



Contrastive vowel length and segment duration in production and perception in Djambarrpuyŋu

Kathleen Jepson, School of Languages and Cultures, University of Queensland, Brisbane, Australia; Institute for Phonetics and Speech Processing, Ludwig Maximilian University of Munich, Germany, kjepson@uq.edu.au

This paper deals with segment duration and phonological length in production and perception in the Australian language Djambarrpuyŋu. It provides an acoustic phonetic account of vowel and consonant duration to ascertain whether the phonetic evidence supports the analysis of contrastive vowel length in production. From a perception perspective, how Djambarrpuyŋu listeners use segment duration in categorising a pair of words that are proposed to be phonologically distinct by vowel length is tested. In the production study, phonemically long vowels were found to be approximately twice as long as short vowels, and consonants following short vowels were somewhat longer than when following long vowels, though the patterns were affected by syllable structure. In perception, listeners relied on vowel duration to distinguish between words that differed phonologically by vowel length, though consonant duration did influence the perceptual crossover point between words. The results are considered to address the questions of what evidence supports the vowel length contrast in Djambarrpuyŋu, and identifying the underlying mechanisms that result in the observed segment duration patterns.



1. Introduction

Djambarrpuyŋu,¹ an Australian Indigenous language, has been proposed to have contrastive vowel length (Wilkinson, 2012). The term contrastive length is used here to mean a segment duration difference that is determined by membership in a phonological category. In Djambarrpuyŋu, the vowel length contrast is restricted to the first syllable of the word, a position which is also the putative location of fixed word stress (Wilkinson, 2012). In closely related languages, both the stressed vowel and the following consonant's duration are reported to vary (e.g., Heath, 1980; Morphy, 1983; Waters, 1979, 1989; Wood, 1978). While the variation in vowel and consonant duration is attributed to different sources in the related languages' analyses (discussed in Section 2.3), the general pattern is that they vary inversely: Vowels of shorter duration are followed by consonants of longer duration, and vowels of longer duration are followed by consonants of shorter duration.

Relationships between vowel duration and consonant duration are observed in a diverse group of languages, and the observed durational patterns are also accounted for by different mechanisms. In Thai for example, vowels are analysed as having a phonemic length contrast, with consonant duration predictable from the length of the preceding vowel (Abramson, 1960). In Italian, on the other hand, consonant length is analysed as being phonemic, making the preceding (stressed) vowel's duration predictable (Hajek et al., 2007). This paper explores the acoustic realisation and perception of the purported vowel length contrast in Djambarrpuyŋu in terms of acoustic features of the vowel and the following consonant.

The Djambarrpuyŋu language and its speakers are introduced in Section 1.1. In Section 2, a cross-linguistic overview of vowel length including acoustic, phonological and perception analyses is presented. The subsections focus on studies on the relationship or complementarity in duration between neighbouring segments. Section 3 presents the research questions which are addressed in the production and perception studies in Sections 4 and 5, respectively. These sections each include the motivations, methods and results of the studies. Conclusions that can be drawn from the studies are presented and discussed in Section 6, considering what mechanisms might underlie the observed length contrast, implications for what we know about vowel length and segment duration relationships, and how the findings fit within the typological understanding of vowel length contrasts.

1.1. Djambarrpuyŋu

Djambarrpuyŋu is spoken by approximately 4,000 people in northeast Arnhem Land (Australian Bureau of Statistics, 2021a). It is a member of the Yolŋu subgroup of the Pama-Nyungan language family (Bown, 2023), and is one of a small number of Australian Indigenous languages that

¹ ISO 639-3: djr; Glottocode: djam1256.

are used in day-to-day life and learned by children at home (see Koch & Nordlinger, 2014 for discussion). Today, speakers live in a number of communities including Galiwin'ku, Gapuwiyak, Milingimbi, and Ramingining (Wilkinson, 2012; Yunupingu, 1996). Djambarrpuyŋu is now a main community language of these communities, which are not all traditional Djambarrpuyŋu clan and language affiliation areas. Speakers are often multilingual and multidialectal with knowledge of related varieties, other Australian languages, Australian English, and Kriol.

The vowel system proposed by Wilkinson (2012) in her grammatical analysis of Djambarrpuyŋu is symmetrical, consisting of six vowels distinguished by height, backness, and—in the first (stressed) syllable of the word—length. To reflect the tendency for vowels in Australian Indigenous languages to have less extreme cardinal-like articulation and more variation within each quality category, the following symbols are used here: /ɪ, ɪ:, ɐ, ɐ:, ʊ, ʊ:/ (Busby, 1980; Fletcher & Butcher, 2014). The length contrast is exemplified in the following near minimal pairs: *weti* /wɛ:ɪ/ “wallaby,” *witiŋ* /wɪ:ɪŋ/ “olive python”; *gorrum* /kɔ:rum/ “to be high,” *gurrum* /kɔrɔm/ “soft”; and *gäna* /kə:nə/ “alone,” *gana* /kənə/ “enough” (Wilkinson, 2012). The restriction of contrastive vowel length to the initial syllable (i.e., the stressed syllable) is commonly reported for Pama-Nyungan languages that have a length contrast (Baker, 2014; Dixon, 2002; Fletcher & Butcher, 2014).

A preliminary acoustic study of Djambarrpuyŋu vowel duration from words in isolation confirmed that duration significantly differed between phonemically short and long vowels in the first syllable of CV.CV words with short vowels having a mean duration of 126 ms and long vowels a duration of 211 ms (Jepson & Stoakes, 2015). Wilkinson (2012) notes, however, that the phonetic contrast can be lost in fast connected speech, which results in the collapsing of minimal pairs. She additionally noted that in reduplicated forms, reduplicants do not retain the long vowel, for example, *yolŋu'-yulŋu* /jɔ:lŋɔʔjɔlŋɔ/ “people” (Wilkinson, 2012; see also Heath, 1980 on Ritharrŋu).²

There are 25 consonants in Djambarrpuyŋu: Six places of articulation for stops, for which there are fortis and lenis series (with the former represented by length in the following), /p:, p, t:, t, t:, t, c:, c, k:, k/, and nasals /m, n, ŋ, j, ŋ/, as well as one trill /r/, three central approximants /w, ɹ, j/, two lateral approximants /l, ɭ/ and glottal stop /ʔ/. The fortis/lenis contrast is yet to be acoustically analysed. Note that the transcription convention adopted here of using the length diacritic for fortis stops does not imply ambisyllabic association, rather it is used to capture that these stops are hypothesised to differ primarily in closure duration, rather than voicing (Butcher, 2004; Round, 2023). Round summarises that languages of Arnhem Land (including Yolŋu languages) with two series of stops usually make the distinction by closure

² As a reviewer pointed out, because of this process of reduplication, vowel length could be interpreted as being a dynamic process, rather than a lexical specification.

duration, in contrast to languages of Cape York, which primarily use voicing. Further, this convention was adopted in previous phonetic research on the language (e.g., Jepson et al., 2021). Phonologically, stops are limited in where they are considered contrastive in Djambarrpuyŋu—only in morpheme internal intercontinuant position. The contrast between the fortis and lenis stop series has a very low functional load, though near minimal pairs do exist (Wilkinson, 2012). For example: *bäba* /pɛ:pɛ/ “gall,” *bäpa* /pɛ:pɛ/ “father”; *gadharra* /kɛɽere/ “coral,” *watharr* /wɛɽɛr/ “white”; and *räga* /ɾɛ:kɛ/ “white berry bush,” *räkay* /ɾɛ:kɛj/ “Eleocharis dulcis.” These specific minimal pairs were not a focus of data collection and the differences between the fortis and lenis stop series are not investigated acoustically in the current paper.

Djambarrpuyŋu allows for syllables that are maximally CV(C)(C)(?); all syllables (and words) are consonant initial with the exception of a very small number of words, predominantly in the “baby talk” register (Wilkinson, 2012, p. 45), for example, *anyany* /ɛɽɛɽ/ “cute.”

2. Background

2.1. Contrastive vowel length

Cross-linguistically, contrastive vowel length is most commonly reported with two levels of contrast; generally, “long” vowels are distinguished from “short” (Odden, 2011). A pattern of three levels is rare (Remijsen, 2014 on Dinka). Phonetically, long and short categories are correlated with larger or smaller duration values respectively, with the exact values for each category being language specific (Odden, 2011, p. 465). The raw duration values can vary considerably. For example, in Thai utterance-medial words, short vowels are, on average, 160 ms, and long vowels are 320 ms (Roengpitya, 2001, p. 30), while in Hungarian accented utterance-medial words, short vowels are 54 ms and long vowels are 94 ms (White & Mády, 2008). The ratio or proportional difference between vowel length categories can also show substantial variation. For example, in Marshallese the ratio is approximately 1:2.8 (Choi, 1992, p. 64), but in Czech, the ratio is 1:1.63, averaged across five vowel pairs (Podlipský et al., 2009, p. 134).

Vowel duration is well known to also be conditioned by several other factors, such as metrical strength, accentuation, voicing and manner of the following consonant, vowel quality, syllable structure, number of syllables in the target word, whether uttered in isolation or in an utterance, and the vowel’s position within the target word (e.g., Lehiste, 1970; Port & Dalby, 1982; Turk, 2012). These factors can affect the duration of long and short vowels differently, and consequently, the proportional differences between the categories may also change (e.g., de Jong & Zawaydeh, 2002, p. 62, their Figure 1 on Ammani-Jordanian Arabic and effects of stress; White & Mády, 2008 on Hungarian and syllables per word). Of particular interest here is syllable structure; position within the utterance is also considered.

Cross-linguistically, vowels in closed syllables are found to be phonetically shorter than in open syllables (Maddieson, 1985). This effect has been observed in languages that have contrastive

vowel length such as Dutch (Jongman, 1998; Rietveld & Frauenfelder, 1987), Arabic (Broselow et al., 1997; de Jong & Zawaydeh, 1999; Khattab, 2007), and Malayalam (Broselow et al., 1997), as well as those which do not such as Italian (Farnetani & Kori, 1986; Hajek et al., 2007). The position of the word within the utterance is also found to have an effect on segment duration, with vowels in words uttered in utterance-final position having longer duration than comparable vowels in words in other positions (White & Mády, 2008 on Hungarian). For Hungarian, final-lengthening interacts with vowel length category such that final long vowels show a greater effect of position than short vowels, being lengthened approximately 40 ms, while short vowels only lengthened an estimated 21 ms (White & Mády, 2008). See also Paschen et al. (2022) for a discussion of the interaction between vowel length and finality.

Beyond duration, contrastive vowel length is also correlated with spectral differences for monophthongs. Typically, membership in the long category corresponds to being more peripheral in the vowel space, whereas being short corresponds to a more centralised acoustic realisation. This is based on a more general assumption that longer duration of a vowel corresponds with more extreme articulation because articulators are able to move to more peripheral positions, while in shorter periods of time the movement that is possible is restricted (Cho, 2004; Lehiste, 1970, p. 31; Lindblom, 1963, p. 1780). However, vowel length is not always found to affect formant structure (cf. Behne et al., 1996, on Norwegian). And, the prediction does not hold for short versus long vowels in some Australian languages (Fletcher & Butcher, 2014). Of relevance here, however, is Graetzer's (2012) analysis of vowels in Gupapuyŋu, a Yolŋu variety closely related to Djambarrpuyŋu, in which she found that long vowel centroids tended to be more peripheral in the vowel space in at least one dimension: the pair /ʊ/ and /ʊ:/ differed in F2, while the pairs /ɪ/ and /ɪ:/ and /e/ and /e:/ differed in F1 (Graetzer, 2012, pp. 187–188). It is possible that there are also minimal vowel formant differences in Djambarrpuyŋu due to articulatory mechanisms that result in differences which are not sufficient to posit vowels of different qualities.

2.2. Phonetics and phonology of vowel-consonant duration relationships

Not infrequently, a complementary duration relationship between a vowel and the following consonant is described, most often when the vowel is stressed. The duration relationship is inverse, with consonant duration being longer when the vowel is shorter, and shorter when the vowel is longer. This is reported to occur in, for example, Arabic (Khattab, 2007), the Shetland dialect of English (Sundkvist & Gao, 2015; van Leyden, 2004), German dialects (Kleber, 2020), Italian dialects (Baroni & Vanelli, 2000), Norwegian (Behne et al., 1996), Swedish dialects (Elert, 1965; Schaeffler, 2005), Thai (Abramson, 1960), Washo (Yu, 2008), and Guienagati and Ozolotepec Zapotec (Benn, 2021; Leander, 2008). There are a range of phonological, phonetic, and prosodic motivations associated with this type of durational relationship; sometimes the pattern is attributed to the vowel, sometimes to the consonant, and sometimes to another syllable weight or duration requirement.

For Washo, “[p]ost-tonic consonant gemination after short stressed vowels may be viewed as a compensatory response to the monomoraicity of short vowels” (Yu, 2008, p. 517). This interpretation of preserving moraic weight is discussed within phonological theory (Gordon, 2016; Hayes, 1989) and is observed in some Australian languages (Baker, 2014; see also Harvey & Borowsky, 1999 on Warray). For Ngalakgan, an Australian language of the Gunwinyguan family, Baker (2008, pp. 75–82) describes a bimoraic minimum for monosyllabic words which can be achieved through a closed syllable containing a short vowel, or the phonetic lengthening of vowels if the syllable is open. In the moraic theory view of compensatory lengthening, the lengthening of a vowel conditioned by the loss of a coda consonant would not occur if the language’s codas are non-moraic for other purposes, such as stress assignment or minimal word requirements (Gordon, 2016, p. 160). Most Australian languages are described to be weight insensitive regarding stress assignment (Baker, 2014; Fletcher & Butcher, 2014; Jepson & Ennever, 2023), irrespective of whether consonants are moraic or not. There is a preference for minimally bimoraic monosyllabic words in Djambarrpuyŋu, despite a small number of exceptions (discussed in Section 2.3). The possibility of a phonetic bimoraic minimum requirement for stressed syllables is discussed in sections 4.3 and 6.

For the Shetland dialect of English, van Leyden (2004) found that consonants were lengthened in a compensatory manner with the shortening of vowel duration. It was reported that a 100 ms shortening in vowel duration was reflected by a 49 ms lengthening of the following consonant. Van Leyden (2004, p. 39) concluded that, while there is an inverse relationship between vowel and consonant duration in the Shetland dialect, it is strongest for some sets of words considered to constitute pairs, particularly local dialect words.

In Thai (Abramson, 1960; Abramson & Ren, 1990), semivowels and nasals in syllables containing short vowels have longer duration than those in syllables containing long vowels, but again, durational differences are not as great as they are for vowels and do not fully compensate for the difference in duration between long and short vowels: /n/ in /sɯn/ was 150 ms while the /n/ in /sɯ:n/ was 110 ms long (Abramson, 1960, p. 132), whereas short and long vowels were on average 85 ms and 165 ms, respectively (Abramson & Ren, 1990). In languages with geminate consonants, such as Makasar for example, singleton /n/ is found to be 100 ms while geminate /n/ is 205 ms on average (Tabain & Jukes, 2016); that is, it is on par with durational differences observed for contrastive vowel length.

