten timesteps into three bins, roughly corresponding to the beginning third (beg.), middle third (mid.)

and final third (end) of the vowel tokens and conducted pairwise comparisons between non-focused and focused syllables within each of these time intervals using the *emmeans()* function. The results are provided in Table 16, and as the visual suggests, the effect size of focus on f0 increases over the course of the vowel. Specifically, f0 falls from approximately 127 Hz in the first third of the vowel to 121 Hz in the final portion of the vowel. For focused tokens, however, f0 remains relatively steady across the length of the vowel at around 134-137 Hz. These numbers suggest that, as expected, f0 acts as an acoustic correlate of FOCUS in Mankiyali, with focused tokens exhibiting a significantly higher f0 than non-focused tokens.

Table 16: Pairwise comparisons showing the correlation between FOCUS and f0 for three timesteps, corresponding to the beginning, middle, and final third of the vowel tokens: f0~FOCUS + (1 + FOCUS|speaker) + (1|target.syllable).

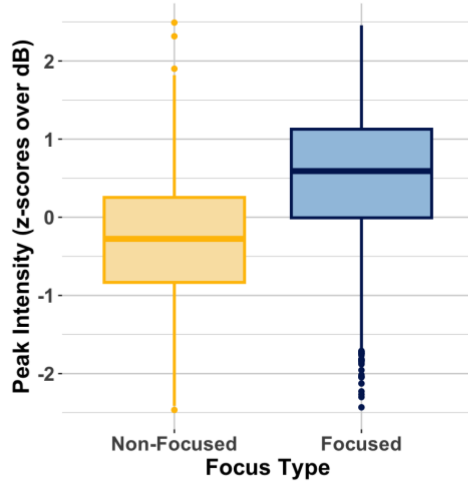
step	nonfocused est. (Hz)	95% CI	focus effect (Hz)	focus effect 95% CI	t-value	p-value
beg	**127**	**[120, 134]**	**10 (8%)**	**[8, 12]**	**10.92**	**< .001 *****
mid	**124**	**[117, 131]**	**11 (9%)**	**[9, 13]**	**12.27**	**< .001 *****
end	**121**	**[114, 128]**	**13 (11%)**	**[11, 15]**	**14**	**< .001 *****

4.2.2 Intensity

The presence of narrow focus also correlated with a simple main effect on peak intensity (formula = PeakIntensity~FOCUS + (1 + FOCUS|speaker) + (1|target.syllable), N = 2,546, intercept (non-focused) = 68.95 dB, Est. (focused) = 4.05 dB, 95% CI = [3.53, 4.57], t = 15.48, p < .001***).[4] As illustrated by the boxplots in Figure 9, focused tokens were, on average, about 1 standard deviation (~4 dB) louder than non-focused tokens with only word-level stress.

Figure 9: Boxplots depicting peak intensity differences between non-focused and focused syllables.

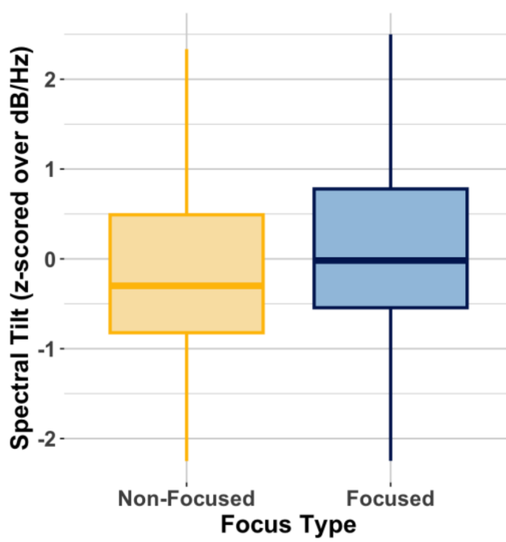
[4] Corresponding l*mer* models examining the effect of FOCUS on **z-scored** peak intensity and **z-scored** spectral tilt values for this and the following section are provided in appendix A.



4.2.3 Spectral Tilt

The simple pairwise comparison of focused and non-focused syllables showed a significant change in spectral tilt of equal magnitude across all syllable types: formula = SpectralTilt~FOCUS + (1 + FOCUS|speaker) + (1|target.syllable), N = 2,546, intercept (non-focused) = -.007 dB/Hz, Est. (focused) = .0008 dB/Hz, 95% CI = [.0006, .001], t = 7.83, p < .001***. This means that focused syllables exhibit a greater increase in intensity at higher frequencies (which is the potential cause of the flatter spectral tilt on stressed syllables) in Mankiyali.

Figure 10: Boxplots depicting spectral tilt differences between non-focused and focused syllables.



4.2.4 Duration

In addition to f0 and intensity, duration was also significantly impacted by the main effect of FOCUS. As shown by the summary of main effects and interaction effects in Table 17, the interaction of FOCUS and SYLLABLE TYPE were significant and positive for both {CVː} and {CVːC} but significant and negative for the three short vowel syllables. These opposite effect sizes indicate that the magnitude of the effect of FOCUS on vowel duration significantly increased for both {CVː} and {CVːC} and likely decreased for the three syllable types with short vowels compared to the main effect of FOCUS across all syllable types. Such interaction effects are unsurprising, given that long vowels need to increase by a greater raw duration between focused and non-focused syllables to maintain a similar proportional increase as short vowels (Hogoboom, 2013).

Table 28: z-scored *lmer* model summary of the effects of FOCUS and SYLLABLE TYPE on duration. Duration ~ FOCUS*SYLLABLE TYPE + (1 + FOCUS|speaker) + (1|target.syllable).

Factors	**Estimate (z-scored over ms)**	**95% CI**	**t-value**	**p-value**
Intercept (grand μ)	.07	[-.04, .19]	1.27	.22
Focused	**.19**	**[.17, .22]**	**14.02**	**< .001 *** **
CV	**-.74**	**[-.96, -.52]**	**-6.56**	**< .001 *** **
CVC	**-.70**	**[-.95, -.46]**	**-5.71**	**< .001 *** **
CVCC	**-.64**	**[-.86, -.42]**	**-5.67**	**< .001 *** **
CVː	**.80**	**[.58, 1.02]**	**7.09**	**< .001 *** **
CVːC	**1.29**	**[1.06, 1.51]**	**11.37**	**< .001 *** **
Focused:CV	**-.06**	**[-.09, -.04]**	**-5.35**	**< .001 *** **
Focused:CVC	**-.05**	**[-.08, -.03]**	**-3.77**	**< .001 *** **
Focused:CVCC	**-.07**	**[-.1, -.04]**	**-5.24**	**< .001 *** **
Focused:CVː	**.09**	**[.07, .11]**	**7.19**	**< .001 *** **
Focused:CVːC	**.10**	**[.08, .12]**	**8.15**	**< .001 *** **

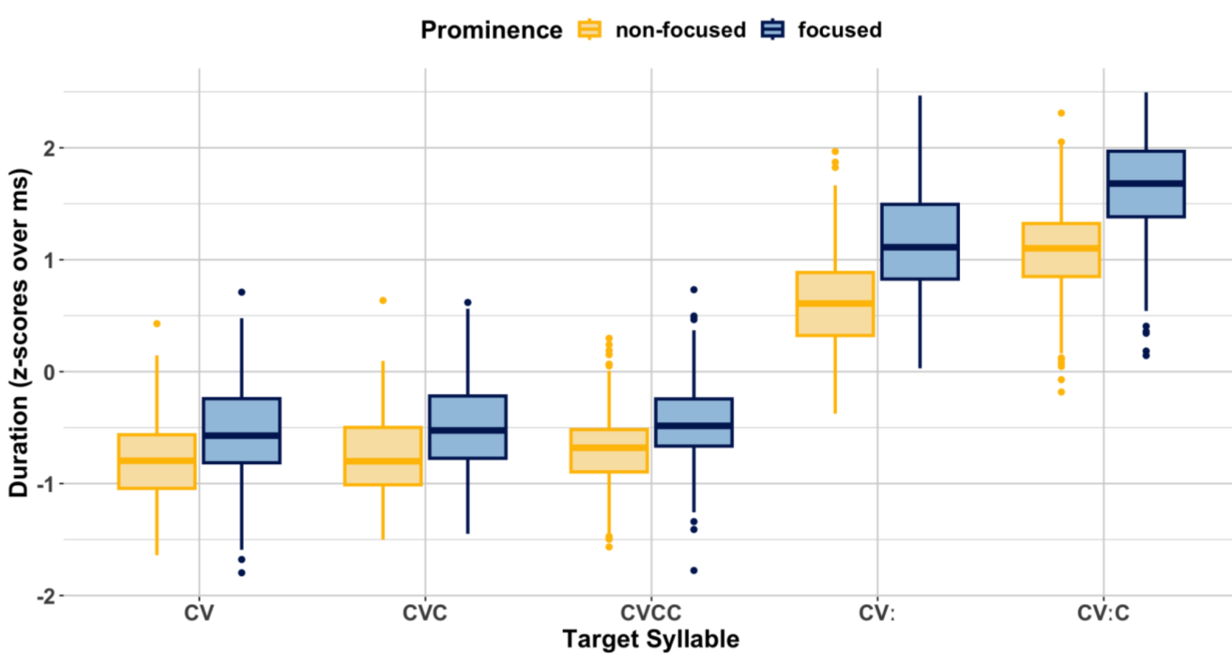
Interestingly, whereas the interaction of word-level stress and syllable type was not significant between long and short vowel syllables for *word-level stress* (Table 17), the magnitude of durational differences *was* significant between long and short vowel syllable types with FOCUS as a main effect. In other words, the inherently longer duration of long vowels led to a larger absolute increase in the effect size caused by the presence of focus-related prominence, but this same effect size was not found for word-level stress.