When longer vowels are only observed with shorter consonants and shorter vowels with longer consonants, difficulty can arise in determining whether the vowel or the consonant has contrastive length. In Thai and the Shetland dialect of English there is a smaller difference in duration between consonants, leading to vowel length being analysed as the primary determiner of vowel duration and, consequently, determining the duration of the following consonant. However, this is not always so clear cut (e.g., in Swedish Schaeffler, 2005). Central Bavarian

presents an interesting case showing the difficulty of determining the motives for varying duration or neighbouring segments (Kleber, 2020). In Central Bavarian, tense (long) vowels occur before lenis stops, and lax (short) vowels occur before fortis stops. Prior to Kleber (2020), there were two proposed phonological solutions to this: Firstly, that the fortis/lenis contrast is phonological while vowel duration is predictable; secondly, that there is a contrast between the complementary length of VC sequences and that prosodic quantity specifies either the vowel or stop to be long, and the other is then predictably short. Kleber concluded that, because lengthening was observed for sonorants (which do not otherwise have a length contrast) and not exclusively for the stops, that vowel length is phonemic in Central Bavarian, and the consonant duration contributes to marking the vowel length contrast. On Washo, Yu (2008) acknowledges that his analysis rests on an assumption of vowel length being contrastive and that the issue is still under debate.

2.3. Segment duration in Yolŋu languages

As mentioned in Section 1, an inverse duration relationship is reported for other Yolŋu languages, with longer vowels being followed by shorter consonants, and vice versa. This type of durational relationship is proposed to be a phonetic feature of most Australian languages that have contrastive vowel length (Dixon, 2002, p. 591). When contrastive vowel length, which had historically been more commonly observed around Australia (Dixon, 2002, p. 639), is lost, consonant duration is hypothesised to be phonologised as being contrastive (Dixon, 2002, pp. 591, 595). Though, for a number of Yolŋu languages, two series of stops are attested, as well as contrastive vowel length (the historical conditions that could account for this are discussed by Dixon, 2002, pp. 604, 611). It appears that this has contributed in part to the diverse analyses for the similar phonetic outcomes in Yolŋu languages, the details for which are discussed here.

Of the varieties and languages discussed below, Djambarrpuyŋu is most closely related to Djapu—both are within the same language group of Dhuwal—while it is more distantly related to Ritharrŋu, Djinang, and Gälpu which are in different language groups (Bower, 2023; Wilkinson, 2012). In some descriptions, sonorants are the focus of lengthening processes, while in others, obstruents are primarily discussed.

Morphy (1983) reports for Djapu that in disyllabic words where the first/stressed syllable is open, the duration of the following consonant varies according to the length of the stressed vowel. The relationship is predictable and complementary, with short vowels being followed by longer consonants, and long vowels followed by shorter consonants. She illustrates this with the pair *wajŋa* /wɛ:ŋɛ/ “to talk/speak” and *wäŋa* /wɛ:ŋɛ/ “home/place” (Morphy, 1983, p. 25). It is argued that the lengthening or “gemination” of the consonant following the long vowel occurs to ensure that the stressed syllable is heavy. In Morphy’s analysis, a syllable is heavy as the result of either a long vowel in the stressed syllable, due to a coda consonant, or, when the following consonant is the onset in the next syllable, the lengthening of that consonant which

is consequentially analysed as being ambisyllabic (to fulfil the requirement that all syllables are consonant initial). So, in the above examples, *wəŋa* /wəŋə/ [wəŋ:ə] is syllabified as [wəŋ.ŋə], while *wəŋa* /wə:ŋə/ [wə:ŋə] is syllabified as [wə:ŋə]. Morphy notes that the durational difference could be analysed as originating from either the vowel or the consonant, but argues that to motivate the lengthening of the consonant is simpler. She suggests there would be no way to motivate lengthening of vowels in words like *yolŋu* /jɔ:lŋɔ/ “person,” which would be a heavy syllable by nature of being closed.

Heath (1980) posits a length contrast for vowels in stressed syllables in Ritharrŋu, though notes that there is a tendency for long vowels to be phonetically shortened in words of three or more syllables as well as in reduplicants (p. 12), as Wilkinson (2012) has also reported for Djambarrpuyŋu (Section 1.1). Lengthened consonants, specifically liquids and nasals, are reported to occur intervocalically when following short vowels. Heath (1980, p. 12) had originally posited a length contrast for consonants in addition to vowels. He provides a discussion of his changing analysis, citing the occurrence of long vowels before consonant clusters, and gives examples such as the word *yalu* “nest” which was originally transcribed as [jəl:ɔ] (orig. [yallu] in text), explaining consonant duration varies phonetically, without phonological implications. Ritharrŋu is analysed as having fortis and lenis stop series though the effect of vowel length on those consonants is not discussed.

In Djinang (Waters, 1979, 1989), a similar phonetic pattern is described, though the phonological analysis and underlying mechanism differs to Djapu and Ritharrŋu. Waters proposes that vowel length is not contrastive. Rather, vowels optionally lengthen in stressed syllables, especially when the syllable is open (see discussion in Ryan, 2019, p. 36), and that consonants, specifically voiceless stops (i.e., fortis stops in other analyses), geminate following stressed syllables when the stressed vowel is phonetically short.

Wood (1978), on the other hand, proposes for Gälpu (also Gaalpu) that vowel length is determined by a “syllable feature” of length—short or long. Long vowels occur in syllables with the long feature, while short vowels occur in short syllables. Stops lengthen in intercontinuant environments, and lengthening is especially “pronounced” when the vowel is short and the consonant is the onset to the second syllable (i.e., is intervocalic) (Wood, 1978, p. 64). He terms the analysis “a prosodic solution” in contrast to a segmental solution in which vowel length would be analysed as contrastive (i.e., as in Morphy, 1983). These two options are both considered valid analyses by Wood, but he argues that the prosodic solution is better supported as it reflects the analysis of the glottal stop as syllable feature (first proposed by Bernhard Schebeck in a paper in 1974, cited in Wood, 1978), possibly historically associated with stress (Wood, 1978, p. 98).

For Djambarrpuyŋu, Wilkinson (2012) posits contrastive vowel length, but does not mention a durational relationship with the following consonant. Jepson and Stoakes (2015) examined intervocalic consonants in CV.CV Djambarrpuyŋu words spoken in isolation, and found that their

duration differed significantly depending on whether they occurred following a phonemically (and durationally) short or long vowel, with consonants following short vowels being 236 ms in duration and consonants following long vowels being 191 ms. This study, also discussed in Section 1.1, supported the phonetic facts of what has been reported for Yolŋu languages: that the vowel in the first syllable of words varies in duration, that the duration of the following consonant varies too, and these are often in a complementary relationship. However, as only CV.CV words were examined, the analysis of segments in words with different syllable structures is necessary to understand the segment duration patterns of Djambarrpuyŋu.

All of these analyses rely to a degree on stress, and a requirement for stressed syllables to be a certain weight (i.e., heavy), or privileged to some degree. That is, contrastive length occurring only in stressed syllables could be interpreted as showing positional prominence, with more segmental contrasts maintained in stressed syllables than unstressed syllables in these languages (Hyman, 2014). On the other hand, as mentioned above (Section 2.2), it is generally assumed that stress is not quantity sensitive in most Australian languages, with most systems having fixed stress either to a word or root morpheme boundary (Baker, 2014; Fletcher & Butcher, 2014; Jepson & Ennever, 2023). The preference for heavy stressed syllables could suggest that some version of a Weight-to-Stress principle, which specifies heavy syllables are stressed, could be at play (Gordon, 2011; Prince, 1990). It is noted that this principle is violated in so much as unstressed syllables can also be heavy (e.g., /^lrɛj.məl/ “temple, side of head”; /^lŋu.mən/ “smell, give off smell”), but it could motivate a preference for phonetically heavy stressed syllables. Djambarrpuyŋu allows monosyllabic words of the structure CV, for example /kɛ/ “and,” or the present and today future form of the imperfective auxiliary (Wilkinson, 2012, p. 44). However, there is a preference for minimally bimoraic roots in Djambarrpuyŋu, and from descriptions of other languages with prosodic minimality constraints, function words are observed to be prosodically deficient such as Ixtayutla Mixtec (Penner, 2019). So, it may be that minimality constraints related to word structure apply in Djambarrpuyŋu, and a similar minimum may also apply specifically to stressed syllables in Djambarrpuyŋu.

Considering other motives for lengthening, Solé and Ohala (2010, pp. 608, 611, *inter alia*) suggest that language-specific secondary properties that covary with a contrastive dimension could serve to enhance the contrast (see also Schertz & Clare, 2020). That could be the case for compensatory lengthening of the type proposed for Yolŋu and other Australian languages, to enhance the contrast between short and long vowels. This too is in conflict with the prevailing understanding of the phonetics of Australian languages. The large consonant inventories with many places of articulation that are found in most Australian languages, it is argued, require a maximisation of acoustic cues to consonant place at the expense of vowel contrasts (Butcher, 2006). This tendency for preserving or enhancing the place information of consonants over enhancing vowels has been termed the “Place of Articulation Imperative” by Butcher (2006)

and is phonetically realised as the “lengthening and strengthening” of, in particular, post-tonic consonants in some Australian languages, serving dual purposes of maintaining and possibly enhancing paradigmatic consonant contrasts on the one hand and signalling prosodic prominence on the other (Butcher, 2006; Fletcher & Butcher, 2014). This preference is not observed across the board in Djambarrpuyŋu, but is found in words of two syllables (Jepson et al., 2021).

In Jepson et al. (2021), the duration of intervocalic consonants was measured in words of two to six syllables in length, and post-tonic consonants (i.e., following the stressed vowel) in disyllabic words were found to be significantly longer than post-tonic consonants in words with more syllables and non-post-tonic consonants. The words analysed in Jepson et al. (2021) contained only short stressed vowels. Therefore, it is not entirely clear if the duration of the consonant was conditioned by post-tonic lengthening (a focus of Jepson et al., 2021) or due to following a short vowel, as explored in this paper (for more on post-tonic consonant lengthening in other Australian languages see Butcher & Harrington, 2003a; Butcher & Harrington, 2003b; Fletcher et al., 2010; Fletcher et al., 2015). Disyllabic words are mentioned explicitly in Morphy’s (1983) analysis of vowel length in Djapu, and because they partake in other segment duration processes in Djambarrpuyŋu, it appears that two syllable words could present an interesting and distinct word structure in Yolŋu languages, which may show processes that are otherwise not observed in words of other structures.

2.4. Perception of segment duration and vowel-consonant duration relationships

Vowel duration is the primary cue used by listeners cross-linguistically when categorising vowels as “short” or “long” in perception studies (e.g., German: Lehnert-LeHouillier, 2007, 2010; Tomaschek et al., 2011; Thai: Abramson & Ren, 1990; Lehnert-LeHouillier, 2007, 2010; Roengpitya, 2001; Japanese: Arai et al., 1999; Behne et al., 1999; Kinoshita et al., 2002; Lehnert-LeHouillier, 2007, 2010). Listeners of languages with a phonemic vowel length contrast use their knowledge of the sounds in their language to categorise stimuli into existing categories in categorisation experiments (Holt & Lotto, 2010; Liberman et al., 1957).

However, vowel duration can be one of many cues that lead to the categorisation of the phone to a length category. A cue, albeit the primary one, is neither independent of the influence of other cues on the perception of a phone, nor solely responsible for it. (Repp, 1982). Further, in perception studies (and in day-to-day speech), there are also stimuli with duration values that fall between the typical durational range of “short” and “long” categories found in the production data for a language. Stimuli within this range are ambiguous as they do not provide durational cues that are sufficient for listeners to categorise the phone. In vowel length perception, stimuli containing vowels in the language-specific ambiguous region can be more difficult for listeners

to process, which has been reflected in reaction times (Eerola et al., 2012, for Finnish listeners). Further, vowels that are between duration categories are subject to the influence of secondary cues.

Covarying of secondary features may be used as reliable indicators of the primary contrast (Solé & Ohala, 2010, p. 611), and when the primary acoustic cue is ambiguous, secondary cues can influence listeners' perception (Hillenbrand & Clark, 2000, pp. 3020–3021; Lehnert-LeHouillier, 2010; Whalen et al., 1993). Secondary features are not ignored by listeners when the primary cue is not ambiguous; these cues are used by listeners even when phonologically redundant (Whalen et al., 1993). This type of effect can alter the location of the crossover point between categories. In vowel length perception, spectral information and f_0 are most commonly investigated as secondary cues. These cues are used differently across languages in perception, as they are in production (Lehnert-LeHouillier, 2010). Language-specific formant quality changes can also result in varying crossover points in the perception of length contrasts (Abramson & Ren, 1990). This is to say, altering of one cue can be compensated for by altering another cue, thus maintaining the original phonetic percept in a cue trading relation (see Raphael, 2005; Repp, 1982).

Durational information beyond the vocalic element has also been shown to be a cue to vowel length in perception. This can operate between the vowel and the following consonant. For example, in Thai, longer post-vocalic consonant duration was found to condition a higher proportion of short vowel responses than post-vocalic consonants with shorter durations (Roengpitya, 2001). Similarly, van Dommelen (1999) found for Norwegian that the boundary between the categories of long and short for vowels was influenced by the duration of the following consonant. Overall, there was a shift in where the boundary occurred, such that a longer consonant resulted in a later boundary between the vowel categories than a shorter consonant, suggesting longer consonants resulted in perceptual shortening of the vowel.

3. The present studies

This paper presents two studies of vowel length in Djambarrpuyŋu, one from a production perspective and one from a perception perspective, both developed from the author's PhD thesis (Jepson, 2019a) and related conference proceedings papers (Jepson, 2019b; Jepson & Stoakes, 2015). It continues a line of research into segment duration in Djambarrpuyŋu (Jepson et al., 2021). Specifically, the present production study builds on Jepson and Stoakes (2015) by investigating other acoustic measures such as formants, as well as by incorporating syllable structure into the analysis, which is cross-linguistically found to affect segment duration and the duration ratio between short and long vowels. The focus on consonants here is also complementary to Jepson et al. (2021) in so much as the issue of vowel length was avoided in that study, due in part to the hypothesised durational relationship.

Together, the current studies provide an opportunity to further discuss the underlying processes for the phonetic patterns of segment duration observed in Djambarrpuyŋu which may be relevant in other Australian languages. They also advance the cross-linguistic understanding of vowel length in production and perception by the inclusion of Djambarrpuyŋu, a language typologically dissimilar in many respects from the languages reported on to date.

This paper sets out to explore the proposed vowel length contrast in Djambarrpuyŋu, taking into account the complexity that comes with durational relationships. From the previous phonological analyses of Djambarrpuyŋu (Wilkinson, 2012) and other Yolŋu languages (Heath, 1980; Morphy, 1983), it is assumed here that vowel length is contrastive, and this paper investigates other cues to the length contrast, in addition to vowel duration. Or, considered another way, what effects vowel length has on the production of vowels and neighbouring consonants. While intervocalic consonants following stressed vowels were reported to vary inversely with the vowel's length in Djapu and Ritharrŋu (Heath, 1980; Morphy, 1983), Wilkinson (2012) does not mention such a pattern for Djambarrpuyŋu; however, the omission is not considered evidence as to whether the durational pattern occurs in the language or not. This paper therefore deals with a similar issue to Yu (2008), van Leyden (2004), and others, in considering the sources of the durational alternation.

This paper considers three main questions:

1. How does vowel length affect vowel duration, vowel formants, and duration of the following consonant in Djambarrpuyŋu?
2. What durational cues do listeners use to categorise words that vary phonemically by vowel length?
3. Do the production and perception results support vowel length being contrastive in Djambarrpuyŋu?

Hypotheses addressing these questions are provided in the introductions to the two studies in sections 4 and 5.

These studies also aim to contribute to topics that are beyond the scope of this paper. For example, enhancing our understanding of the acoustic characteristics of vowels in Djambarrpuyŋu is relevant for undertaking an analysis of word stress, which is a controversial topic in the research on Australian languages (see e.g., Tabain et al., 2014; Jepson & Ennever, 2023). The paper is also motivated by a desire to improve understanding of Indigenous languages' sound systems more generally; there are a handful of linguistic perception studies conducted with Indigenous listeners to date focused predominantly on consonants (e.g., Anderson, 1997; Bundgaard-Nielsen & Baker, 2015; Bundgaard-Nielsen et al., 2012; Stoakes et al., 2012; Wang et al., 2024).

4. Study 1: Vowel length production

In this study, all vowel qualities are examined, and word length and syllable structure are controlled for (discussed in Section 4.1.2). Other factors, such as stress, are mutually exclusive (vowels specified for contrastive length are necessarily stressed in Wilkinson’s 2012 analysis), or were not targeted in the experimental data (variation in accentual prominence—all words were accented). The possible complementarity between vowels in the first syllable of words and the following consonant is also considered in this study.

Four hypotheses are tested, based on the literature covered in previous sections:

- H1. Vowel duration will vary consistently with the phonological categorisation of short vs. long, with long vowels having longer duration than short vowels.
- H2. Vowel formants will be affected by vowel length, with long vowels being more peripheral in the vowel space than short vowels.
- H3. Intervocalic consonant duration will vary inversely with the phonological length categorisation of the preceding vowel.
- H4. Syllable structure will have an effect on the duration of both vowels and consonants. Vowels in open syllables will be longer than vowels in closed syllables. Intervocalic consonants’ duration will be affected by the preceding vowel’s length, while coda consonants’ duration will not.

4.1. Methods

This study is based on a dataset drawn from a larger database produced as part of the author’s doctoral research (in the process of being archived in PARADISEC). The project had ethics approval from the University of Melbourne Human Research Ethics Committee (HREC number: 1544581). Further details of data collection and the database can be found in Jepson (2019a) and Jepson et al. (2021).

4.1.1. Participants

Seven native Djambarrpuyŋu speakers were recorded in Milingimbi, a small island off the coast of northeast Arnhem Land, Northern Territory, Australia, with a population of ~1,100 (Australian Bureau of Statistics, 2021b). Four women and three men participated, aged between 32 and 68 years (mean 48.6) at the time of recording. Variation due to age is not investigated in this dataset. All participants identified as speaking Djambarrpuyŋu and had knowledge of a number of other Yolŋu languages and other non-Yolŋu Australian languages. All participants were familiar with Australian English. Speakers were representative of their community in

terms of age and linguistic repertoire, though may not identify themselves as belonging to the Djambarrpuyŋu clan specifically. Participants gave informed consent to participate in the study, have their recordings archived, and for the recordings to be analysed and results presented. They were paid for their time.

4.1.2. Materials

To control for effects of word length in this study, disyllabic words were selected for analysis from the larger corpus (discussed further in Section 4.1.4; see also Jepson et al., 2021). Items had one of four structures — C1V1.C2V2, C1V1.C2V2C3, C1V1C2.C3V2, C1V1C2.C3V2C4, where a full stop represents a syllable break. As noted in Section 1, vowel length is proposed to be contrastive only in the initial syllable of words in Djambarrpuyŋu; therefore, of interest in the present study are V1 and C2. The identity of C2 was a fortis stop (i.e., obstruent) or a nasal, trill, approximant, or lateral (i.e., sonorant). In total, 107 lexical items were included in the dataset.

The data, summarised in **Table 1**, are not balanced in terms of how many of each vowel phoneme occurred, nor how many of each phonotactic structure were included.

Vowel	Open (<i>n</i>)	Closed (<i>n</i>)
/ɪ/	86	26
/ɪː/	80	0
/e/	372	213
/eː/	205	35
/ʊ/	186	108
/ʊː/	96	21

Table 1: Count summary of data for vowel length analysis by syllable type.

4.1.3. Elicitation and recording procedure

The wordlist items were elicited in three frame sentences in which the target word was syntactically in utterance-initial, -medial, or -final position. The items in the wordlist were discussed with each speaker before the recording session. In the recording session, each item was presented verbally in English, Djambarrpuyŋu, or through an explanation in English. Prompting by the researcher was adopted for consistency across speakers because individuals had differing levels of comfort in undertaking literacy-dependent tasks. Speakers said each item in the three frame sentences one time.

Recordings were made using a Zoom H6 digital recorder and a Countryman headset microphone with a hypercardioid pattern directional capsule covered by a wind shield, and

were of 24-bit bit-depth and 48-kHz sample rate (16-bit bit-depth for analysis). Recording sessions were conducted inside a house with overhead fans and air-conditioning units turned off. Speakers were seated either at a wooden table facing parallel to a wall, or on a sofa. There are two exceptions: One male speaker was seated on a secluded veranda facing away from the house, another male speaker was seated in the shade outside and subsequently on a veranda.

4.1.4. Data processing, measurements, and analysis

Utterance segmentation and transcription were conducted in Praat (Boersma & Weenink, 2019) using a modified Djambarrpuyŋu orthography. Data were forced aligned on two separate occasions and so used two slightly differing methods. For the first set of data, the Munich Automatic Segmentation System (Schiel et al., 2011) implemented in R (R Core Team, 2023) was employed, using a modified SAMPA (language-independent) parameter definition that included acoustic models for retroflex stops, laterals, and nasals. For the second set of data, the web-based Munich Automatic Segmentation System (Kisler et al., 2017) was employed, using the language-independent model. All segmentation was manually corrected in Praat using waveform and spectral information. For stops, the occlusion and release phrases were measured together. Details of segmentation for all segments can be found in Jepson (2019a, pp. 83–85; see also Jepson et al., 2021).

A hierarchically associated Emu SDMS database was created for the full corpus using the emuR package in R (Jochim et al., 2023). The database was queried in R using the emuR suite of commands, to produce a dataset containing disyllabic words of the structures C1V1.C2V2, C1V1.C2V2C3, C1V1C2.C3V2, C1V1C2.C3V2C4, as discussed in Section 4.1.2. Durational values were extracted for V1 and C2. First and second formant values were extracted at the temporal midpoint of the vowel using the `get_trackdata()` function from the emuR package. This function drew upon extracted acoustic signal files that were saved in the database. Values for the first four formants were calculated using the formant estimate function `forest()` from the wrassp package (Winkelmann et al., 2023). For the `forest()` function, default settings were used, therefore nominal F1 was set to 500 Hz, and a window length of 30 ms. Formants were not manually corrected, nor were any values excluded in the analysis.

Linear Mixed-Effects Models were used to statistically model and test the data. Statistical analysis was performed in R using the `lmer()` function from the lme4 package (Bates et al., 2015). From ANOVA comparisons between models, Chi squared and *p*-values are reported in this paper. Differences of Least Square Means results are reported in the results section (4.2) (estimated difference, standard error), and post-hoc significance (*p*-) values for individual comparisons were obtained using the emmeans package (Lenth, 2023) and were bonferroni corrected. *P*-values < 0.05 are considered to be significant. Visualisations were created using the ggplot2 package in R

(Wickham, 2016). For formants, statistical analyses were performed on raw Hz values; Lobanov normalised (z -score) values were used for visualisation purposes (Lobanov, 1971).

The models reported were theoretically motivated to allow for the investigation of issues relevant to vowel length and durational relationships that are discussed in the literature. See individual results sections 4.2.1, 4.2.2, and 4.2.3 for details of the models.

4.2. Results

4.2.1. Vowel duration

Figure 1 shows the distribution of values for long and short vowels in the dataset. A clear pattern can be seen in the data: Vowels that are phonemically long (in white) had longer duration values than their phonemically short counterparts (in grey). This observation is reflected in the results of the statistical analysis. Duration of the target vowels was statistically tested in a mixed effects model that included vowel length category (two levels: long, short), vowel quality (three levels: open central /e/, close front /i/, close back /u/), syllable type (two levels: open, closed), utterance frame (three levels: frame initial, frame medial, and frame final), and C2 consonant category (i.e., “consonant category,” two levels: obstruent, sonorant) as fixed effects with an interaction between vowel length category and syllable type, and vowel length and utterance frame. Random effects included random intercepts for speaker and word, as well as by-speaker random slopes for the effect of vowel length category. **Table 2** presents a summary of selected post-hoc comparisons.

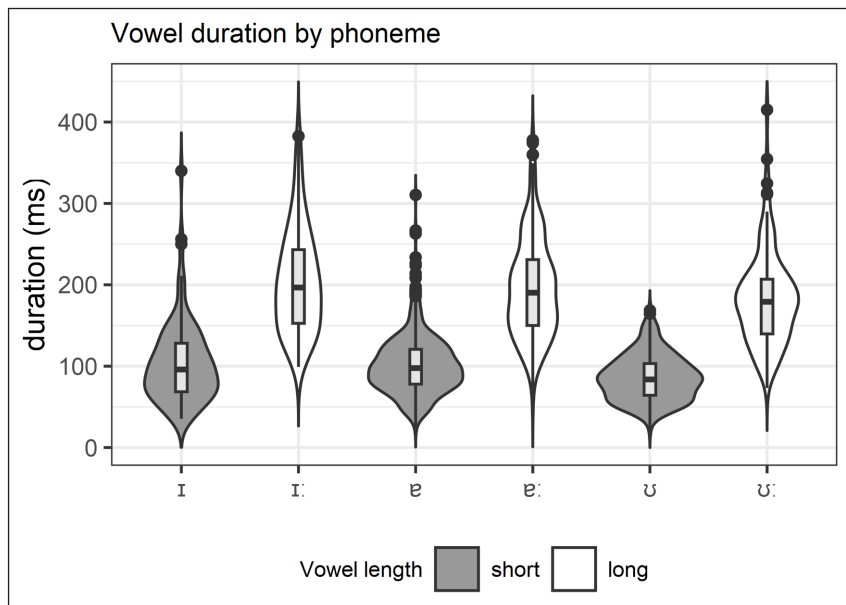


Figure 1: Duration values (ms) of the six Djambarrpuyngu vowels. Phonemically short vowels coloured grey, phonemically long vowels coloured white.

Vowel	<i>n</i>	Mean duration ms (SD)	Min. duration ms	Max. duration ms	Ratio of mean (long:short)
/ɪ/	112	103 (49)	35	340	1.99:1
/ɪ:/	80	205 (65)	99	383	
/e/	589	102 (37)	22	310	1.92:1
/e:/	240	196 (61)	53	378	
/ʊ/	294	86 (28)	24	169	2.13:1
/ʊ:/	117	183 (57)	73	415	
Length					
short	991	97 (37)	22	340	2:1
long	437	194 (61)	53	415	

Table 2: Summary of mean duration values (ms), standard deviation (ms), minimum duration (ms), and maximum duration (ms) for vowels by phonemic category and length, and ratios between relevant pairs (long:short).

Quality had a significant effect on vowel duration ($\chi^2(2) = 16.95, p < 0.001$). There was a small difference in duration for the three vowel qualities, between approximately 4 ms and 20 ms (see **Table 3**); however, only the comparison between the open central and the close back vowels was significant.

Comparison	Estimated difference (ms)	Standard error (ms)	<i>p</i> -value
<i>vowel length</i>			
short ~ long	-70.3	11.1	<0.0001
<i>vowel quality</i>			
/e/ ~ /ɪ/	3.6	6.2	1
/e/ ~ /ʊ/	18.3	4.4	0.0002
/ɪ/ ~ /ʊ/	14.7	6.7	0.092
<i>cons. category</i>			
obs ~ son	-33.4	4.9	<0.0001
<i>vowel length:syllable type</i>			
V open ~ V: open	-89.8	11	<0.0001
V open ~ V closed	14.2	5	0.0323

(Contd.)

Comparison	Estimated difference (ms)	Standard error (ms)	<i>p</i> -value
V open ~ V: closed	-36.6	13.2	0.0631
V: open ~ V: closed	53.2	9.2	<0.0001
V: open ~ V closed	104.1	11.2	<0.0001
V closed ~ V: closed	-50.8	13.3	0.0046
<i>vowel length:frame</i>			
V initial ~ V: initial	-62.6	11.3	0.0013
V medial ~ V: medial	-64.7	11.4	0.0009
V final ~ V: final	-83.7	11.3	0.0001
V final ~ V initial	-1.6	2.2	1
V final ~ V medial	4.1	2.3	1
V initial ~ V medial	5.7	2.3	0.2371
V: final ~ V: initial	19.4	3.3	<0.0001
V: final ~ V: medial	23	3.5	<0.0001
V: initial ~ V: medial	3.6	3.5	1

Table 3: Selected results from the post-hoc analysis comparing vowel duration ($n = 1431$).

The category of the following consonant also had an effect on duration ($\chi^2(1) = 40.24$, $p < 0.0001$), with vowels preceding obstruents an estimated 33 ms (SE = 5 ms, $p < 0.0001$) shorter than vowels preceding sonorants.

The interaction between vowel length category and syllable type had a significant effect on vowel duration ($\chi^2(1) = 14.83$, $p < 0.001$). **Figure 2** shows the distribution of duration values for short and long vowels faceted by whether they occurred in open or closed syllables. While the duration of the short vowels in both syllable types was similar, the duration values of the long vowels differed between open and closed syllables, and there was more overlap between the categories in closed syllables; long vowels are an estimated 53 ms (SE = 9 ms, $p < 0.0001$) longer when in open than in closed syllables. Short vowels, on the other hand, were only an estimated 14 ms (SE = 5 ms, $p = 0.0323$) longer in open syllables than in closed syllables. Because syllable structure affected long and short vowels differently, the differences between the two vowel length categories differed between the two syllable structures: When the syllable is open, long vowels are an estimated 90 ms (SE = 11 ms, $p < 0.0001$) longer than short vowels, but 51 ms (SE = 13 ms, $p = 0.0046$) longer when the syllable is closed. See ratio values for raw duration values in **Table 2**.