Table 18 reports pairwise comparisons of duration for each syllable type with corresponding boxplots in Figure 11. The important takeaways for our purposes are that focused and non-focused syllables show a significant mean duration difference of 9-10 ms for short vowel syllables, and this difference is exacerbated by an additional 12-14 ms for long vowel syllables.

Table 18: Pairwise comparisons showing the correlation of FOCUS and vowel duration for each of the five syllable types (CV, CVC, CVCC, CV:, CV:C) in the experiment.

syllable type	nonfocused est. (ms)	95% CI	focus effect (ms)	focus effect 95% CI	t-value	p-value
CV	67	[56, 78]	9 (14%)	[6, 12]	5.92	< .001 ***
CVC	68	[55, 81]	10 (14%)	[6, 13]	5.24	< .001 ***
CVCC	71	[59, 82]	9 (12%)	[5, 12]	5.08	< .001 ***
CV:	122	[111, 134]	22 (18%)	[19, 25]	13.85	< .001 ***
CV:C	142	[131, 153]	23 (16%)	[20, 26]	14.62	< .001 ***

Figure 11: *Boxplots grouped by syllable type depicting vowel duration differences between focused and non-focused syllables.*



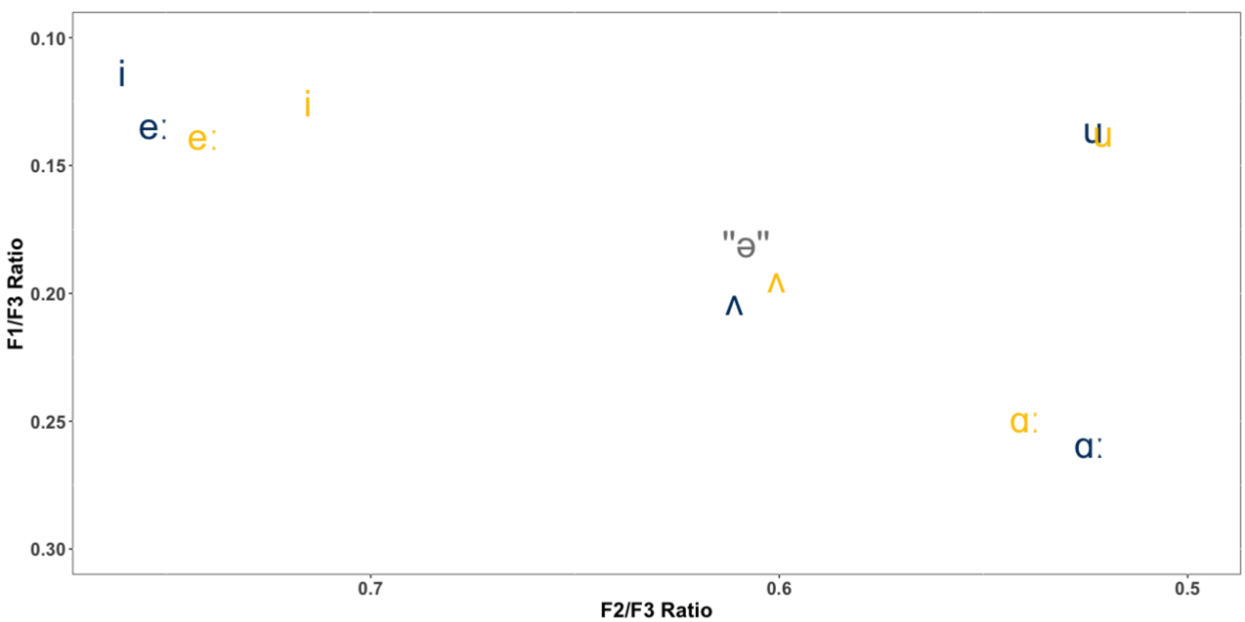
4.2.4 Vowel Space Expansion

To determine whether the presence of focus-related prominence alters vowel quality, the Euclidean distance of each token from a speaker's calculated center point in the F1/F3-F2/F3 vowel space was compared across two levels of prominence: focused tokens (N = 1295) and word-level stressed tokens that lack focus-related prominence (N = 1251). As summarized by the *lmer* in Table 19, a main effect of focus was found for /ɑ:/, /e:/, /ʌ/, and /i/ but not for /u/. These results are visualized by the F1/F3-F2/F3 vowel space plot in figure 12. These results suggest that vowel space expansion is a plausible correlate of focus-related prominence in Mankiyali, though its expression may be limited to certain vowel qualities.

Table 19: Pairwise comparisons showing the correlation between FOCUS and Euclidean Distance from a neutral schwa for each vowel quality. Euclidean Distance is calculated in an F1/F3-F2/F3 space, so estimates are not in Hertz but with respect to the first two formants as ratios of F3. EucDist~FOCUS + (1|speaker) + (1|target.syllable).

vowel	nonfocused est. (EucDist)	95% CI	focus effect (EucDist)	focus effect 95% CI	t-value	p-value
ɑː	**.11**	**[.1, .12]**	**.01 (9%)**	**[.004, .016]**	**3.31**	**= .001 ****
eː	**.14**	**[.13, .16]**	**.01 (7%)**	**[.003, .018]**	**2.71**	**.007 ****
u	.11	[.1, .13]	-.005 (4%)	[-.014, .005]	-.96	.34
ʌ	**.09**	**[.08, .1]**	**-.006 (7%)**	**[-.011, -.001]**	**-2.46**	**.02 ***
i	**.12**	**[.1, .15]**	**.04 (33%)**	**[.029, .053]**	**6.65**	**< .001 *****

Figure 12: F1/F3-F2/F3 vowel space of the mean formant ratio values of all 30 speakers for stressed focused (dark blue) and non-focused (yellow) tokens compared to the mean formant ratio value of the hypothetical neutral schwa (grey) across all speakers.



4.2.5 Summary

This section compared the acoustic properties of word-level stressed syllables in the focused position of a sentence to the acoustic properties of word-level stressed syllables in the non-focused position of a sentence. Whereas duration was the only correlate found to consistently covary as a function of word-level stress, all five acoustic properties measured in the study significantly changed as a function of focus. To sum up, tokens marked with focus-related prominence are not only produced with an increased f0, as predicted, but with the blanket expansion of all phonetic properties measured in the study.

5 Discussion

5.1 Word-level stress correlates

As demonstrated in Section 4.1, duration is the only acoustic property measured in this study found to consistently correlate with word-level stress across syllable types. Nevertheless, while the change is significant for all five syllable types examined, the effect size remains relatively small, ranging from 10-15 ms. Other studies that control for phrasal prominence and cite duration as a correlate to word-level stress have found much larger effect sizes. For instance, stressed vowels in Papiamentu, a Caribbean creole language with contrastive word-level stress, are, on average, 31 ms (39%) longer than their unstressed counterparts (Remijsen & van Heuven, 2002). This durational difference is more than double the durational difference found between unstressed and stressed vowels in the current study, which may lead one to interpret the results as an effect of something other than stress. Nevertheless, at least three reasons exist to support the interpretation of the detected durational effect as a correlate of word-level stress. First, impressionistic judgments of native Mankiyali speakers strongly and consistently agree on stress placement analyses that support the following scale: CVV(C) > CVC(C) > CV.[5] And as Hayes (1995, p.9) asserts, if a group of native speakers agrees on a specific linguistic observation (e.g., presence of stress), "we are justified in taking that agreement as a datum." The reasoning behind this assertion is clear concerning stress: while the acoustic manifestation of prominence may be inconsistent cross-linguistically, the regular identification of its presence provides perceptual evidence in favor of its existence. Moreover, while no controlled experiments have been conducted to confirm these perceptual impressions quantitatively, the potential correlates that correspond with speakers' judgments must be ascertained before such an experiment can be undertaken. Thus, determining the acoustic correlates of stress is a necessary precursor to a perceptual study. In the meantime, speaker impressions provide support for the presence of stress in Mankiyali.

---

[5] See section 2.3 for a discussion of the other two weight levels.