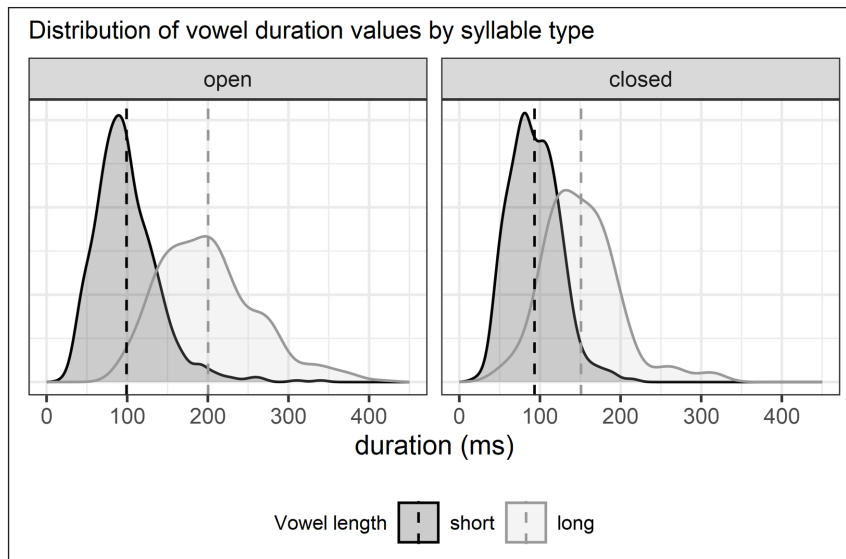


Figure 2: Distribution of duration values (ms) of vowels faceted by syllable type, with the vertical lines representing the group means. Phonemically short vowels coloured dark grey, phonemically long vowels coloured light grey.

Lastly, the interaction between vowel length and position in utterance was statistically significant ($\chi^2(2) = 32.54, p < 0.0001$). Differences between long and short vowels remained significant in all frame positions, but long vowels were significantly lengthened when the word was in utterance final position, compared with long vowels in either initial or medial frames. Short vowels did not show an effect of lengthening when the word was in utterance-final position.

4.2.2. Vowel formants

Figure 3 plots Lobanov normalised F1 and F2 values extracted at the vowel temporal midpoint, plotted separately for each speaker. Centroids are plotted with 95% confidence intervals; short vowel centroids are plotted in dark grey and ellipses are represented with a solid line; long vowel centroids are plotted in light grey and ellipses are represented by a dashed line. There is considerable overlap of the short and long vowel pairs for all speakers. However, the centroid for the long vowel in each pair is often more peripheral in the vowel space than the short vowel centroid in either the F1 or F2 dimension. F1 is generally relevant for the open vowel pair /e, e:/ with /e:/ having higher values than /e/, while F2 is relevant for close vowels /i, i:/ and /u, u:/, with long vowels having higher or lower values, respectively, than their short counterparts. The long vowels also have smaller ellipses than their short counterparts. It is possible that there is less variability in formant frequency values of these vowels; however, there are also fewer tokens. Mean F1 and F2 values (Hz), and standard deviation values are presented in **Table 4**.

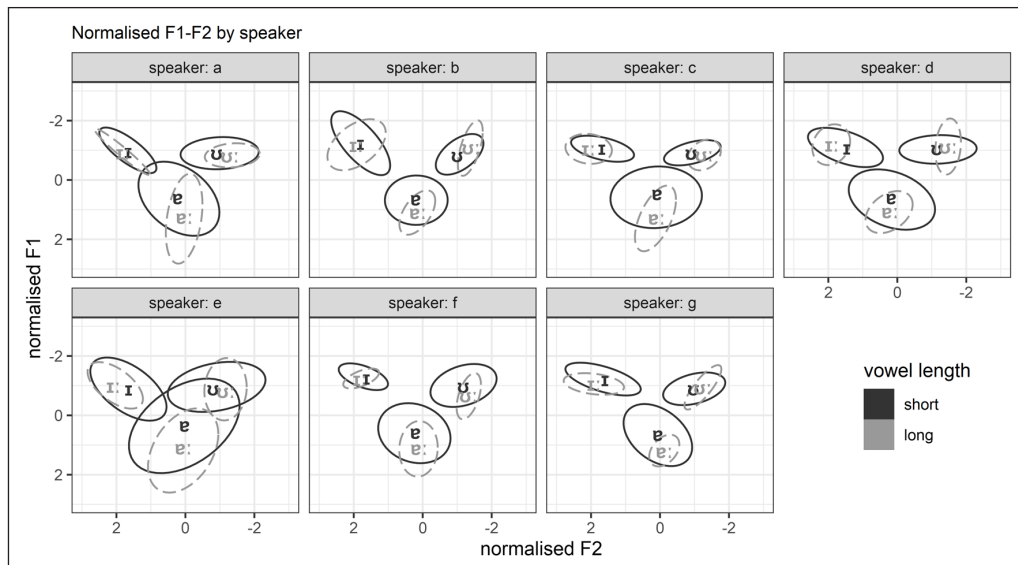


Figure 3: Mean first and second formant values (z -score; Lobanov normalised by speaker) at temporal midpoints of vowels with ellipses representing the 95% confidence intervals, plotted for each speaker (a–g). Phonemically short vowels are coloured dark grey with full line ellipses, phonemically long vowels are coloured light grey with dashed line ellipses.

	Vowel	n	F1 mean Hz (SD)	n	F2 mean Hz (SD)
<i>Women</i>					
	/ɪ/	61	399 (46)	59	2214 (165)
	/ɪ:/	41	399 (41)	41	2335 (126)
	/ɛ/	322	687 (115)	322	1588 (262)
	/ɛ:/	123	799 (117)	123	1561 (159)
	/ʊ/	158	420 (46)	158	1067 (212)
	/ʊ:/	64	437 (56)	64	928 (120)
<i>Men</i>					
	/ɪ/	51	373 (68)	51	2106 (223)
	/ɪ:/	39	386 (58)	39	2252 (185)
	/ɛ/	263	632 (145)	263	1434 (258)
	/ɛ:/	117	723 (141)	117	1404 (147)
	/ʊ/	136	416 (88)	136	1015 (231)
	/ʊ:/	53	408 (77)	53	866 (97)

Table 4: Summary of mean F1 and F2 frequencies (Hz) and standard deviations at vowel midpoints, for women and men (note: F2 for two /ɪ/ tokens could not be extracted).

Two models predicted F1 and F2 by vowel length category (two levels: long, short), vowel quality (three levels: open central, close front, close back), and gender (two levels: women, men) as fixed effects with an interaction between vowel length and vowel quality. Random effects included random intercepts for speaker, place of the following consonant, and word. Gender was included to account for variance in the data due to speaker sex, though is not examined in detail here.

Table 5 and **Table 6** present selected post-hoc comparisons from the statistical analyses of F1 and F2 values, respectively. For F1, the interaction between length and quality was found to have a significant effect ($\chi^2(2) = 41.11$, $p < 0.0001$). The open vowel pair /e, e:/ differed significantly in F1, with F1 being significantly lower for the short category vowel /e/ than the long vowel /e:/ by an estimated -97 Hz (SE = 9 Hz, $p < 0.0001$). For F2, the interaction between vowel length and quality was again found to be significant ($\chi^2(2) = 11.72$, $p = 0.0028$). For the /ɪ, ɪ:/ and /ʊ, ʊ:/ pairs, F2 was found to differ significantly in the expected directions, with /ɪ/ having significantly lower F2 than /ɪ:/ by an estimated -218 Hz (SE = 95 Hz, $p = 0.025$), and /ʊ/ having significantly higher F2 than /ʊ:/ by an estimated 179 Hz (SE = 62 Hz, $p = 0.005$).

Comparison	Estimated difference (Hz)	Standard error (Hz)	p-value
length:vowel quality			
/ɪ/ ~ /ɪ:/	20.51	18.60	0.2741
/e/ ~ /e:/	-96.68	9.43	<0.0001
/ʊ/ ~ /ʊ:/	-4.18	13.28	0.7535

Table 5: Selected results from the post-hoc analysis comparing F1 frequency (Hz) at vowel midpoints ($n = 1431$).

Comparison	Estimated difference (Hz)	Standard error (Hz)	p-value
length:vowel quality			
/ɪ/ ~ /ɪ:/	-218.4	94.9	0.0236
/e/ ~ /e:/	37.3	38.7	0.3367
/ʊ/ ~ /ʊ:/	179.0	61.8	0.0046

Table 6: Selected results from the post-hoc analysis comparing F2 frequency (Hz) at vowel midpoints ($n = 1429$).

4.2.3. Consonant duration

Duration of the C2 target consonants was statistically tested in a mixed effects model which included vowel length category (two levels: short, long), syllable type (two levels: open, closed,

i.e., where the consonant was intervocalic, or the coda, respectively), utterance frame (three levels: frame initial, frame medial, and frame final), and consonant category (two levels: obstruent, sonorant) as fixed effects with a three-way interaction between vowel length category, consonant category, and syllable type. Random effects included random intercepts for speaker and word, as well as by-speaker random slopes for the effect of vowel length category. Mean duration values for consonants and standard deviation values are presented in **Table 7**. This model was fitted specifically to test the three-way interaction between length of the vowel preceding the consonant, consonant category, and syllable structure.

		<i>n</i>	Duration mean ms (SD)	Ratio of mean (after long:after short)
	<i>Vowel length</i>			
	short	991	155 (93)	1:1.45
	long	437	107 (49)	
	<i>Consonant category</i>			
	obstruent	319	254 (88)	
	sonorant	1109	107 (48)	
	<i>Position in syllable</i>			
	intervocalic	1025	156 (92)	
	coda	403	99 (46)	
<i>Summary relevant for interaction between vowel length, syllable structure, and consonant category</i>				
	<i>C₂ cat.; vowel length</i>			
<i>Intervocalic</i>	<i>Obstruent</i>			
	after short vowel	261	269 (88)	1:1.37
	after long vowel	27	196 (50)	
	<i>Sonorant</i>			
	after short vowel	383	129 (54)	1:1.3
	after long vowel	354	99 (42)	
<i>Coda</i>	<i>Obstruent</i>			
	after short vowel	19	185 (56)	1:1.08
	after long vowel	12	172 (52)	
	<i>Sonorant</i>			
	after short vowel	328	92 (38)	1:0.94
	after long vowel	44	98 (36)	

Table 7: Summary of post-vocalic consonant duration values (ms), standard deviations (ms), and ratios between relevant pairs (after long:after short).

As shown in **Figure 4**, obstruents overall had longer duration values than sonorants. Intervocalic consonants were affected by the length of the preceding vowel, with consonants following short vowels being longer than those following long vowels. Consonants in coda position do not show an effect of the preceding vowel's length.

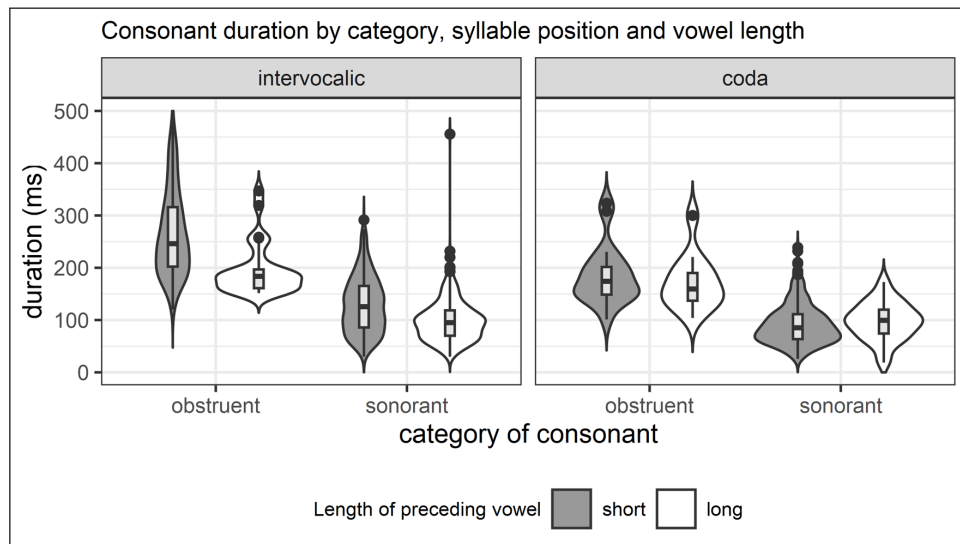


Figure 4: Duration values (ms) of post-vocalic consonants grouped by category, length of the preceding vowel, and syllable structure. Violins are coloured by preceding vowel length: consonants after short vowels coloured grey, after long vowels coloured white.

These observations were confirmed in the statistical analysis (see selected results in **Table 8**). There was a significant difference in duration between consonants following short vowels and consonants following long vowels, with an estimated difference of 33 ms (SE = 11 ms, $p < 0.005$). For obstruents there was a difference of 45 ms (SE = 18 ms, $p < 0.015$), and 20 ms (SE = 11 ms, $p < 0.057$) for sonorants. Therefore, without considering the structure of the syllable, a difference due to the length of the preceding vowel is observed only for obstruents. However, there was a significant effect of the three-way interaction between vowel length category, consonant category, and syllables type ($\chi^2(4) = 13.62$, $p = 0.0086$). The duration of consonants was found to differ consistently by preceding vowel length in open syllables, with sonorants following short vowels longer than those following long vowels by an estimated 34 ms (SE = 10 ms, $p < 0.036$). Obstruent duration was affected in the same way, with obstruents in open syllables following short vowels longer than those following long vowels by an estimated 79 ms (SE = 21 ms, $p < 0.008$). Duration differences were not significant when the consonants were in closed syllables ($p = 1$). Position in the utterance frame also had a significant effect on consonant duration ($\chi^2(2) = 16.55$, $p < 0.001$), with consonants in utterance-final words significantly longer than when in utterance-initial or -medial words.

Comparison	Estimated difference (ms)	Standard error (ms)	p-value
<i>vowel length</i>			
after V: ~ after V	32.8	11.3	0.0049
<i>vowel length:consonant category</i>			
after V:.obs ~ after V.obs	45.3	18.4	0.0151
after V:.son ~ after V.son	20.3	10.5	0.0571
<i>syllable structure</i>			
open ~ closed	40.8	10	0.0001
<i>vowel length:consonant category:syllable structure</i>			
after V.obs.open ~ after V:.obs.open	78.74	21.08	0.0081
after V.obs.closed ~ after V:.obs.closed	11.89	29.19	1
after V.obs.open ~ after V.obs.closed	89.32	19.06	0.0002
after V:.obs.open ~ after V:.obs.closed	22.48	29.61	1
after V.son.open ~ after V:.son.open	34.2	9.95	0.0366
after V.son.closed ~ after V:.son.closed	6.47	16.68	1
after V.son.open ~ after V.son.closed	39.62	8.43	0.0002
after V:.son.open ~ after V:.son.closed	11.89	16.11	1

Table 8: Selected results from the post-hoc analysis comparing post-vocalic consonant duration values (Hz) ($n = 1431$).

4.3. Summary and discussion

As hypothesised, vowel duration varied consistently with respect to the length categories of short and long. Long vowels, overall, were twice as long as short vowels. The third hypothesis, that intervocalic consonants would vary with the phonological length of the preceding vowel, was also confirmed. The ratio between the two consonant categories following short and long vowels was relatively similar, though overall, obstruents had longer duration than sonorants. One might have thought the putative fortis/lenis contrast would have resulted in less durational variation for obstruents, due to duration most likely being used as a means for distinguishing between the two stop series (Round, 2023). However, it could have also been the case that these fortis stops were lengthened to an even greater degree, as there would be no upper duration limit that could result in an overlap in categories. As lenis stops were not examined here, comments on their behaviour cannot be made. The finding that vowels are, overall, shorter before obstruents than sonorants may reflect a broader segment duration relationship in the language (see Section 4.2.1).

Overall, consonant duration varied to a lesser degree than vowel duration. For both consonant categories, the lengthening does not wholly make up for the shorter duration of short vowels; on average, short vowels were 97 ms and long vowels were 194 ms, while consonants were 155 ms after short vowels and 107 ms after long vowels. That is to say, as in the Shetland variety of English and Washo (van Leyden, 2004; Yu, 2008), while consonant duration is longer after short vowels by approximately 45% (see **Table 7**, overall ratio), this does not account for short vowels being only half the duration of long vowels. These findings contribute to the conclusion that contrastive vowel length, and not consonant length, remains the most plausible phonological explanation for the observed segment duration values.

The lengthening of the C2 consonant appears to result in the main stressed syllable in Djambarrpuyŋu being heavy, irrespective of the phonological analysis of consonants or vowels (as in Swedish, Thai, Washo, and the Shetland variety of English (Schaeffler, 2005; van Leyden, 2004; Yu, 2008)). The consonant lengthening pattern also does not appear to be conditioned by another word-level requirement, for example, bimoraic minimum, as the minimum was already met in this dataset because all the words were disyllabic. It could be, however, that there is a phonetic bimoraic minimum requirement for stressed syllables. The complementarity between vowel and consonant duration is taken up in the perception study, presented in Section 5.

With respect to the second hypothesis, formants were affected by vowel length; long vowels were more peripheral than short vowels. Close vowels varied in F2 (higher values for /ɪ:/ than /ɪ/, lower values for /ʊ:/ than /ʊ/), and the open vowel varied in F1 (higher values for /ɛ:/ than /ɛ/), reflecting findings reported for Gupapuyŋu (Graetzer, 2012) and tendencies for the relationship between duration and formant structure cross-linguistically (Cho, 2004; Lehiste, 1970, p. 31; Lindblom, 1963, p. 1780).³ The relatively compressed vowel spaces with considerable overlap between short/long pairs is observed in many Australian languages (Fletcher & Butcher, 2014).

The final hypothesis, that syllable structure would affect vowel and consonant duration, was supported. Long vowels were shorter in closed syllables than in open syllables, while short vowels remained relatively constant in duration. This suggests there may be a threshold beyond which vowels cannot be further compressed (Siddins et al., 2013; see also Nooteboom, 1997, pp. 656–658). The effect of syllable structure on long and short vowels consequently affected the ratio between categories (see **Table 2**). The duration of intervocalic C2 consonants varied inversely with the length of the preceding vowel, while the duration of coda C2 consonants

³ The long close front vowel had the longest duration, followed by the long open vowel, though the difference is not statistically significant for the two qualities overall (cf. Keating, 1985). As suggested by Christian DiCanio, this could be because vowel tokens as a function of coda type are not balanced, which could bias the duration of particular vowel categories. The effect of consonant category on vowel duration offers some support for that suggestion.

did not vary significantly due to vowel length. In Lebanese Arabic, it has been found that only long vowel duration is affected by following geminate consonants; they have shorter duration compared to when they are followed by singleton consonants, while short vowels remain the same duration in both conditions (Khattab & Al-Tamimi, 2014). That finding, while conditioned by a different process, echoes the Djambarrpuyŋu findings. Syllable association also operates in a different way in Dutch: C2 in CV1C2.CVC words was found to be of significantly longer duration when the V1 was short than when it was long, but in CV1.C2VC words C2 duration did not differ significantly (Jongman, 1998). Jongman argued that the status of the C2 consonant as ambisyllabic or tautosyllabic determines whether lengthening occurs, with only tautosyllabic consonants participating in the lengthening pattern. If the duration variation of the consonants was to enhance the vowel length contrast in Djambarrpuyŋu (in a way described by e.g., Solé & Ohala, 2010), we might expect that consonant duration values in Djambarrpuyŋu followed the Dutch pattern, as there is considerably more overlap in duration between the short and long vowel categories in closed syllables. However, we see no significant duration relationship between vowel and consonant duration when the consonant is coda to the first syllable. The opposing results agree that syllable structure “has a direct influence on the ... acoustic realization of individual segments” (Jongman, 1998, p. 219); however, it does not necessarily have the same influence.

5. Study 2: Vowel length perception

To follow up the acoustic analysis, this study furthers our understanding of the segmental contrasts in Djambarrpuyŋu through investigating Djambarrpuyŋu listeners perception of the vowel length contrast in a categorisation task. This study examines a reduced selection of possible factors contributing to the perception of vowel length. That is to say, V1 and C2 duration were considered in a minimal pair of words, *wəŋa* /wəŋe/ “to talk/speak” and *wāŋa* /wə:ŋe/ “home/place,” discussed further in Section 5.1.2.

This study tests two hypotheses:

- H1. Vowel length perception is categorical in the pair of words *wəŋa* /wəŋe/ “to talk/speak” and *wāŋa* /wə:ŋe/ “home/place.”
- H2. C2 consonant duration is used by listeners as a secondary cue to vowel length. Therefore,
 - i) consonant duration affects listeners’ perception when vowel duration is in the durationally ambiguous region in the pair of words *wəŋa* /wəŋe/ “to talk/speak” and *wāŋa* /wə:ŋe/ “home/place”;
 - ii) consonant duration affects the boundary between vowel length categories such that longer consonant duration conditions a later crossover boundary whereas shorter consonant duration conditions an earlier crossover boundary.

5.1. Methods

5.1.1. Listeners

Twenty native Djambarrpuyŋu listeners (19 analysed), living in Milingimbi, completed the experiment. Participants included ten women and ten men with an age range of 25 to 61 years (mean 41 years). The current study does not investigate age-related differences. All participants were familiar with a number of related Yolŋu languages, other Australian languages, and Australian English. Participants were recruited by the author or were suggested by participants who had already completed the task. Participants gave informed consent to participate in the study and have their responses analysed, and they were paid for their time.

One male participant's data were excluded from analysis due to irregular responses attributable to his extreme, near-total hearing loss. Five of the remaining participants reported having hearing difficulties in one or both ears, either permanently or intermittently. Hearing loss is not an uncommon issue in remote communities in Australia for Indigenous people, as childhood ear diseases such as otitis media and resultant complications occur at a considerably higher rate for Indigenous children than for non-Indigenous Australian children (O'Connor et al., 2009). It is believed that the listeners represent their wider speech community, and the 19 analysed participants' data show consistent patterns, suggesting that hearing difficulties did not affect participants' performance in this experiment.

5.1.2. Stimuli

The minimal pair selected for investigation was *waŋa* /wɛŋɛ/ “to talk/speak” and *wäŋa* /wɛ:ŋɛ/ “home/place.” The selection of this pair was motivated by two factors. Firstly, there is only a small number of minimal pairs that vary only in vowel length in Djambarrpuyŋu, and this minimal pair is cited in linguistic research (e.g., Morphy, 1983) and by Djambarrpuyŋu speakers as demonstrative of the vowel length contrast. Secondly, these items are believed to be common in daily speech.

Due to the absence of an acoustic analysis of the fortis/lenis contrast in the stops, a perception study including stop duration in relation to vowel duration is not possible at this stage (also, a four-way minimal set that contains phonemic short and long vowels with fortis and lenis stops is not known to the author).

Durational measures from a subset of the production data were used to determine the range of segment duration values for the experiment. In those data, the durational range for the short vowel in /wɛŋɛ/ target words was 102 ms to 158 ms, and for the long vowel in /wɛ:ŋɛ/ target words was 132 ms to 286 ms, while the durational range for the nasal in /wɛŋɛ/ target words was 100 ms to 211 ms, and for the nasal in /wɛ:ŋɛ/ target words was 67 ms to 129 ms.

A single token of /wɛ:ŋɛ/ from the production data, spoken by a male speaker, 39 years of age, was selected to create the audio stimuli. This vowel is plotted in the F1–F2 space of all vowels to illustrate the quality of the selected token (**Figure 5**); it is in the overlapped area for the short and long open vowel counterparts. In this token, the vowel was 181 ms in duration, and the nasal was 83 ms in duration. The steady state portions of the vowel and nasal were selected for manipulation to preserve transitional information that occurred at the segment boundaries. This resulted in the vowel having 97 ms that was transitional from the preceding approximant and into the following nasal, and the nasal having 35 ms that was considered as transitional to/from the surrounding vowels. Therefore, in the token, 84 ms of the vowel and 48 ms of the nasal were manipulated.

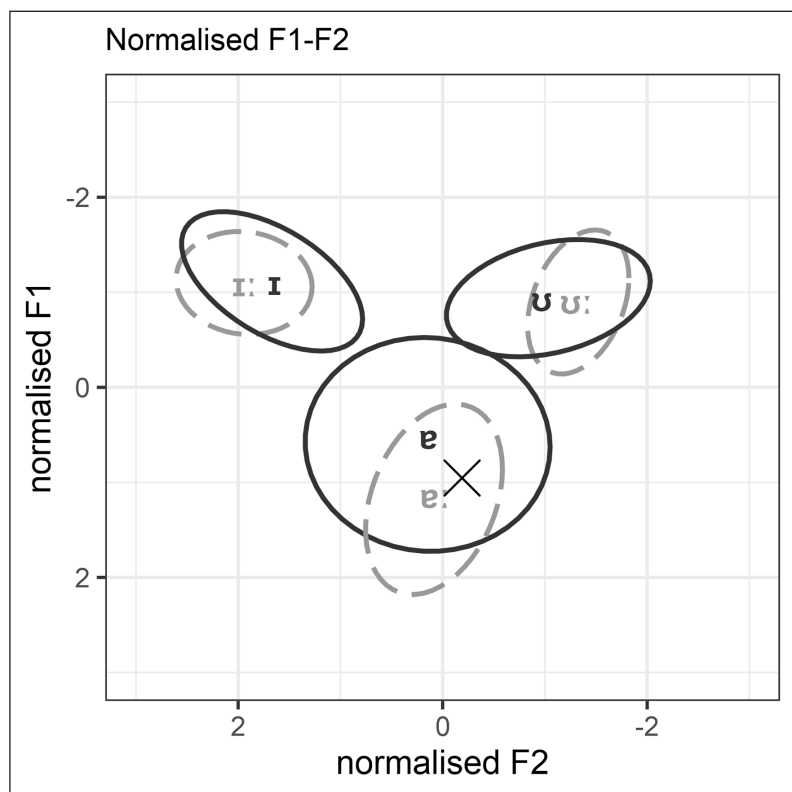


Figure 5: Plotted normalised first and second formant values (z-score; Lobanov normalised) at the temporal midpoint of the *wāŋa* /wɛ:ŋɛ/ “home/place” token used to create the stimuli for the perception experiment (marked with ×), plotted against all data for all speakers; ellipses represent the 95% confidence intervals.

The duration of the vowel and the nasal segment was manipulated independently. Continua of seven equidistant steps were created for the vowel and for the nasal (following Lehnert-LeHouillier, 2010). Stimuli were created in Praat. The experiment was full factorial in its design;

therefore, stimuli contained all combinations of the manipulated vowel steps and nasal steps, resulting in 49 unique stimuli. In the stimuli, vowels ranged between 110 ms and 260 ms, with 25 ms difference between steps. The consonant duration ranged from 90 ms to 200 ms, with 18 ms difference between steps. Due to the full factorial design, some stimuli contained vowel and consonant duration combinations that would infrequently, if ever, occur in natural speech. See **Table 9** for the duration of steps in the continua.

Step	Vowel (ms)	Nasal (ms)
1	110	90
2	135	108
3	160	127
4	185	145
5	210	163
6	235	182
7	260	200

Table 9: Duration (ms) of vowel and nasal steps in stimuli. Experimental design was full factorial resulting in a stimuli set in which each vowel step and nasal step combination occurred (total: 49 unique stimuli).

5.1.3. Procedure

The experiment was constructed and presented in OpenSesame (v. 3.1; Mathôt et al., 2012) using a laptop computer. The legacy backend option was selected, which made use of PyGame. Stimuli were presented in randomised order in four blocks that contained 49 trials (i.e., all stimuli occurred once in each block), resulting in each participant completing a total of 196 trials. The first block of trials was originally included as a training block. Its results were consistent with the following three blocks and so responses from the first block have been included in the analysis. Variation due to block is discussed in the results where relevant.

Participants' categorisation of stimuli was obtained through a forced-choice task in which participants were asked to identify the test stimulus as either *wəŋa* /wəŋə/ "to talk/speak" or *wāŋa* /wə:ŋə/ "home/place." An image represented each word onscreen (see **Figure 6**). In each trial, a stimulus was played once, and there was no constraint on the time to respond. Participants made their selection by choosing either the LEFT SHIFT key or RIGHT SHIFT key on the laptop keyboard, which corresponded to the onscreen visual representations of the words. The keyboard prompts remained the same throughout the trials and blocks, as did the arrangement of images onscreen. The participant either made the selection themselves or indicated their selection (by pointing to the left or to the right) to the researcher, who pressed

the indicated SHIFT key. Participants used the latter method if they were unfamiliar or not confident with using computers. For this reason, and because participants were not asked to respond as quickly as possible, reaction times are not examined. Responses were collected using the keyboard response collection function in OpenSesame. Between each trial, a fixation dot appeared in the centre of the screen to indicate a new trial was about to commence. Between each block, there was a “break time” screen which allowed participants to have a break of a duration of their choosing. The experiment took approximately 20 minutes to complete, including preliminary discussion about the experiment, signing the consent form, and within-experiment breaks.



Figure 6: Visual representations of *wəŋa* /wəŋe/ “to talk/speak” (left) and *wāŋa* /wə:ŋe/ “home/place” (right) used onscreen in the perception experiment.

Instructions for the task were translated from English to Djambarrpuyŋu by Albena Buyanggirr, a Djambarrpuyŋu speaker and translator. The instructions were presented on the laptop in OpenSesame. The researcher sat with each participant and they read through the instructions together. The participant indicated their willingness to participate in the study and indicated if they had hearing difficulties through buttons onscreen, selected using the laptop trackpad.

Participants were told that the researcher wanted to learn about how Yolŋu (i.e., Indigenous people who speak Yolŋu languages like Djambarrpuyŋu) listened to the sounds in words. Participants were instructed to listen carefully, as each stimulus would be played only one time per trial, and to respond thoughtfully, as the accuracy of their response was important, but not the speed. It was explained that the two SHIFT keys on the laptop keyboard corresponded to the pictures on screen which represented the words *wəŋa* and *wāŋa*, and upon making a decision, the participant needed to select one of the SHIFT keys to indicate the word they heard. Finally,

participants were told the experiment was not a test, and therefore there was no right or wrong answer.