Second, while the effect of stress on duration is small, the directionality of the effect is consistently positive, statistically significant, and likely perceptually detectable across syllable types. That is, all five syllable types measured in the study show an *increase* in duration when stress is present. If stress did not exist, we might expect either no effect at all or, at the very least, random directionality where some syllables exhibit a negative effect such that duration decreases in the presence of stress. Furthermore, the durational effects on all five syllable types meet or exceed the Just Noticeable Difference (JND) value as explained by Stevens (2000, p.228-229), which he defines as the minimum proportion of duration a speech sound must change by to be perceived by speakers at least 50% of the time. For the unstressed vowels analyzed in this study, the JND is about equal to the square root of the length of the vowel for vowels up to 100 ms. For example, the JND of a 100 ms vowel is approximately $\sqrt{100}$ or 10 ms. For vowels consistently longer than 100 ms, the JND is about 10% of the base duration. These JND calculations are reported by Stevens (2000, p.203) to apply cross-linguistically, as they stem from general physiological traits of the human auditory system. Table 19 shows the calculated mean duration of word-level unstressed vowels in Mankiyali, the JND for each syllable type based on that mean duration, and the actual mean durational change between word-level unstressed vowels and word-level stressed vowels observed in this study. The effect size of word-level stress eclipses the JND for all syllable types except {CV:}, which is about equal with the predicted JND. It is possible that, since all {CV:} syllables that do not receive primary stress are marked with secondary stress, the durational difference between primary and secondary stressed {CV:} syllables is diminished compared to the other syllable types. Conversely, for {CV:C}, this attenuation was likely counteracted by the analysis of {CV:C} in the word-final position, which coincides with word-final lengthening effects.

Table 19: Results from the present study on word-level stress showing the mean duration of unstressed vowels in each of the five syllable types examined, the predicted JND based on those durations, and the actual durational change correlating with word-level stress.

Syllable Type	mean duration (ms)	JND (ms)	observed change (ms)
CV	56	7 (13%)	11 (19%)
CVC	57	8 (14%)	11 (18%)
CVCC	60	8 (13%)	10 (17%)

CV:	112	11 (10%)	11 (10%)
CV:C	128	13 (10%)	15 (11%)

Finally, though 10-15 ms constitutes a relatively small effect of stress on duration compared to effects found for languages like Papiamentu cited above, such an increase is within the normal range of durational differences reported in the literature in studies that control for prominence at different levels. For instance, in Gordon & Roettger's (2021) cross-linguistic survey of published studies examining the acoustic correlates of stress, they found that only 19 of 85 studies in the survey successfully measured word-level stress in positions without focus-related prominence and away from phrasal edges where pitch excursions are rampant. Out of those 19 studies, only a single language - Jordanian Arabic (De Jong & Zawaydeh, 1999, 2002) - utilizes a weight-sensitive stress system, and the durational difference between unstressed and stressed vowels in the non-focused position was ~12 ms (23%) for short vowels and ~27 ms (14%) for long vowels. They do not report that long vowels without primary stress receive secondary stress in Jordanian Arabic, so the difference between their results for long vowels and the ones reported here should be interpreted with this in mind. Taking the effects of secondary stress into account, the impact of word-level stress on duration in Mankiyali is similar in magnitude to the impact of stress on duration in Jordanian Arabic. Thus, it is possible that correlates undergo more dramatic changes for languages in which stress placement plays a role in lexical identification (i.e., lexically contrastive stress languages like Papiamentu). Conversely, in languages with weight-sensitive stress systems like Mankiyali and Jordanian Arabic, where stress placement is predictable, the magnitudes of the correlates of stress are dampened.

Taken together, we assume that the effect size of duration found in this study is large and consistent enough for speakers to rely on as a correlate of word-level stress. Moreover, though a comprehensive survey of various stress systems is needed, the durational effect size in Mankiyali could be small because the stress system is weight-based and predictable as opposed to lexically contrastive.

A strict version of the FLH, predicts that duration should not be called upon as a primary acoustic correlate of prominence in Mankiyali. Contrary to these predictions, not only was duration found to be a significant acoustic correlate of both word-level stress and focus-related prominence, but it was also the *only* acoustic property found to consistently correlate with word-level stress. In light of these results, it seems that the FLH, in its strongest form, does not hold up to scrutiny. Rather, an increase in vowel duration due to word-level stress does not close the gap between short and long vowels in any meaningful way. Specifically, the mean difference between short vowels with word-level stress and long vowels without word-level stress ranges from 38 ms on the low end to 62 ms on the high end, and a durational effect size of ~10 ms due to word-level stress still allows speakers to easily distinguish between the two vowel types. The use of duration as the primary acoustic correlate of word-level stress in Mankiyali, then, argues against the claim that acoustic properties utilized in phonologically contrastive domains cannot be used to indicate the location of stress.

5.2 Durational differences from word-level stress vs. compression effects

One may argue that the durational differences we have found between stressed and unstressed syllables may, in fact, reflect a compression effect – in which additional segments result in the shortening of other segments – rather than a correlate of stress. The idea, similar to the concept of polysyllabic shortening discussed in (Lehiste, 1972), is that the addition of another segment to a word (e.g., adding a final consonant to a CV.CV word) results in the duration of the other segments compressing to compensate. Since the measurements of the effect of stress on duration in this study were typically examined by adding an additional segment to the final syllable to shift stress, our finding that duration correlates with stress may be confounded with the compression effect. To test whether this is the case, we conducted a pairwise comparison, evaluating the effect of stress on duration between the penultimate vowels in [kʌ.ˈt͡sʌɽ] and [ˈpʌtʰ.re], using every sentence three token for each of these two words from all speakers in the experiment (N = 110).

Both words contain the same number of onsets, nuclei, and codas, so the durational difference should not arise if it was caused by word-internal compression effects. Additionally, the vowel quality of the target syllables is the same, and both vowels are immediately preceded and followed by voiceless stops/affricates. These similarities should mitigate any potential contextual or vowel quality effects on duration. If stress is responsible for the durational difference rather than a compression effect, we expect to observe an effect size similar to the results of the larger *lmer* models from section 4.1.1, in which the presence of stress correlated with about a 10 ms increase in vowel duration. Specfically, we expect the stressed penultimate vowel in [ˈpʌtʰ.re] to be about 10 ms longer than the unstressed penultimate vowel in [kʌ.ˈt͡sʌɽ]. Conversely, if compression effects were responsible for the observed effects in section 4.1.1, this effect size should disappear between these two words.

Moreover, Waals (1999) argues that compression effects can occur *syllable*-internally, such that, all else equal, closed syllables will exhibit shorter vowel durations than open syllables. If correct, this compression effect at the syllable level would counteract any durational increases due to stress. As such, if penultimate vowels in [ˈpʌtʰ.re] are significantly longer than penultimate vowels in [kʌ.ˈt͡sʌɽ], it is likely that the effect is stress-related rather than caused by a compression effect.

Results of the lmer model are given in Table 20 and highlight that, as predicted if stress were giving rise to the durational increase, the unstressed penultimate vowel in [kʌ.ˈt͡sʌɽ] is indeed approximately 10 ms shorter than the stressed penultimate vowel in [ˈpʌtʰ.re]. This is consistent with the interpretation that stress is the root cause of the durational effect and contradicts the view that a compression effect lies behind the durational differences. In sum, while segments may compress durationally in proportion to the number of segments in a word, compression does not seem to explain the durational effects reported in this study.

Table 20: simple pairwise comparison *lmer* model summary comparing the effect of STRESS on duration between unstressed penultimate [**kʌ**.ˈt͡sʌɽ] and stressed penultimate [ˈ**pʌtʰ**.re]: Duration~STRESS + (1|speaker).

Factors	Estimate (ms)	95% CI	t-value	p-value
Intercept (**kʌ**.ˈt͡sʌɽ)	57	[52.78, 60.63]	28.34	< .001 ***
ˈ**pʌtʰ**.re	**10**	**[7.11, 13.28]**	**6.48**	**< .001 *** **

5.3 Correlates of focus-related prominence

Whereas duration was the only acoustic property found to consistently correlate with word-level stress, focus-related prominence correlates include all the acoustic properties measured in this study. We take it that the importance of highlighting new information in a dialogue requires the recruitment of a host of correlates to ensure the information is perceived correctly. While many studies have argued that f0 is the primary correlate of focus, acoustic properties seem to be mobilized across the board in Mankiyali to highlight the words in narrow focus.

Additionally, the difference in f0 between focused and non-focused constituents is smaller than we expected, but the effect size falls in line with other studies that measure the effect of narrow focus on f0. For instance, Greek shows a mean f0 difference of 8 Hertz between non-focused and focused constituents, and Hungarian shows a difference of 17 Hertz (Vogel et. al., 2016).