Participants listened to stimuli using Philips SHL3060BK over-ear headphones. Eighteen participants (including the participant whose data were excluded), completed the experiment in an outside location, and two participants were seated inside a house while completing the experiment. Background noises and distractions existed for all participants. If a participant was unable to make a decision or was distracted for a particular trial, the researcher indicated this by using the “P” key to pass the trial. The researcher only used the “P” key following direct instruction from the participant. This response was selected extremely infrequently ($n = 6$), and those responses were not included in the analysis presented in 5.2.

5.1.4. Data processing and analysis

A logistic mixed model was used to test the data in R using the `glmer()` function from the `lme4` package (Bates et al., 2015). The model reported on included response (*waja/wāja*) as the dependent variable, vowel duration step (seven levels), nasal duration step (seven levels), and block (four levels) as fixed effects with an interaction between vowel duration step and nasal duration step. The random effects structure included intercepts for participant (19 levels). This model was selected based on theoretically motivated assumptions and to address specific, motivated questions. Results for vowel and nasal duration are the focus of Section 5.2, with block mentioned briefly. Emmeans was used to calculate pairwise comparisons (bonferroni corrected). From ANOVA comparisons between the selected model and null models, Chi squared and p -values are reported in this paper. For pairwise comparisons from the post-hoc analysis, selected z -values and p -values are reported. In the results, the phonemic representations of the words are used.

5.2. Results

Figure 7 shows the proportion of /wɛŋɛ/ and /wɛ:ŋɛ/ responses to each stimulus, pooled for the 19 participants. Along the x-axis is vowel step, and on the y-axis is the nasal step. Each cell represents one stimulus. The colour of the cell represents the count of /wɛ:ŋɛ/ responses: A lighter colour reflects more /wɛ:ŋɛ/ responses, the darker the cell, the fewer /wɛ:ŋɛ/ responses (i.e., more /wɛŋɛ/ responses). Each stimulus has maximally 76 responses.

Of the 3,718 total valid responses, 1,378 were /wɛŋɛ/ and 2,340 were /wɛ:ŋɛ/. The responses, therefore, are skewed towards /wɛ:ŋɛ/ (63%). Possible reasons for this are discussed further in Section 5.3. Overall, participants selected /wɛ:ŋɛ/ less frequently when a stimulus contained a vowel at step 1 or 2 in the continuum, and more frequently when a stimulus contained a vowel at step 4, 5, 6 or 7 in the continuum. As can be seen in **Figure 7**, stimuli with vowel step 3

form an intermediate group for which /wɛ:ŋɛ/ and /wɛ:ŋɛ/ are selected with relatively equal frequency. Nasal step, on the other hand, does not appear to have a consistent effect on listeners' responses.

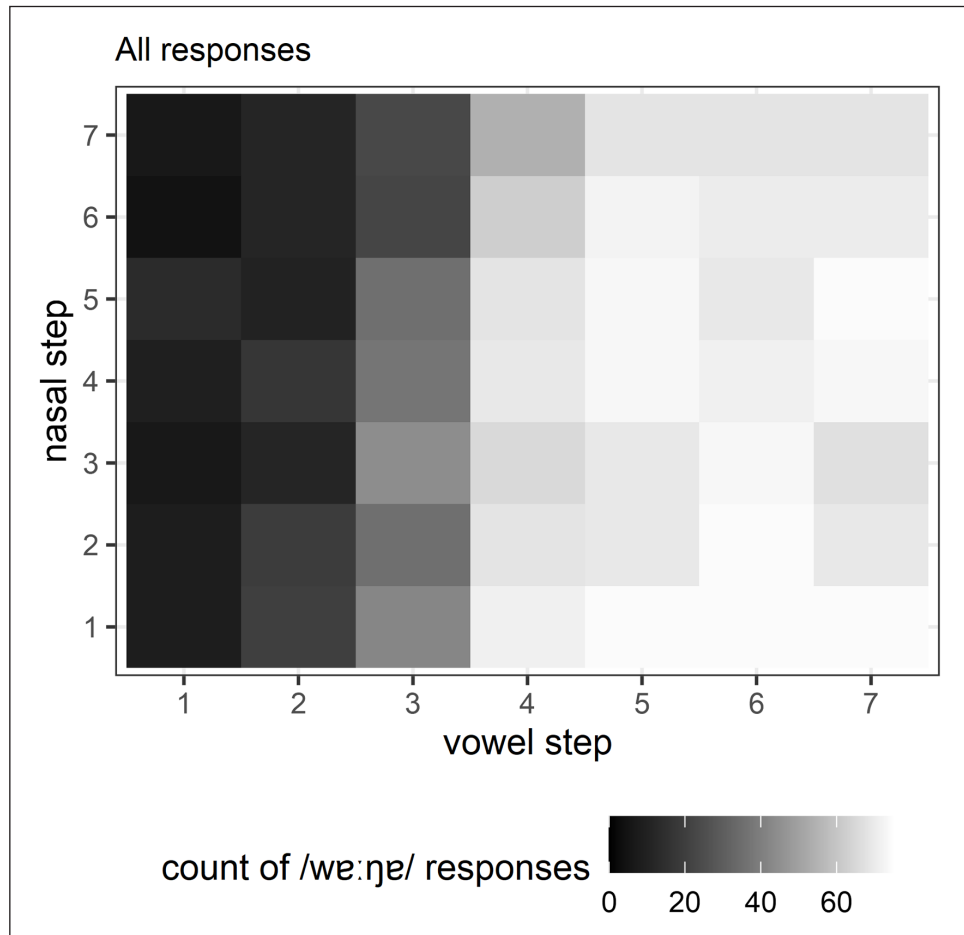


Figure 7: Summary of /wɛ:ŋɛ/ responses to each stimulus pooled for all participants. The more /wɛ:ŋɛ/ responses to a stimulus, the lighter the cell colour.

5.2.1. Vowel duration

Figure 8 presents the proportion of /wɛ:ŋɛ/ responses as a function of vowel step. Individual listeners' response frequencies are plotted by points to provide an indication of the range of responses in the data; points are overlapped so a darker point indicates more overlapping listener responses. There is a sharp categorical boundary between steps 2 and 4, the stimuli either side of which only show a small amount of variation. For stimuli containing vowels of step 5, 6 or 7, there is little sensitivity to the changes in vowel duration. Note also the greater dispersion of

individual listener's responses to the vowel step 3 stimuli compared with other step categories. Taken together, it appears that the crossover point between the short and long vowel categories is around vowel step 3, where participants overall performed at chance (Lehnert-LeHouillier, 2010; Liberman et al., 1957).

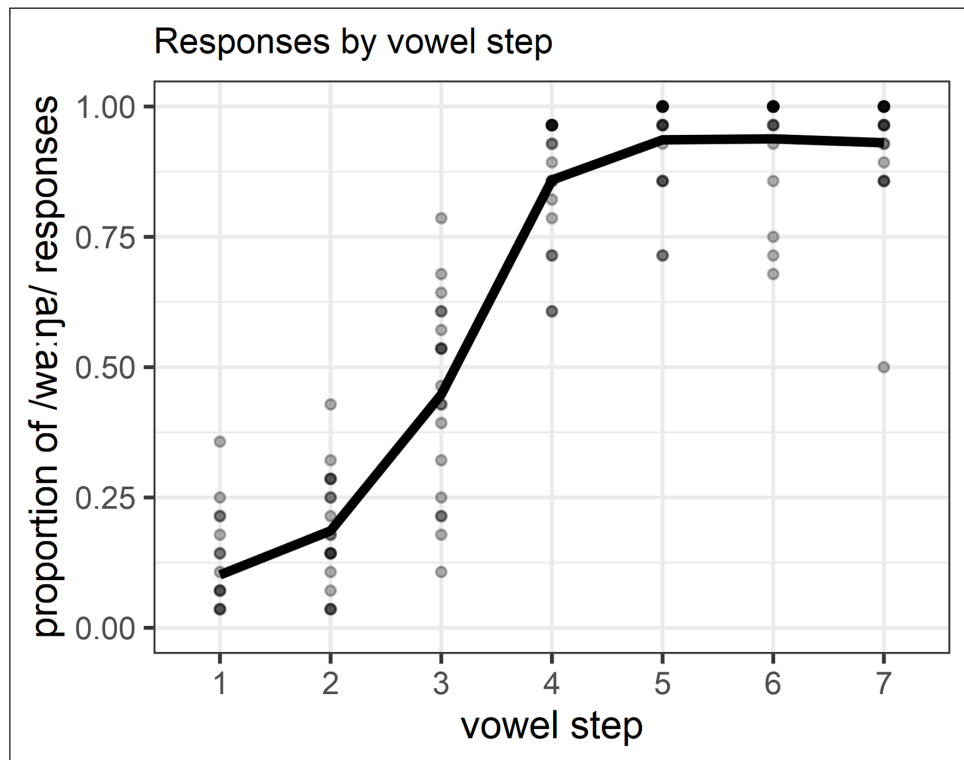


Figure 8: Proportion of /we:ŋe/ responses as a function of vowel step. Individual listeners' response frequencies represented by points.

Post-hoc comparisons between the vowel steps supported the pattern observed in **Figure 7** and **Figure 8**. Comparisons of response to stimuli with vowel duration steps at the higher end of the continuum were not significant: that is, comparisons between vowel steps 5~6, 5~7, 6~7 ($p > 0.05$) (see **Table 10**). All other comparisons were significant.

The proportion of /wɛŋe/ and /we:ŋe/ responses to each stimulus in each block of trials, pooled for the 19 participants, remained similar for each block, though the statistical analysis showed that there was a significant effect of block on listener's categorisation ($\chi^2(3) = 28.794$, $p < 0.0001$). Pairwise comparisons revealed that block 1 was significantly different from blocks 2–4, which did not differ significantly from one another. Specifically, more /we:ŋe/ responses were observed in block 1.

Comparison	z-value	p-value
<i>vowel step</i>		
1 ~ 2	-3.950	0.0016
1 ~ 3	-11.775	<.0001
1 ~ 4	-20.789	<.0001
1 ~ 5	-20.646	<.0001
1 ~ 6	-20.405	<.0001
1 ~ 7	-20.316	<.0001
2 ~ 3	-9.048	<.0001
2 ~ 4	-19.555	<.0001
2 ~ 5	-19.231	<.0001
2 ~ 6	-18.965	<.0001
2 ~ 7	-18.871	<.0001
3 ~ 4	-13.509	<.0001
3 ~ 5	-14.292	<.0001
3 ~ 6	-14.117	<.0001
3 ~ 7	-13.948	<.0001
4 ~ 5	-3.686	0.0048
4 ~ 6	-3.803	0.0030
4 ~ 7	-3.474	0.0108
5 ~ 6	-0.176	1
5 ~ 7	0.155	1
6 ~ 7	0.327	1

Table 10: Selected results from the post-hoc analysis comparing vowel step ($n = 3718$).

5.2.2. Consonant duration

Figure 9 presents the proportion of /wɛ:ŋɛ/ responses as a function of nasal step. It is clear that nasal duration did not have as great of an influence on listeners' perception compared with vowel duration. However, post-hoc tests showed listeners' categorisation was significantly affected by nasal step (**Table 11**). Differences were significant ($p < 0.05$) for all comparisons that included nasal duration step 7, except 3~7 and 6~7. The comparison between nasal steps 1~6 was also found to be significant. This reflects the decrease in /wɛ:ŋɛ/ responses observed for nasals at the longer end of the nasal duration step continuum compared with stimuli containing nasals at the shorter end of the continuum in **Figure 9**.

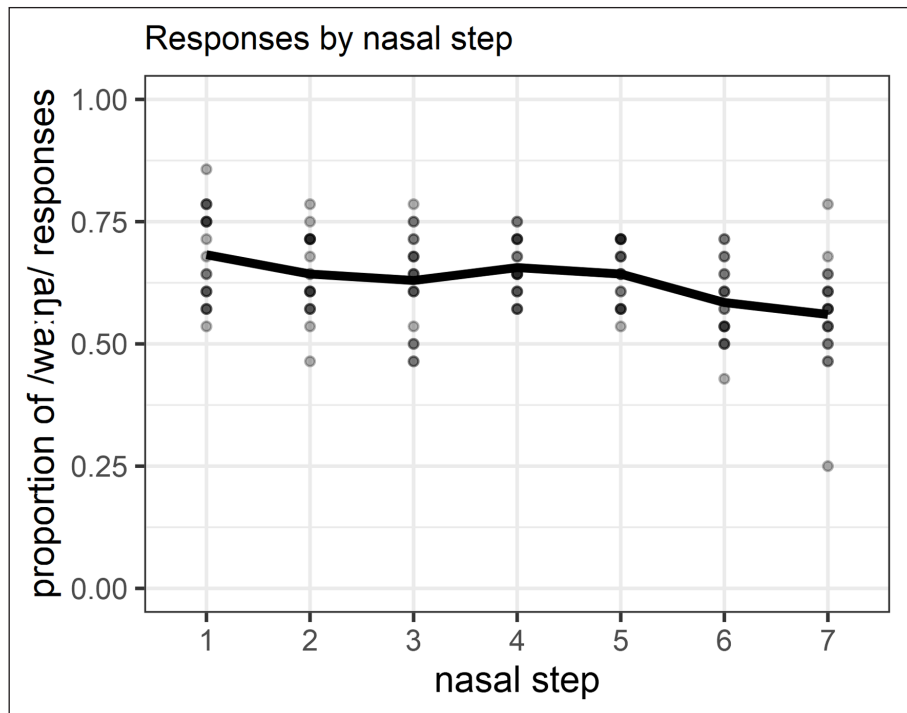


Figure 9: Overall proportion of /we:ŋe/ responses as a function of nasal step. Individual listeners' response frequencies represented by points.

Comparison	z-value	p-value
<i>nasal step</i>		
1 ~ 2	2.015	0.9215
1 ~ 3	2.986	0.0593
1 ~ 4	1.393	1
1 ~ 5	1.600	1
1 ~ 6	3.913	0.0019
1 ~ 7	4.915	<0.0001
2 ~ 3	1.064	1
2 ~ 4	0.678	1
2 ~ 5	0.409	1
2 ~ 6	2.152	0.6593

(Contd.)

Comparison	z-value	p-value
2 ~ 7	3.238	0.0253
3 ~ 4	1.744	1
3 ~ 5	1.446	1
3 ~ 6	1.140	1
3 ~ 7	2.196	0.5896
4 ~ 5	0.253	1
4 ~ 6	2.803	0.1064
4 ~ 7	3.908	0.0020
5 ~ 6	2.489	0.2690
5 ~ 7	3.546	0.0082
6 ~ 7	0.954	1

Table 11: Selected results from the post-hoc analysis comparing nasal step ($n = 3718$).

In **Figure 7**, it can be seen that for stimuli containing a vowel at either end of the vowel continuum, listeners reliably selected one of the following responses: /wəŋɐ/ (i.e., short vowel word) when the vowel was at the shortest point in the continuum (vowel step 1), and conversely selected /wɛ:ŋɐ/ (i.e., long vowel word) when the vowel was at its longest point in the continuum (vowel step 7), irrespective of nasal duration. The vowel step 3 data presents a different case, in which listeners equally often selected a short or long vowel for the first five nasal duration steps. However, at steps 6 and 7, when the nasal was at the two longest points in the nasal continuum, there was a decline in /wɛ:ŋɐ/ responses (indicated by the darker colour in those cells in **Figure 7**). That is, when stimuli containing a vowel of step 3 also contained a nasal at the longest or second longest point in the nasal duration continuum, listeners selected the long vowel word /wɛ:ŋɐ/ approximately 30% of the time. However, the interaction between vowel step and nasal step was not significant ($\chi^2(36, N = 3718) = 38.34, p = 0.36$). Nevertheless, there are further trends that can be observed.

Figure 10 presents responses to stimuli that contained nasal step 1 (full black line), nasal step 7 (dashed black line), and all data (thin grey line). If considering stimuli that only contained nasal step 1, listeners' crossover points between categories occurred earlier than the average across all the data. When stimuli contained a nasal at step 7, listeners' crossover point between categories occurred later.

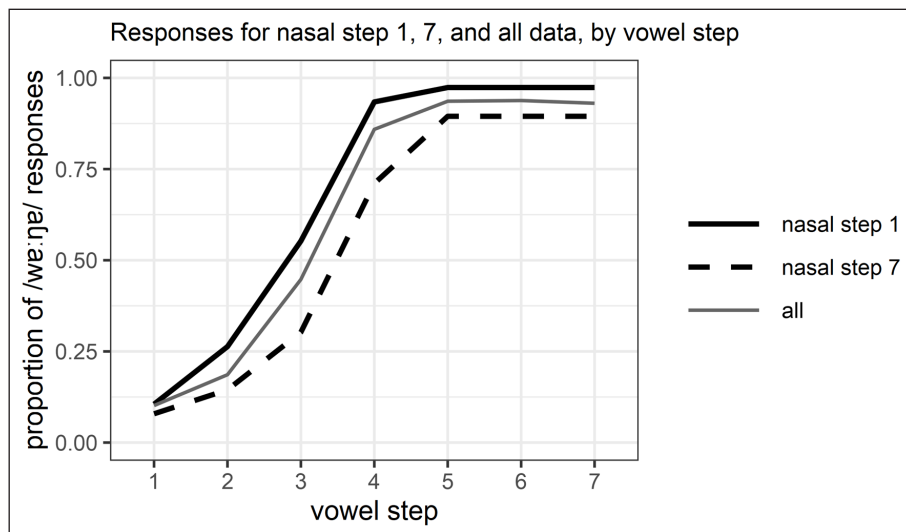


Figure 10: Proportion of /wɛ:ŋɛ/ responses to stimuli containing a step 1 nasal (full black line) and step 7 nasal (dashed black line), as well as across all the data (thin grey line) as a function of vowel step.

5.3. Summary and discussion

It was hypothesised that vowel length perception is categorical in the pair of words /wɛŋɛ/ and /wɛ:ŋɛ/. The results from the analysis show that listeners' selection in this forced-choice task were strongly influenced by vowel duration. There was a shift between /wɛŋɛ/ and /wɛ:ŋɛ/ at around vowel step 3 when considering data pooled across the 19 participants. Vowels at step 3 were ambiguous as to their length category as evidenced by participants performing at chance for stimuli containing those vowels (Lehnert-LeHouillier, 2010). Recall that in the corpus, the duration of the short vowel in /wɛŋɛ/ ranged from 102 to 158 ms (median = 127), and 132 to 286 ms (median = 214 ms) in /wɛ:ŋɛ/. Therefore, the location for the crossover point between short and long vowel words in perception, around vowel step 3 in the vowel continuum (160 ms), is in line with the production data in terms of the durational ranges for short and long categories.

The duration of the nasal in C2 affected listeners responses in two ways. Firstly, it was hypothesised that the consonant's duration would have the strongest effect when the vowel's duration was ambiguous. **Figure 7** suggests that consonant duration is used by listeners as a secondary cue to vowel length when vowel duration is in the durationally ambiguous region between short and long (vowel step 3). Secondly, an overall effect of consonant duration on vowel length perception was hypothesised, which is observed by the leftwards and rightwards shift in the vowel category crossover points in **Figure 10** for the nasal step 1 and 7 data. Specifically, shorter consonant duration conditioned an earlier crossover between the length categories, while longer consonant duration conditioned a later crossover. There were also fewer /wɛ:ŋɛ/ responses to stimuli with nasal step 7, even when the vowel was at its longest. Together, these findings

suggest listeners made use of secondary cues when the primary cue was ambiguous (Lehnert-LeHouillier, 2010), as well as phonologically redundant phonetic information in their perception of the vowel length contrast in this minimal pair (Whalen et al., 1993). The pattern observed in the figures corresponds to results for Thai (Roengpitya, 2001; see also Abramson & Ren, 1990, on formants) and Norwegian (van Dommelen, 1999), which showed similar movement in the crossover point in vowel categorisation due to the duration of the following consonant.

Because nasal duration did not have a strong effect on listeners' responses whereas vowel duration showed a pattern consistent with categorical perception, the results of the perception study support the analysis that only vowel length is contrastive in the language. The results from this study do suggest that the inverse lengthening of consonants may perceptually enhance the vowel length contrast for this minimal pair and possibly other words. It is possible that this finding would not be observed in a study comparing words with obstruents. That is because the nasal steps are shorter in duration than the vowel steps in the perception task, and sonorants overall are shorter than obstruents in the language. Both of these factors might mean that the nasal (i.e., sonorant) duration is more of a secondary perceptual cue than if there had been a perceptual comparison between words with obstruents.

With respect to block, recall that it also had a significant effect. One possibility for this result is that listeners became more familiar with the range of durational values contained in the data over time. Inspection of the data showed that listeners' decisions became more consistent throughout the blocks, especially so for stimuli containing vowel step 1. Overall, responses in block 1 were significantly different from blocks 2, 3, and 4, while blocks 2, 3, and 4 did not differ significantly.

The overall greater proportion of /wɛ:ŋɛ/ responses may be due to a number of factors, some of which are discussed here. It could be that the continuum did not include vowels of short enough duration. This is supported in the lack of a level state at the lower end of the vowel continuum in **Figure 8**. Another possibility is that because the original token contained a long vowel and listeners had access to a range of acoustic cues other than duration, these cues conditioned more /wɛ:ŋɛ/ responses. For example, the small difference in formants may have affected listeners, at least in the trials in block 1, though the manipulated token was within the overlapped areas for short and long vowels (**Figure 5**). An additional factor that could be utilised by listeners that may be examined in the future is the f_0 contour (see e.g., Lehnert-LeHouillier, 2010; Sanker, 2019), which has been found to differ between short and long vowels in Djambarrpuyŋu, but has not been subject to thorough investigation (Jepson, 2019a, pp. 104–109). The presentation of the stimuli could have also been a factor in the skewed responses as the keyboard responses and on-screen images were not counter-balanced throughout the trials or between participants. This may have led to a bias to select the right-side response button (i.e., selecting the long vowel word /wɛ:ŋɛ/) (Richardson, et al., 2020). These factors will be taken into account in future perception studies with Yolŋu listeners.

6. Conclusions

Two studies on vowel length in Djambarrpuyŋu were presented in this paper. The first study provided quantification of acoustic correlates of the vowel length contrast in Djambarrpuyŋu, addressing the first research question: How does vowel length affect vowel duration, vowel formants, and duration of the following consonant in Djambarrpuyŋu? The second study explored the influence of vowel duration and post-vocalic nasal duration on vowel length categorisation in perception, showing that vowel duration is the primary cue listeners use to categorise words even though both vowel and consonant duration varied in the stimuli. This addresses the second research question: What durational cues do listeners use to categorise words that vary phonemically by vowel length?

In the production study, it was shown that the duration contrast between long and short vowels was substantial, with a ratio of about 2:1 (~194 ms vs. ~97 ms). While this pattern was affected by syllable structure (long vowels had shorter duration in closed than open syllables), it remained robust in the two-syllable structure conditions. Consonant duration differed by length category of the preceding vowel, though the durational differences were considerably smaller than vowels for both obstruent and sonorant categories. The production study results concur with anecdotal reports for related languages in which consonants have been proposed to lengthen following stressed short vowels in open syllables.

The details of the results most strongly resemble those found for the Shetland variety of English (van Leyden, 2004), in part because the overall consonant duration pattern is relatively weak. There are also similarities to results for Washo (Yu, 2008) and Thai (Abramson, 1960); however, consonant duration variation is greater in those languages. As discussed in Section 2.2, Yu (2008) noted that the complementarity between vowel and consonant duration in Washo can make the interpretation of duration results difficult. The results for Djambarrpuyŋu, however, show much more consistent and substantial vowel differences than consonant differences. This supports the conclusion that vowel duration is phonological and addresses the third research question: Do the production and perception results support vowel length being contrastive in Djambarrpuyŋu?

The effect of syllable structure on vowel length was as one might expect, with long vowels being shorter in closed syllables than open syllables. Other cross-linguistic patterns of vowel duration, such as long vowels being affected by finality to a greater degree than short vowels, was also observed in these data. As well as durational differences, the first and second formants of vowels were also found to vary by vowel length, with long vowels being more peripheral in the vowel space than short vowels, though considerable overlap would suggest short and long vowels functionally have the same quality. Considering analyses of other small vowel systems with a reported length contrast, there appears to be a trend that long and short vowels differ

somewhat in their quality, but the difference is not large or consistent enough to justify two phonemic categories; for instance, this is the case for Chickasaw (Gordon et al., 2000). It is possible that this is due to the relatively large acoustic spaces these vowels occupy, relative to languages with more vowel qualities, which necessarily must be produced with less variability.

In the perception study, vowel length perception was shown to be categorical, and the perceptual categories align with the ranges for these categories in production. Consonant duration affected listeners' decisions when the vowel's duration was ambiguous and also had an overall effect of moving the category crossover boundary either slightly towards more short or long vowel word perceptions. These findings support the acoustic results and again corroborate the analysis that vowel length is phonological.

While both the production and perception studies answer “yes” to the third research question, understanding the motivations behind the phonetic patterns observed is not as clear. It can be said with some confidence that vowel duration varies because vowel length is contrastive and so the duration of a vowel in the stressed syllable is determined by its length category. Regarding consonant duration variation, it can be said that for intervocalic consonants following stressed vowels, that vowel length predicts a complementarity in duration with longer consonants after short vowels, and shorter consonants after long vowels. But a further phonological motivation is more difficult to determine. Positing a requirement for the stressed syllable of words to be phonologically heavy, as described for Djapu and Swedish (Morphy, 1983; Schaeffler, 2005), either through having a long vowel or being closed (by a coda, or lengthening of the following consonant and its possible designation as being ambisyllabic) would explain the patterns observed. This suggests a weight requirement of some type. One option could be weight sensitivity. However, there does not appear to be evidence for stress moving to the second syllable of a word, for example, if heavy when the first syllable is phonemically light. Unlike in other languages with fixed stress and vowel length such as Finnish (Suomi et al., 2013), in Djambarrpuyŋu vowel length is not contrastive outside of stressed syllables. Another option could be a requirement for stressed syllables to be minimally bimoraic, which is achieved phonetically by long vowels or consonant lengthening within the syllable (e.g., Prince, 1990). This avenue requires further consideration from a phonological perspective but appears to be supported by the phonetic findings presented here.

Another possible explanation for this pattern, supported by this pair of studies, is that consonant duration variation phonetically enhances the vowel length contrast (Solé & Ohala, 2010). As discussed in Section 2.2, consonants are proposed to be given priority over vowels in Australian languages due to the large number of places of articulation (Butcher, 2006). While findings suggest the consonant duration varies in support of the vowel's length, it is also possible that the vowels investigated here offer additional acoustic cues to the consonant's place

of articulation, supporting the place of articulation imperative, for example, through formant transitions. The two possibilities of consonant duration enhancing the length contrast, and aspects of vowels (e.g., transitions) enhancing place features for consonants are not mutually exclusive, though require further investigation.

These findings also highlight a confound in Jepson et al. (2021), in which only disyllabic words showed an effect of post-tonic consonant lengthening. In that study, only words containing short stressed vowels were considered, that is, words which also present the environment for consonant lengthening in the present production study. Jepson et al. (2021) raise the possibility that for consonants shortened after long vowels—based on their findings that in words longer than three syllables in length—intervocalic C2 consonants are no longer in duration than intervocalic consonants elsewhere in the word. Do we expect that consonants following long vowels in multisyllabic words are shortened further, compared with those in disyllabic words? Or, could it be that the durational lengthening of these particular consonants due to vowel length applies only within disyllabic words? It may be that the small durational differences for consonants reduced further due to effects of polysyllabic shortening (White & Turk, 2010). Additional investigation is required to understand the interactions between these different factors contributing to the duration of both vowels and consonants.

This paper contributes to our understanding of the phonetics and phonology of Djambarrpuyngu and expands the typological understanding of contrastive vowel length, and segment duration relationships. It is still unknown how the vowel length contrast is realised in words longer than two syllables. Additionally, the duration of vowels outside of the first syllable (which are non-contrastive for length) remain to be acoustically examined to assess if they pattern with the short or long vowels, or if they form an intermediate non-contrastive category. Further acoustic and, ideally, perception studies of other Australian languages with a vowel length contrast will help reveal the phonetic durational patterns of speech segments in these languages.

Data accessibility statement

The data tables for the two studies and the scripts used to analyse and plot the data are available on Open Science Framework: <https://osf.io/qke2r/>.

Acknowledgements

I wholeheartedly thank the Yolŋu who participated in these studies. Many thanks to †Paula Madiwirr who assisted in creating the frames for the production study, and to Albena Buyangirr for translating the instructions for the perception experiment from English to Djambarrpuyŋu. Cultural and linguistic information in this paper has been shared with permission from Yolŋu community members, and I acknowledge that this information remains the Indigenous Cultural Intellectual Property of Yolŋu. I thank Janet Fletcher, Hywel Stoakes, Ruth Singer, Melanie Wilkinson, Caroline Jones, Bert Remijsen, Jonathan Harrington, Felicitas Kleber, Josiane Riverin-Coutlée, Rasmus Puggaard-Rode, and Jonáš Podlipský along with other attendees of the Vowel and Consonant Quantity workshop in Zurich for comments and suggestions, as well as two anonymous reviewers and Associate Editor Christian DiCanio.

Funding information

This project received funding from The University of Melbourne (PhD Fieldwork Grant), and through the ARC Centre of Excellence for the Dynamics of Language (Project ID: CE140100041). The author received funding from the Alexander von Humboldt Foundation while writing the manuscript.

Competing interests

The author has no competing interests to declare.