5.4 The Mankiyali Stress Criterion

As discussed earlier, Paramore (2021) analyzes the Mankiyali stress system as involving three distinct levels of syllable weight: {CVː(C)} > {CVC(C)} > {CV}. In contrast, the present study hypothesized that Mankiyali stress is sensitive to a more fine-grained, five-level weight hierarchy: {CVːC} > {CVː} > {CVCC} > {CVC} > {CV}. Our goal was to determine whether the acoustic correlates of stress found in previously established weight distinctions (e.g., {CVC} > {CV}, {CVː} > {CVC}) also extend to weight distinctions that have not been previously reported (e.g., {CVːC} > {CVː} and {CVCC} > {CVC}). The results in Section 4.1.1 reveal a systematic relationship between vowel duration and the hypothesized stress scale. Specifically, we found significantly increased duration of penultimate vowels in {ˈCV.CV} words compared to {CV.ˈCVC} words, supporting the claim that {CVC} outweighs {CV}. Similarly, {ˈCVC.CV} words showed greater penultimate vowel duration than {CVC.ˈCVː} words, reinforcing that {CVː} outweighs {CVC}. Finally, {ˈCVː.CVː} words showed significantly greater penultimate vowel duration than {CVː.ˈCVːC} words, aligning with the hypothesis that {CVːC} > {CVː}.

While the magnitude of these durational effects alone is not sufficient to *prove* the proposed five-tiered stress scale, it does provide added support, especially when coupled with native speaker judgements about the placement of primary stress. This study, then, lends quantitative support to the impressionistic view that the Mankiyali stress criterion employs at least four levels of weight: {CVːC} > {CVː} > {CVCC, CVC} > {CV}.

Unfortunately, we were unable to find near minimal pairs of the shape {ˈCVC.CVC} vs. {CVC.ˈCVCC} to test whether {CVCC} outweighs {CVC}. Nevertheless, we assume that – if superheavy {CVːC} outweighs heavy {CVː} in the language – the same is true of superheavy {CVCC} outweighing heavy {CVC}. In sum, the data provided in this paper provides an empirical basis for analyzing Mankiyali's stress system as containing five distinct levels of stress: {CVːC} > {CVː} > {CVCC} > {CVC} > {CV}. Such a scale is, to our knowledge, unattested in the world's languages, so – while further research is needed to confirm these results – the Mankiyali stress system represents a unique system that is highly complex in its weight divisions.

6 Conclusion

This paper reported results from an experiment seeking to determine the acoustic correlates of word-level stress and focus-related prominence in Mankiyali. Tokens were analyzed in two sentence types, one in which the word was focused and the other in which the word was non-focused.

The results from the experiment reveal that duration is the sole acoustic property measured in this study found to conistently correlate with word-level stress across syllable types. For focus-related prominence, all acoustic properties measured covaried with its presence or absence. Moreover, we found weak evidence suggesting that different syllable structures call upon different acoustic properties to varying degrees to signal word-level stress.

Finally, the results of the present study lend support to a 5-way stress criterion in Mankiyali: {CVːC} > {CVː} > {CVCC} > {CVC} > {CV}. Such a fine-grained and complex criterion is extremely rare and perhaps otherwise unattested cross-linguistically. As such, while further research is needed to confirm the conclusions drawn here, the Mankiyali stress system may extend the limits of how intricate weight-sensitive stress systems can be.

In terms of future directions, the analysis of focus-related prominence in the present study is limited in scope, exploring acoustic patterns in the domain of the stressed syllable rather than across entire sentences and phrases. In subsequent work, a detailed examination of intonational phrases and their associated boundaries, along with different kinds of focus-related prominence, should be performed. In addition, a perception study examining speakers' sensitivity to the observed changes in duration in determining word-level stress placement is welcomed to add support to the findings presented here.

Appendix A

Table 21: *Experiment tokens with glosses*

Word	Gloss	Word	Gloss
ˈkʰʌb-ʌ	left-M.SG	kʰʌˈbɑːɽ	rope.NOM.F.SG
ˈmʌz-ʌ	enjoyment-M.SG	mʌˈzʌɽ	inside.SG.POST
ˈkʌt͡s-ʌ	unripe-M.SG	kʌˈt͡sʌɽ	mule.F.SG

ˈjʌk-ʌ	one-OBL	jʌˈkʌɽ	wicked.SG
ˈt͡suk-i	sour-M.PL	t͡suˈkiː-ɳ	a weed-GEN.F.SG/PL
ˈgʌɽk-u	to thunder-3SG.PST	gʌɽˈk-uː	thunder-M.SG
ˈmuski	smile.F.SG.PST	musˈkiː	smile.F.SG.PST.Q
ˈbelt-i	bucket-F.PL	belˈt-iːz	bucket.F.PL-DAT
ˈkutr-e	to chop-1PL.PRS	kutˈr-eːz	to chop-INF
ˈpʌtʰr-e	to lay-1PL.PRS	pʌtʰˈr-eː-z	to lay-INF
ˈbẽːɣĩː	rooster.M.PL	bẽːˈɣĩː-z	rooster.M.PL-DAT
ˈpeːkiː	barber's razor.F.PL	peːˈkiː-z	barber's razor.F.PL-DAT
ˈkɑːɣ-ɑː	crow-M.PL	kɑːˈɣɑːz	paper.M.SG
ˈɖeːkiː	postman.M.PL	ɖeːˈkiː-z	postman.M.PL-DAT
ˈt͡ʃeːbiː	key.F.PL	t͡ʃeːˈbiː-z	key.F.PL-DAT
ˈsʌŋg-jo	friend-ERG.PL	sʌŋg-ˈtoːb	friend-NMLZ
ˈgʌnd-jo	dirty-OBL.PL	gʌnd-ˈgiː	dirty-NMLZ
ˈmistri-jo	mason-ERG.PL	mistˈriː	mason.SG
ˈistri-jo	iron-OBL.PL	istˈriː	iron.F.SG/PL
ˈd͡ʒʌntri-jo	calendar-OBL.PL	d͡ʒʌntˈriː	calendar.F.SG/PL
lʌŋˈgɑːr	hand grinder.F.SG	ˈɖɑːŋgɑːr	bull.M.SG
kʌˈdɑːr	respect.F.SG	ˈkɑːrdɑːr	leader.SG
goˈdɑːɽ	knapsack.F.SG	ˈd͡ʒoːndɑːr	wheat crop weed.M.SG
mʌzˈdɑːr	tasty.SG	ˈdɑːɣdɑːr	stained.SG
bʌzˈvɑːɳ	part of a game.M.SG	ˈɑːzvɑːɳ	sky.M.SG

Leipzig Glossing Abbreviations

Abbreviations used throughout the paper follow the Leipzig Glossing Rules (LGR, http://www.eva.mpg.de/lingua/resources/glossing-rules.php) and are summarize below.

3	third person
DAT	dative
ERG	ergative
F	feminine
GEN	genitive
IMP	imperative
INF	infinitive
M	masculine
NMLZ	nominalizer/nominalization
OBL	oblique
PL	plural
POST	postposition
PST	past
Q	question particle
SBJV	subjunctive
SG	singular

Focus change lmer model

The following table provides the full lmer model from the analysis of focus change in section 3.2 for speaker AZ1.

Table 22: *lmer* model showing the effect of sentence type on the mean f0 of target word vowels for speaker AZ1: token_MeanF0 ~ sʌŋgi_MeanF0 + Focus + (1 | target_token).

Factors	Estimate (Hz)	95% CI	t-value	p-value
Intercept (grand μ)	152	[148, 156]	67.35	< .001 ***
Focused target vowel	**8**	**[6, 10]**	**9.81**	**< .001 *** **
Non-focused target vowel	**-8**	**[-10, -6]**	**-9.81**	**< .001 *** **
[sʌŋgi] mean f0	**0.63**	**[0.36, 0.89]**	**4.60**	**< .001 *** **

z-scored lmer models

The following z-scored lmer models correspond to the lmer models analyzing raw values of peak intensity in section 4.2.2 and raw values of spectral tilt in 4.2.3.

Table 23: z-scored pairwise comparison *lmer* model summary showing the effect FOCUS on Peak Intensity: PeakIntensity~FOCUS + (1 + FOCUS|speaker) + (1|target.syllable).

Factors	Estimate (z-scored over dB)	95% CI	t-value	p-value
Intercept (non-focused)	-.4	[-.63, -.16]	-3.3	.003 .
focused	**.89**	**[.79, .99]**	**17.08**	**< .001 *** **

Table 24: z-scored pairwise comparison *lmer* model summary showing the effect FOCUS on Spectral Tilt: SpectralTilt~FOCUS + (1 + FOCUS|speaker) + (1|target.syllable).

Factors	Estimate (z-scored over dB/Hz)	95% CI	t-value	p-value
Intercept (non-focused)	-.10	[-.43, .23]	-.58	.57
focused	**.32**	**[.25, .40]**	**8.36**	**< .001 *** **

Appendix B: Visuals and lmers grouped by syllable type and near-minimal pairs for word-level duration.

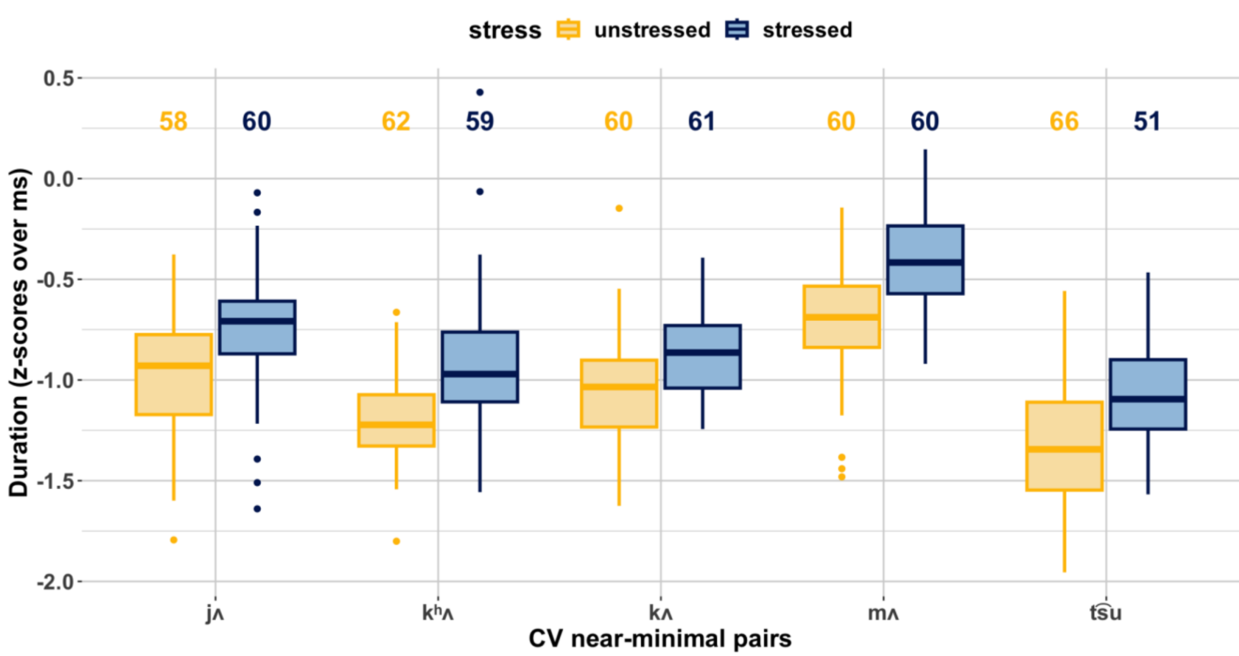
Each plot visualizes a comparison between unstressed vs. stressed near-minimal pairs for one of the five syllable types. The total count of each token is provided above the corresponding boxplot. Near minimal pairs are labeled based on their similar target syllable, as shown in table XX. Pairwise comparison lmer tables for each set of near-minimal pairs in a syllable type are provided below each plot. The following model specification was used for all lmer models comparing near minimal pairs: Duration~STRESS + (1|speaker).

Table 25: 25 sets of near minimal pairs. The target syllable for each token is bolded, and the notation used in the below plots and lmers for each set of near-minimal pairs is placed to the left of each set.

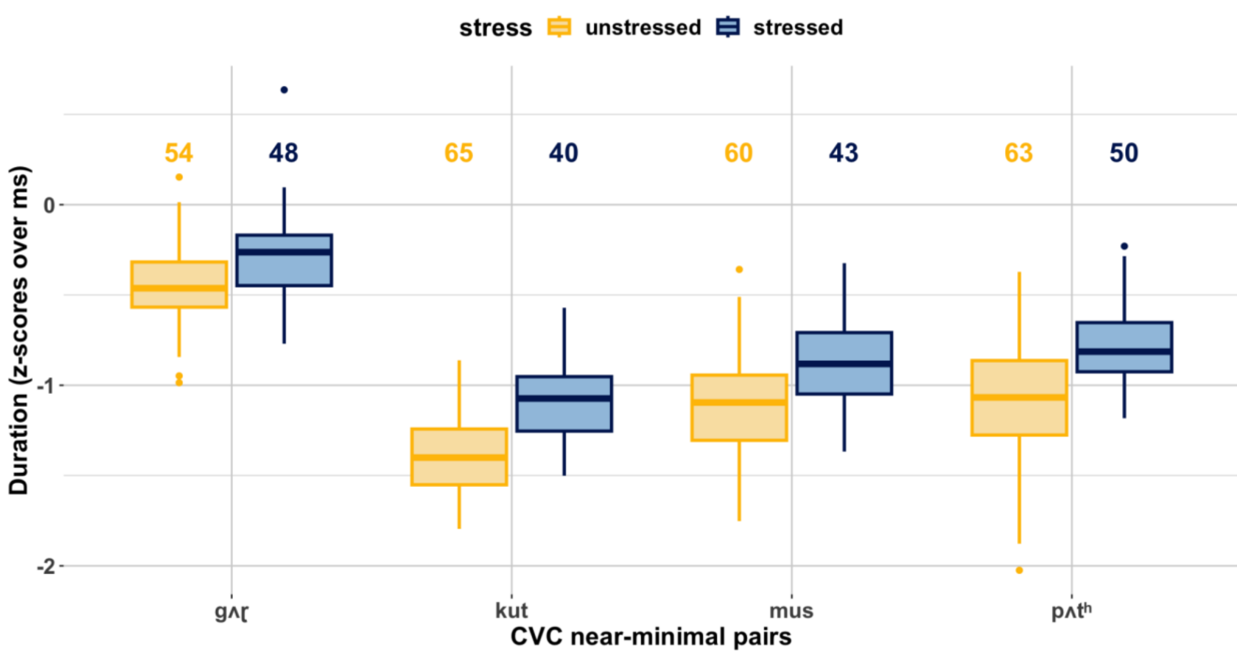
CV			**CVC**			**CVː**		
kʰʌ:	ˈ**kʰʌ**.bʌ	**kʰʌ**.ˈbɑːɽ	**gʌɽ:**	ˈ**gʌɽ**.ku	**gʌɽ**.ˈkuː	**bẽː:**	ˈ**bẽː**.ˌɣĩː	ˌ**bẽː**.ˈɣĩːz
mʌ:	ˈ**mʌ**.zʌ	**mʌ**.ˈzʌɽ	**mus:**	ˈ**mus**.ki	**mus**.ˈkiː	**peː:**	ˈ**peː**.ˌkiː	ˌ**peː**.ˈkiːz
kʌ:	ˈ**kʌ**.t͡sʌ	**kʌ**.ˈt͡sʌɽ	**kut:**	ˈ**kut**.re	**kut**.ˈreːz	**kɑː:**	ˈ**kɑː**.ˌɣɑː	ˌ**kɑː**.ˈɣɑːz
jʌ:	ˈ**jʌ**.kʌ	**jʌ**.ˈkʌɽ	**pʌtʰ:**	ˈ**pʌtʰ**.re	**pʌtʰ**.ˈreːz	**ɖeː:**	ˈ**ɖeː**.ˌkiː	ˌ**ɖeː**.ˈkiːz
t͡su:	ˈ**t͡su**.ki	**t͡su**.ˈkiːɳ				**t͡ʃeː:**	ˈ**t͡ʃeː**.ˌbiː	ˌ**t͡ʃeː**.ˈbiːz

CVCC			**CVːC**		
sʌŋg:	ˈ**sʌŋg**.jo	**sʌŋg**.ˈtoːb	**ŋgɑːr:**	lʌŋ.ˈ**gɑːr**	ˈɖɑːŋ.ˌ**gɑːr**
gʌnd:	ˈ**gʌnd**.jo	**gʌnd**.ˈgiː	**dɑːr:**	kʌ.ˈ**dɑːr**	ˈkɑːr.ˌ**dɑːr**
mist:	ˈ**mist**.ri.ˌjo	**mist**.ˈriː	**o(n)dɑːr:**	go.ˈ**dɑːɽ**	ˈd͡ʒoːn.ˌ**dɑːr**
ist:	ˈ**ist**.ri.ˌjo	**ist**.ˈriː	**zdɑːr:**	mʌz.ˈ**dɑːr**	ˈdɑːɣ.ˌ**dɑːr**
d͡ʒʌnt:	ˈ**d͡ʒʌnt**.ri.ˌjo	**d͡ʒʌnt**.ˈriː	**vɑːɳ:**	bʌz.ˈ**vɑːɳ**	ˈɑːz.ˌ**vɑːɳ**

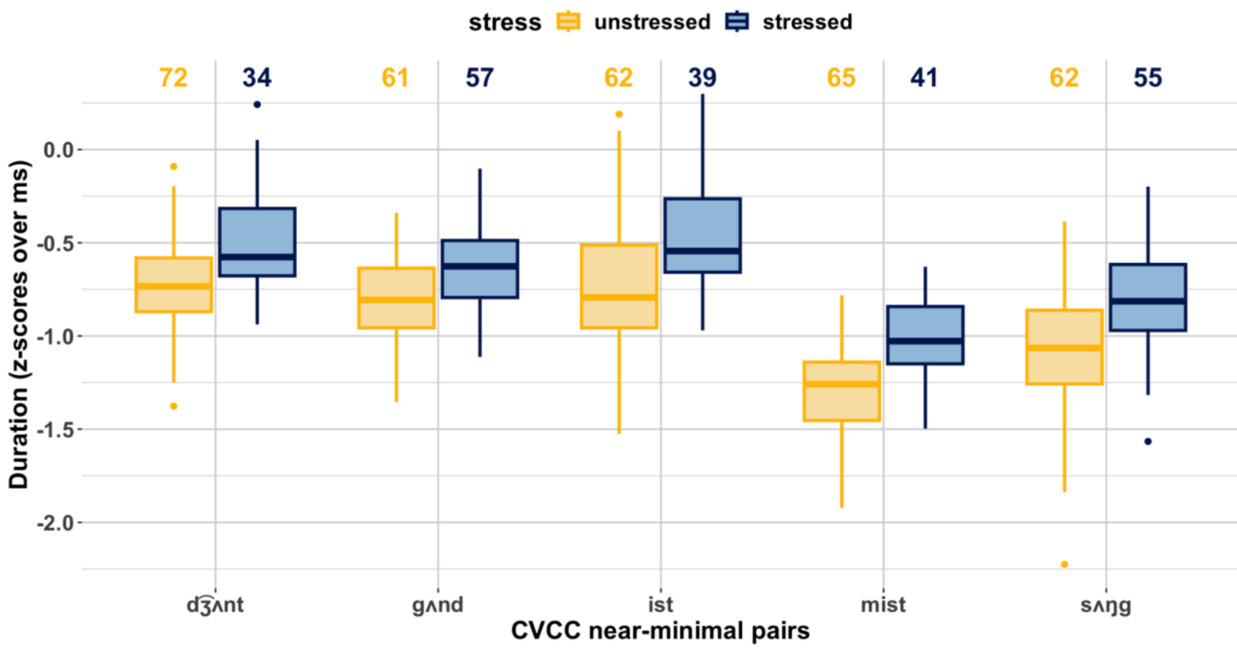
<u>unstressed vs. stressed: Duration</u>



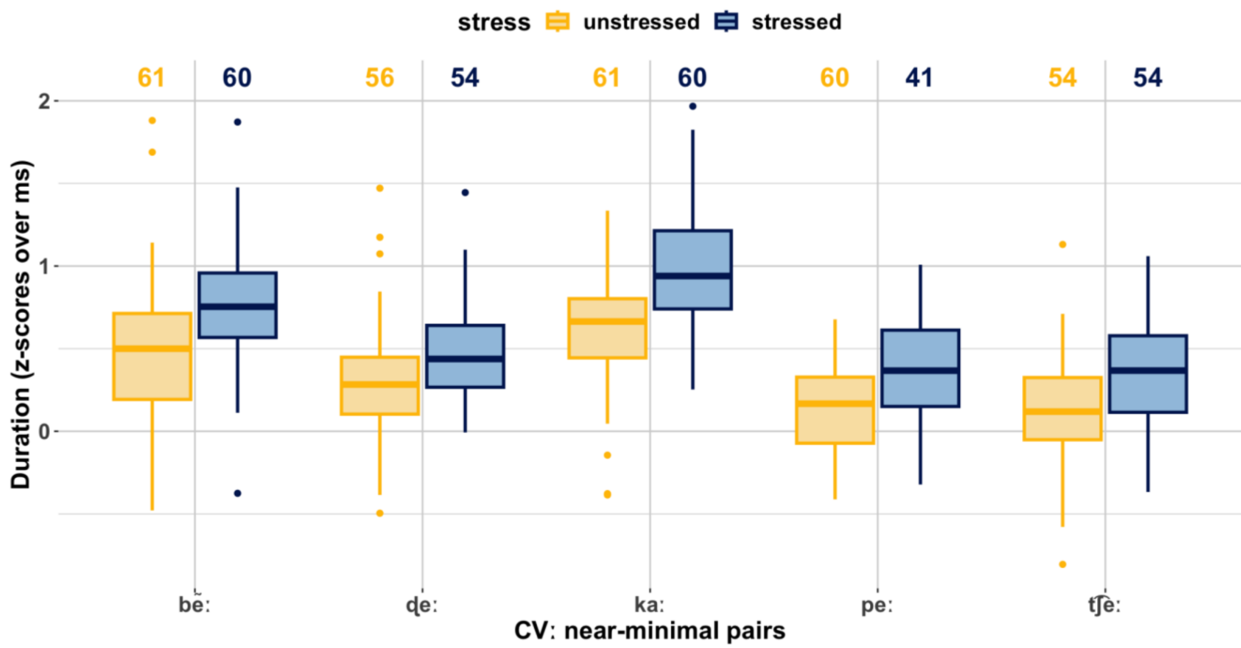
Factors	Estimate (ms)	95% CI	t-value	p-value
Intercept (unstressed kʌ)	57	[52.84, 60.58]	28.72	< .001 ***
stressed kʌ	**7**	**[4.87, 9.61]**	**5.99**	**< .001 *** **
Intercept (unstressed kʰʌ)	51	[46.63, 55.04]	23.67	< .001 ***
stressed kʰʌ	**12**	**[8.7, 14.54]**	**7.8**	**< .001 *** **
Intercept (unstressed mʌ)	69	[65.14, 73.82]	31.36	< .001 ***
stressed mʌ	**13**	**[9.99, 15.43]**	**9.15**	**< .001 *** **
Intercept (unstressed jʌ)	60	[55.13, 64.64]	24.67	< .001 ***
stressed jʌ	**9**	**[6.05, 11.67]**	**6.18**	**< .001 *** **
Intercept (unstressed t͡su)	45	[40.37, 50.38]	17.77	< .001 ***
stressed t͡su	**11**	**[7.46, 14.31]**	**6.24**	**< .001 *** **



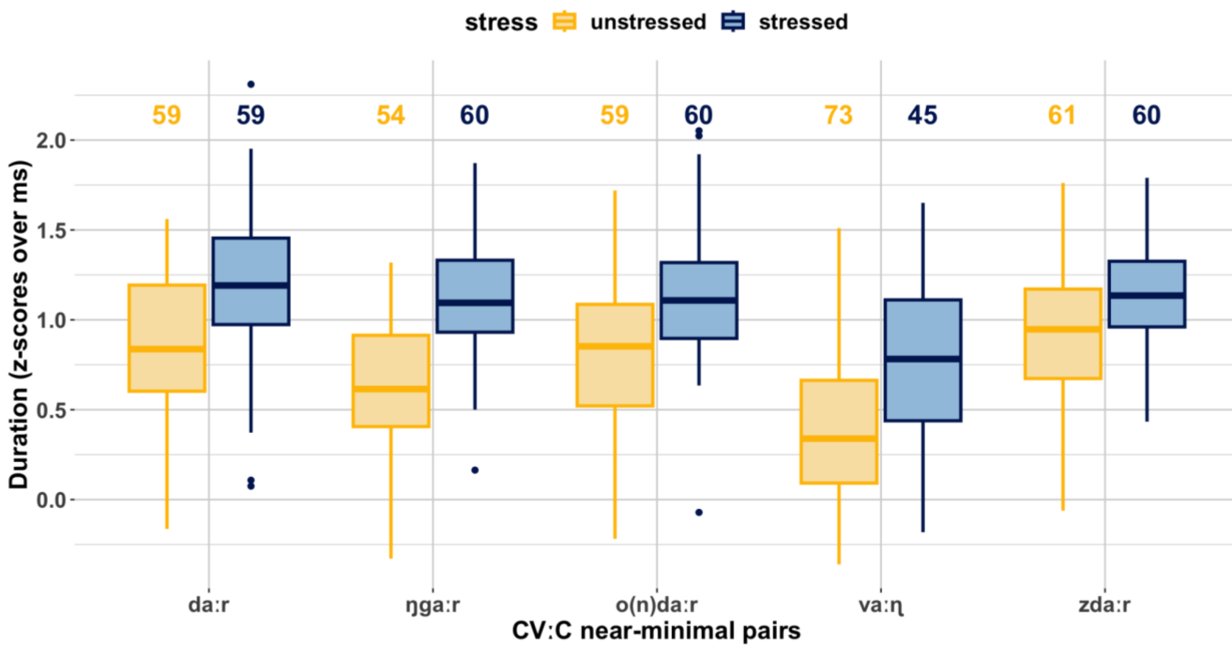
Factors	Estimate (ms)	95% CI	t-value	p-value
Intercept (unstressed gʌɽ)	79	[74.47, 84.01]	32.56	< .001 ***
stressed gʌɽ	**7**	**[3.97, 9.14]**	**4.97**	**< .001 *** **
Intercept (unstressed kut)	43	[38.87, 46.76]	21.26	< .001 ***
stressed kut	**11**	**[8.45, 13.9]**	**8.04**	**< .001 *** **
Intercept (unstressed mus)	53	[48.72, 56.94]	25.2	< .001 ***
stressed mus	**11**	**[7.49, 14.69]**	**6.03**	**< .001 *** **
Intercept (unstressed pʌtʰ)	55	[50.48, 59.39]	24.16	< .001 ***
stressed pʌtʰ	**12**	**[8.49, 14.62]**	**7.39**	**< .001 *** **



Factors	Estimate (ms)	95% CI	t-value	p-value
Intercept (unstressed d͡ʒʌnt)	69	[64.1, 73.28]	29.33	< .001 ***
stressed d͡ʒʌnt	**8**	**[5.23, 11.68]**	**5.13**	**< .001 *** **
Intercept (unstressed gʌnd)	66	[61.43, 69.63]	31.32	< .001 ***
stressed gʌnd	**7**	**[4.31, 9.08]**	**5.5**	**< .001 *** **
Intercept (unstressed ist)	68	[63.33, 73.47]	26.45	< .001 ***
stressed ist	**14**	**[9.9, 17.22]**	**7.26**	**< .001 *** **
Intercept (unstressed mist)	46	[42.03, 50.41]	21.63	< .001 ***
stressed mist	**12**	**[8.89, 15.7]**	**7.08**	**< .001 *** **
Intercept (unstressed sʌŋg)	55	[50.54, 58.69]	26.27	< .001 ***
stressed sʌŋg	**12**	**[8.68, 15.67]**	**6.83**	**< .001 *** **



Factors	Estimate (ms)	95% CI	t-value	p-value
Intercept (unstressed bẽ:)	118	[110.14, 125.42]	30.21	< .001 ***
stressed bẽ:	**13**	**[8.65, 16.53]**	**6.26**	**< .001 *****
Intercept (unstressed ɖe:)	110	[104.88, 115.47]	40.77	< .001 ***
stressed ɖe:	**6**	**[2.84, 10.09]**	**3.5**	**< .001 *****
Intercept (unstressed ka:)	124	[117.31, 130.91]	35.76	< .001 ***
stressed ka:	**14**	**[10.12, 17.3]**	**7.48**	**< .001 *****
Intercept (unstressed pe:)	105	[98.36, 111.07]	32.3	< .001 ***
stressed pe:	**9**	**[5.13, 12.68]**	**4.62**	**< .001 *****
Intercept (unstressed t͡ʃe:)	104	[97.61, 110.02]	32.8	< .001 ***
stressed t͡ʃe:	**9**	**[5.93, 12.15]**	**5.7**	**< .001 *****



Factors	Estimate (ms)	95% CI	t-value	p-value
Intercept (unstressed da:r)	133	[123.4, 143.16]	26.44	< .001 ***
stressed da:r	**14**	**[10.23, 18.61]**	**6.74**	**< .001 *** **
Intercept (unstressed ŋga:r)	124	[115.8, 132.87]	28.56	< .001 ***
stressed ŋga:r	**19**	**[15.56, 23.41]**	**9.73**	**< .001 *** **
Intercept (unstressed o(n)da:r)	132	[121.96, 142.2]	25.58	< .001 ***
stressed o(n)da:r	**13**	**[8.44, 18.46]**	**5.26**	**< .001 *** **
Intercept (unstressed va:ɳ)	114	[105.61, 122.37]	26.66	< .001 ***
stressed va:ɳ	**15**	**[10.19, 20.81]**	**5.72**	**< .001 *** **
Intercept (unstressed zda:r)	135	[126.48, 143.62]	30.89	< .001 ***
stressed zda:r	**10**	**[5.68, 13.58]**	**4.78**	**< .001 *** **

References

Abbasi, A., & S. Hussain. (2015). Phonetic analysis of lexical stress in Sindhi. *Sindh Univ. Research Journal (Science Series)*, 47(4), pp.749-756.

Abbasi, A., Pathan, H. & Channa, M. (2018). Experimental Phonetics and Phonology in Indo-Aryan & European Languages. *Journal of Language and Cultural Education*, 6(3) 21-52. https://doi.org/10.2478/jolace-2018-0023

Abercrombie, David. 1976. Stress and some other terms. *Work in Progress* 9, 51-53.

Almbark, Rana, Nadia Bouchhioua & Sam Hellmuth. 2014. Acquiring the phonetics and phonology of English word stress: Comparing learners from different L1 backgrounds. *Concordia Working Papers in Applied Linguistics* 5, 19–35.

Bates, Douglas, Martin Maechler, Ben Bolker, & Steven Walker. (2023). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software 67(1),* p.1-48.

Berinstein, A. E. (1979). A Cross-Linguistic Study on the Perception and Production of Stress. *UCLA Working Papers in Phonetics* 47.

Boersma, Paul & David Weenink. (2015). Praat: Doing phonetics by computer (version 5.4.21). http://www.fon.hum.uva.nl/praat/.

Bowers, D. (2019). Positional vs. sonority-driven stress in Gujarati: New experimental evidence. *Phonological Data and Analysis*, 1(1), 1–28. https://doi.org/10.3765/pda.v1art1.19

Bruggeman, A., Louriz, N., Almbark, R., & Hellmuth, S. (2021). Acoustic correlates of lexical stress in Moroccan Arabic. *Journal of the International Phonetic Association*, 51(3), 425-449. doi:10.1017/S002510032000002X

Bouchhioua, Nadia. (2008). *The acoustic correlates of stress and accent in Tunisian Arabic: A comparative study with English*. Ph.D. dissertation, Université de 7 Novembre, l'Institut Supérieur des Langues de Tunis.

Butt, Miriam & Tracy H. King. (1996b). Structural topic and focus without movement. Paper presented at *Frist LFG Conference*, Grenoble, France, August 26-28.

Cardona, George & Dhanesh Jain. (2003). *The Indo-Aryan languages.* Routledge.

Caplow, N. (2016). Stress patterns and acoustic correlates of stress in Balti Tibetan. *Himalayan Linguistics*, 15(2). http://dx.doi.org/10.5070/H915232601

CoRSAL. (2023). The Computational Resource for South Asian Languages. corsal.unt.edu

De Jong, Kenneth, & Bushra Adnan Zawaydeh. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics* 27, 3-22.

De Jong, Kenneth, & Bushra Adnan Zawaydeh. (2002). Comparing stress, lexical focus, and segmental focus patterns of variation in Arabic vowel duration. *Journal of Phonetics* 30, 53- 75.

Dyrud, L. O. (2001). *Hindi-Urdu: stress accent or non-stress accent?* Master's Thesis, University of North Dakota, Grand Forks.

Féry, Caroline. (2013). Focus as prosodic alignment. *Natural Language and Linguistic Theory* 31, p. 683-734.

Fry, Dennis B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27(4), 765–768.

Garellek, M., & White, J. (2015). Phonetics of Tongan stress. *Journal of the International Phonetic Association*, 45(1), 13-34. doi:10.1017/S0025100314000206

Goedemans, Rob & Harry Van der Hulst. (2013). Fixed stress locations. In Matthew Dryer & Martin Haspelmath (eds.), *The World Atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/14 (accessed 08 December, 2021).

Gordon, M. (2004). A Phonological and Phonetic Study of Word-Level Stress in Chickasaw. *International Journal of American Linguistics*, 70(1), 1–32.

Gordon, Matthew K. (2011). Stress: Phonotactic and phonetic evidence. In M. Van Oostendorp, C.J. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell Companion to Phonology* (2). https://doi.org/10.1002/9781444335262.wbctp0039

Gordon, Matthew K. (2014). Disentangling stress and pitch accent: A typology of prominence at different prosodic levels. In Van der Hulst, Harry (ed.), *Word Stress: Theoretical and Typological Issues*, pp. 83-118. Oxford University Press.

Gordon, Matthew K. & Ayla Applebaum. (2010). Acoustic correlates of stress in Turkish Karbadian. *Journal of the International Phonetic Association* 40(1), pp. 35-58. doi:10.1017/S0025100309990259

Gordon, Matthew K., & Timo B. Roettger. (2021). Methodological issues in the study of word stress correlates. *Linguistics Vanguard*, *3*(1).

Gussenhoven, Carlos. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Halliday, M. A. K. (1967b). Notes on transitivity and theme in English, part 2. *Foundations of Language* 10:41-53.

Harnsberger, J.D. (1994). *Towards an intonational phonology of Hindi*. Master's thesis, University of Florida.

Hayes, Bruce. (1995). *Metrical stress theory: Principles and case studies.* Chicago, IL & London: The University of Chicago Press.

Heldner, M., & E. Strangert. (2001). Temporal effects of focus in Swedish. *Journal of Phonetics* 29, pp.329-361.

Hogoboom, Anya. (2013). Reanalyzing final consonant extrametricality: a proportional theory of weight. *Journal of Comparative German Linguistics* 16, 1-31.

Hussain, Sarmad. (1997). *Phonetic correlates of stress in Urdu*. Ph.D. dissertation, Northwestern University.

Hyman, L. (2014). Do all languages have word accent? In H. Van der Hulst (Ed.), *Word Stress: Theoretical and Typological Issues* (pp. 56-82). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139600408.004

Ivanov, V.B. & Silanteva, L.G. (2020). Pitch and duration as main factors of Wakhi stress. *European Journal of Natural History* 3, 115-119.

Jackendoff, Ray S. (1972). *Semantic interpretation in generative grammar.* Cambridge, MA: MIT Press.

Jabeen, Farhat. (2017). Position vs. prosody: Focus realization in Urdu/Hindi. Paper presented at *Phonetics and Phonology in Europe (PaPE)* Conference, Koln, Germany.

Jabeen, Farhat. (2019). Interpretation of LH intonation contour in Urdu/Hindi. In *Proceedings of International Congress of Phonetic Science*.

Jabeen, Farhat. (2022). Word order, intonation, and prosodic phrasing: Individual differences in the production and identification of narrow and wide focus in Urdu. *Languages*, 7, 103.

Jabeen, Farhat & Elisabeth Delais-Roussarie. (2019). Towards a phonological analysis of the rising contours in Urdu/Hindi. Paper presented at *Speech Prosody 2018,* Poznan, Poland.

Jabeen, Farhat & Elisabeth Delais-Roussarie. (2020). The accentual phrase in Urdu/Hindi: A prosodic unit at the interplay between rhythm and intonation. Paper presented at *International Conference on Speech Prosody 2020*, Tokyo, Japan.

Keane, Elinor. (2007). Distribution and alignment of f0 contours in Tamil. *Proceedings of the XVIth International Congress of Phonetic Sciences*. Saarbrücken.

Kelkar, Ashok R. (1968). *Studies in Hindi-Urdu, Volume 1: Introduction and Word Phonology*. Postgraduate and Research Institute, Deccan College.

Khan, Sameer ud Dowla. (2008). *Intonational phonology and focus prosody of Bengali*. PhD dissertation, University of California, Los Angeles.

Khan, Sameer ud Dowla. (2014). The intonational phonology of Bengladeshi Standard Bengali. In Sun-Ah Jun (ed.), *Prosodic Typology II: The phonology of intonation and phrasing*. Oxford: Oxford University Press, pp. 81-117.

Kiranpreet, Nara. (2016). Acoustic analysis of Punjabi stress and tone (Doabi dialect), *The Journal of the Acoustical Society of America 139*, 2215.

Kügler, Frank, & Sasha Calhoun. (2021). Prosodic Encoding of Information Structure: A typological perspective. In C. Gussenhoven & A. Chen (Eds.), *The Oxford Handbook of Language Prosody* (pp. 454–467). Oxford University Press.

Kuznetsova, Alexandra, Per B. Brockhoff, & Rune H. B. Christensen. lmerTest Package: Tests in Linear Mixed Effects Model. *Journal of Statistical Software 82(13),* p.1-26.

Ladd, D. Robert. (1980). *The structure of intonational meaning: evidence from English*. Indiana University Press, Cambridge.

Ladd, D. Robert. (2008). *Intonational phonology*, 2nd ed. Cambridge: Cambridge University Press. [First edition 1996]

Ladefoged, Peter. (2003). *Phonetic Data Analysis: an introduction to fieldwork and instrumental techniques*. Wiley-Blackwell Publishing.

Lahiri, Aditi & Jennifer Fitzpatrick-Cole. (1999). Emphatic clitics and focus intonation in Bengali. In *Phrasal Phonology*, René Kager and Wim Zonnevel (eds.). Dordrecht: Foris Publications, p. 119-144.

Lehiste, Ilse. (1972) The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America* 51(6), pp. 2018-2024.

Lenth, Russell V. (2024). *emmeans: Estimated marginal means, aka least-squares means* (R package version 1.8.9). https://cran.r-project.org/package=emmeans.

Lunden, Anya. (2013). Reanalyzing final consonant extrametricality. *Journal of Comparative German Linguistics* 16, 1-31.

Lunden, A., Campbell, J., Hutchens, M., & Kalivoda, N. (2017). Vowel-length contrasts and phonetic cues to stress: An investigation of their relation. *Phonology,* 34(3), 565-580. doi:10.1017/S0952675717000288

Moore, Brian C. J. (2013). *Introduction to the Psychology of Hearing* (6th ed.). Brill.

Monahan, Phillip J. & William J. Idsardi. (2010). Auditory sensitivity to formant ratios: Toward an account of vowel normalization. *Language and cognitive processes,* 25(6), 808-839.

Mumtaz, B., Bögel, T., & Butt, M. (2020). Lexical Stress in Urdu. *Proceedings of Interspeech 2020*, p. 1888-1892.

Nair, R., A. Abbi, R. Gupta, and A. Kidwai. (2001). Acoustic correlates of lexical stress in Hindi. In *Linguistic Structure and Language Dynamics in South Asia—papers from the proceedings of SALA XVIII Roundtable*.

Nakai, Satsuki, Sari Kunnari, Alice Turk, Kari Suomi, and Riikka Ylitalo. (2009). Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics*, 37(1), 29-45.

Nakai, Satsuki, Alice E. Turk, Kari Suomi, Sonia Granlund, Riikka Ylitalo, and Sari Kunnari. (2012). Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics*, 40, 796-807.

Ohala, M. (1986). A search for the phonetic correlates of Hindi stress. *South Asian languages: structure, convergence, and diglossia*, ed. by Bh. Krishnamurti, C. Masica, and A. Sinha, 81-92.

Ohala, M. (1991). Phonological areal features of some Indo-Aryan languages. *Language Sciences*, 13(2), 107-124.

Ortega-Llebaria Marta & Pilar Prieto. (2011). Acoustic correlates of stress in central Catalan and Castilian Spanish. *Language & Speech, 54(1)*, 73-97.

Paramore, Jonathan Charles. 2021. *Mankiyali Phonology: Description and Analysis*. Master's thesis, University of North Texas. UNT Digital Library. https://digital.library.unt.edu

R Core Team. 2016. R: *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/. [Software]

Remijsen, Bert & Vincent van Heuven. (2002). Stress, tone, and discourse prominence in the Curacao dialect of Papiamentu. *Phonology* 22, 205-235.

Rooth, Mats. (1992). A theory of focus interpretation. *Natural Language Semantics* 1: 75-116.

Sbranna, S., Ventura, C., Albert, A., & Grice, M. (2023). Prosodic marking of information status in Italian. *Journal of Phonetics*, 97, 101-212. https://doi.org/10.1016/j.wocn.2023.101212

Shih, Shu-Hao. (2018). On the existence of sonority-driven stress in Gujarati. *Phonology*, 35(2), 327-364. doi:10.1017/S0952675718000064

Sluijter, A. M., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America*, *100*(4 Pt 1), 2471–2485. https://doi.org/10.1121/1.417955

Stevens, Kenneth N. (2000). Auditory processing of speechlike sounds. In Kenneth N Stevens, *Acoustic Phonetics* (ch.6).

Turk, Alice, Satsuki Nakai, & Mariko Sugahara. (2006). Acoustic Segment Durations in Prosodic Research: A Practical Guide. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schließer (Eds.), *Methods in Empirical Prosody Research* (pp. 1-28). Berlin, Boston: De Gruyter.

Van der Hulst, H. (2014). The study of word accent and stress: Past, present, and future. In H. Van der Hulst (Ed.), *Word Stress: Theoretical and Typological Issues* (pp. 3-55). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139600408.003

Van Heuven, V. (2018). Acoustic Correlates and Perceptual Cues of Word and Sentence Stress: Towards a Cross-Linguistic Perspective. In R. Goedemans, J. Heinz, & H. Van der Hulst (Eds.), *The Study of Word Stress and Accent: Theories, Methods and Data* (pp. 15-59). Cambridge: Cambridge University Press. doi:10.1017/9781316683101.002

Van Heuven, Vincent J. & Alice Turk. (2021). Phonetic correlates of word and sentence stress. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford Handbook of Language Prosody* (pp. 150-165). Oxford Unversity Press. DOI: 10.1093/oxfordhb/9780198832232.013.8

Vogel, Irene, Angeliki Athanasopoulou, & Nadya Pincus. (2016). Prominence, contrast, and the Functional Load Hypothesis: An acoustic investigation. In: Heinz J, Goedemans R, Van der Hulst H (eds), *Dimensions of Phonological Stress* (p.123-167). Cambridge: Cambridge University Press.

Waals, Juliette A. J. S. (1999). An experimental view of the Dutch syllable. *Netherlands Graduate School of Linguistics (Landelijke Onderzoekschool Taalwetenschap).*

Yakup, M., & Sereno, J. (2016). Acoustic correlates of lexical stress in Uyghur. *Journal of the International Phonetic Association*, 46(1), 61-77. doi:10.1017/S0025100315000183