References

- Abramson, A. S. (1960). *The vowels and tones of Standard Thai: Acoustical measurements and experiments*. [Doctoral dissertation, Columbia University].
- Abramson, A. S., & Ren, N. (1990). Distinctive vowel length: Duration vs. spectrum in Thai. *Haskins Laboratories Status Report on Speech Research*, 256–268.
- Anderson, V. B. (1997). The perception of coronals in Western Arrernte. *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, 389–392. <https://doi.org/10.21437/Eurospeech.1997-146>
- Arai, T., Behne, D., Czigler, P., & Sullivan, K. (1999). Perceptual cues to vowel quantity: Evidence from Swedish and Japanese. *Proceedings of the Swedish Phonetics Conference (Fonetik)*, 8–11.

- Australian Bureau of Statistics. (2021a). *Cultural diversity: Census*. Retrieved April 4, 2023, from <https://www.abs.gov.au/statistics/people/people-and-communities/cultural-diversity-census/latest-release>
- Australian Bureau of Statistics. (2021b). *Milingimbi: 2021 census community profiles*. Retrieved March 3, 2023, from <https://abs.gov.au/census/find-census-data/community-profiles/2021/ILOC70600701>
- Baker, B. (2008). *Word structure in Ngalakgan*. CSLI Publications.
- Baker, B. (2014). Word structure in Australian languages. In H. Koch & R. Nordlinger (Eds.), *The languages and linguistics of Australia: A comprehensive guide* (pp. 139–213). Mouton de Gruyter.
- Baroni, M., & Vanelli, L. (2000). The relationship between vowel length and consonantal voicing in Friulian. In L. Repetti (Ed.), *Phonological theory and the dialects of Italy* (pp. 13–44). John Benjamins.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Behne, D., Arai, T., Czigler, P., & Sullivan, K. (1999). Vowel duration and spectra as perceptual cues to vowel quantity: A comparison of Japanese and Swedish. *Proceedings of the 14th International Congress of Phonetic Sciences*, 857–860.
- Behne, D., Moxness, B., & Nyland, A. (1996). Acoustic-phonetic evidence of vowel quantity and quality in Norwegian. *Speech, Music and Hearing: Quarterly Progress and Status Report*, 37(2), 13–16.
- Benn, J. (2021). *The phonetics, phonology, and historical development of Guienagati Zapotec*. [Doctoral dissertation, State University of New York at Buffalo].
- Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer*. (Version 6.0.46). [Computer software].
- Bowern, C. (2023). Australian language families and linguistic classifications. In C. Bowern (Ed.), *The Oxford guide to Australian languages* (pp. lxvii–xciv). Oxford University Press.
- Broselow, E., Chen, S.-I., & Huffman, M. K. (1997). Syllable weight: Convergence of phonology and phonetics. *Phonology*, 14(1), 47–82. <https://doi.org/10.1017/S095267579700331X>
- Bundgaard-Nielsen, R. L., & Baker, B. (2015, 6–10 September). Perception of voicing in the absence of native voicing experience. *Proceedings of INTERSPEECH 2015*, 2352–2356. <https://doi.org/10.21437/Interspeech.2015-509>
- Bundgaard-Nielsen, R. L., Baker, B., Kroos, C., Harvey, M., & Best, C. T. (2012). Vowel acoustics reliably differentiate three coronal stops of Wubuy across prosodic contexts. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 3(1), 133–161. <https://doi.org/10.1515/lp-2012-0009>
- Busby, P. (1980). The distribution of phonemes in Australian Aboriginal languages. In B. Waters & P. Busby (Eds.), *Papers in Australian linguistics* (Vol. 14, pp. 73–139). Pacific Linguistics.
- Butcher, A. (2004). ‘Fortis/lenis’ revisited one more time: The aerodynamics of some oral stop contrasts in three continents. *Clinical Linguistics and Phonetics*, 18(6–8), 547–557. <https://doi.org/10.1080/02699200410001703565>

- Butcher, A. (2006). Australian Aboriginal languages: Consonant-salient phonologies and the “Place-of-Articulation Imperative.” In J. Harrington & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 187–210). Taylor and Francis.
- Butcher, A., & Harrington, J. (2003a). An acoustic and articulatory analysis of focus and the word/morpheme boundary distinction in Warlpiri. *Proceedings of the 6th International Seminar on Speech Production*, 19–24.
- Butcher, A., & Harrington, J. (2003b). An instrumental analysis of focus and juncture in Warlpiri. *Proceedings of the 15th International Congress of Phonetic Sciences*, 321–324.
- Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 32(2), 141–176. [https://doi.org/10.1016/S0095-4470\(03\)00043-3](https://doi.org/10.1016/S0095-4470(03)00043-3)
- Choi, J. D. (1992). *Phonetic underspecification and target interpolation: An acoustic study of Marshallese vowel allophony* (UCLA Working Papers in Phonetics, Vol. 82). [Doctoral dissertation, University of California, Los Angeles].
- de Jong, K., & Zawaydeh, B. (1999). Stress, duration, and intonation in Arabic word-level prosody. *Journal of Phonetics*, 27(1), 3–22. <https://doi.org/10.1006/jpho.1998.0088>
- de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics*, 30, 53–75. <https://doi.org/10.1006/jpho.2001.0151>
- Dixon, R. M. W. (2002). *Australian languages: Their nature and development*. Cambridge University Press.
- Eerola, O., Savela, J., Laaksonen, J.-P., & Aaltonen, O. (2012). The effect of duration on vowel categorization and perceptual prototypes in quantity languages. *Journal of Phonetics*, 40(2), 315–328. <https://doi.org/10.1016/j.wocn.2011.12.003>
- Elert, C.-C. (1965). *Phonologic studies of quantity in Swedish*. Almqvist & Wiksells.
- Farnetani, E., & Kori, S. (1986). Effects of syllable and word structure on segmental durations in spoken Italian. *Speech Communication*, 5(1), 17–34. [https://doi.org/10.1016/0167-6393\(86\)90027-0](https://doi.org/10.1016/0167-6393(86)90027-0)
- Fletcher, J., & Butcher, A. (2014). Sound patterns of Australian languages. In H. Koch & R. Nordlinger (Eds.), *The languages and linguistics of Australia: A comprehensive guide* (pp. 91–138). Mouton de Gruyter.
- Fletcher, J., Butcher, A., Loakes, D., & Stoakes, H. (2010). Aspects of nasal realization and the place of articulation imperative in Bininj Gun-wok. *Proceedings of the 13th Australasian International Conference on Speech Science and Technology*, 78–81.
- Fletcher, J., Stoakes, H., Loakes, D., & Singer, R. (2015). Accentual prominence and consonant lengthening and strengthening in Mawng. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Gordon, M. (2011). Stress systems. In J. Goldsmith, J. Riggle, & A. C. L. Yu (Eds.), *The handbook of phonological theory* (2nd ed., pp. 141–163). Blackwell Publishing.
- Gordon, M. (2016). *Phonological typology*. Oxford University Press.

- Gordon, M., Munro, P., & Ladefoged, P. (2000). Some phonetic structures of Chickasaw. *Anthropological Linguistics*, 42(3), 366–400.
- Graetzer, S. (2012). *An acoustic study of coarticulation: Consonant-vowel and vowel-to-vowel coarticulation in four Australian languages*. [Doctoral dissertation, University of Melbourne].
- Hajek, J., Stevens, M., & Webster, G. (2007). Vowel duration, compression and lengthening in stressed syllables in Italian. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1057–1060.
- Harvey, M., & Borowsky, T. (1999). The minimum word in Warray. *Australian Journal of Linguistics*, 19(1), 89–99. <https://doi.org/10.1080/07268609908599576>
- Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2), 253–306.
- Heath, J. (1980). *Basic materials in Ritharngu: Grammar, texts and dictionary*. Pacific Linguistics.
- Hillenbrand, J. M., & Clark, M. J. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, 108(6), 3013–3022. <https://doi.org/10.1121/1.1323463>
- Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, 72(5), 1218–1227. <https://doi.org/10.3758/APP.72.5.1218>
- Hyman, L. M. (2014). Do all languages have word accent? In H. van der Hulst (Ed.), *Word stress: Theoretical and typological issues* (pp. 56–82). Cambridge University Press.
- Jepson, K. (2019a). *Prosody, prominence and segments in Djambarrpuyngu* [Doctoral dissertation, University of Melbourne].
- Jepson, K. (2019b). The role of vowel and consonant duration in vowel length categorisation by Djambarrpuyngu listeners. *Proceedings of the 19th International Congress of Phonetic Sciences*, 305–309.
- Jepson, K., & Ennever, T. (2023). Lexical stress. In C. Bower (Ed.), *The Oxford guide to Australian languages* (pp. 145–158). Oxford University Press.
- Jepson, K., Fletcher, J., & Stoakes, H. (2021). Prosodically conditioned consonant duration in Djambarrpuyngu. *Language and Speech, Special Issue: Prosodic prominence—a cross-linguistic perspective*, 64(2), 261–290. <https://doi.org/10.1177/0023830919826607>
- Jepson, K., & Stoakes, H. (2015). Vowel duration and consonant lengthening in Djambarrpuyngu. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Jochim, M., Winkelmann, R., Jansch, K., Cassidy, S., & Harrington, J. (2023). *emuR: Main Package of the EMU Speech Database Management System*. (Version 2.4.0) [Computer software]. <https://CRAN.R-project.org/package=emuR>
- Jongman, A. (1998). Effects of vowel length and syllable structure on segment duration in Dutch. *Journal of Phonetics*, 26, 207–222. <https://doi.org/10.1006/jpho.1998.0075>
- Keating, P. (1985). Universal phonetics and the organization of grammars. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115–132). Academic Press.
- Khattab, G. (2007). A phonetic study of gemination in Lebanese Arabic. *Proceedings of the 16th International Congress of Phonetic Sciences*, 153–158.

- Khattab, G., & Al-Tamimi, J. (2014). Geminate timing in Lebanese Arabic: The relationship between phonetic timing and phonological structure. *Laboratory Phonology*, 5(2), 231–269. <https://doi.org/10.1515/lp-2014-0009>
- Kinoshita, K., Behne, D., & Arai, T. (2002, 16–20 September 2002). Duration and f0 as perceptual cues to Japanese vowel quantity. *Proceedings of the International Conference on Spoken Language Processing (ICSLP), INTERSPEECH2002*, 757–760. <https://doi.org/10.21437/ICSLP.2002-253>
- Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45, 326–347. <https://doi.org/10.1016/j.csl.2017.01.005>
- Kleber, F. (2020). Complementary length in vowel-consonant sequences: Acoustic and perceptual evidence for a sound change in progress in Bavarian German. *Journal of the International Phonetic Association*, 50(1), 1–22. <https://doi.org/10.1017/S0025100317000238>
- Koch, H., & Nordlinger, R. (2014). The languages of Australia in linguistic research: Context and issues. In H. Koch & R. Nordlinger (Eds.), *The languages and linguistics of Australia: A comprehensive guide* (pp. 3–21). Mouton De Gruyter.
- Leander, A. J. (2008). *Acoustic correlates of fortis/lenis in San Francisco Ozolotepec Zapotec*. [Doctoral dissertation, University of North Dakota].
- Lehiste, I. (1970). *Suprasegmentals*. MIT Press.
- Lehnert-LeHouillier, H. (2007). *The perception of vowel quantity: A cross-linguistic investigation*. [Doctoral dissertation, State University of New York, Buffalo].
- Lehnert-LeHouillier, H. (2010). A cross-linguistic investigation of cues to vowel length perception. *Journal of Phonetics*, 38(3), 472–482. <https://doi.org/10.1016/j.wocn.2010.05.003>
- Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. (Version 1.8.8) [Computer software]. <https://CRAN.R-project.org/package=emmeans>
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54(5), 358–368. <https://doi.org/10.1037/h0044417>
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America*, 35(11), 1773–1781. <https://doi.org/10.1121/1.1918816>
- Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of American*, 49(2B), 606–608. <https://doi.org/10.1121/1.1912396>
- Maddieson, I. (1985). Phonetic cues to syllabification. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 203–221). Academic Press.
- Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324. <https://doi.org/10.3758/s13428-011-0168-7>
- Morphy, F. (1983). Djapu, a Yolngu dialect. In B. J. Blake & R. M. W. Dixon (Eds.), *Handbook of Australian languages* (Vol. 3). The Australian National University Press.

- O'Connor, T., Perry, C. F., & Lannigan, F. J. (2009). Complications of otitis media in Indigenous and non-Indigenous children. *Medical Journal of Australia*, 191(9), S60–S64. <https://doi.org/10.5694/j.1326-5377.2009.tb02929.x>
- Odden, D. (2011). The representation of vowel length. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. Blackwell.
- Paschen, L., Fuchs, S., & Seifart, F. (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics*, 94, 1–22. <https://doi.org/10.1016/j.wocn.2022.101179>
- Penner, K. (2019). *Prosodic structure in Ixatyutla Mixtec: Evidence for the foot*. [Doctoral dissertation, University of Alberta].
- Podlipský, V. J., Skarnitzl, R., & Volín, J. (2009). High front vowels in Czech: A contrast in quantity or quality? *Proceedings of INTERSPEECH 2009*, 132–135. <https://doi.org/10.21437/Interspeech.2009-50>
- Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, 32(2), 141–152. <https://doi.org/10.3758/BF03204273>
- Prince, A. (1990). Quantitative consequences of rhythmic organization. In *Papers from the 26th regional meeting of the Chicago Linguistic Society: The parasession on the syllable in phonetics and phonology* (Vol. 2, pp. 355–398). Chicago Linguistic Society.
- R Core Team. (2023). *R: A language and environment for statistical computing*. (Version 4.3.0) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 182–206). Blackwell.
- Remijsen, B. (2014). Evidence for three-level vowel length in Ageer Dinka. In J. Caspers, Y. Chen, W. Heeren, J. Pacilly, N. O. Schiller, & E. van Zanten (Eds.), *Above and beyond the segments: Experimental linguistics and phonetics* (pp. 246–260). John Benjamins.
- Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92(1), 81–110. <https://doi.org/10.1037/0033-2909.92.1.81>
- Richardson, B., Pfister, R., & Fournier, L. R. (2020). Free-choice and forced-choice actions: Shared representations and conservation of cognitive effort. *Attention, perception and psychophysics* (82), 5126–2530. <https://doi.org/10.3758/s13414-020-01986-4>
- Rietveld, T., & Frauenfelder, U. H. (1987). The effects of syllable structure on vowel duration. *Proceedings of the 11th International Congress of Phonetic Sciences*, 28–31.
- Roengpitya, R. (2001). *A study of vowels, diphthongs, and tones in Thai*. [Doctoral dissertation, University of California, Berkeley].
- Round, E. (2023). Segment inventories. In C. Bower (Ed.), *The Oxford guide to Australian languages* (pp. 96–105). Oxford University Press.
- Ryan, K. M. (2019). *Prosodic weight: Categories and continua*. Oxford University Press.
- Sanker, C. (2019). Influence of coda stop features on perceived vowel duration. *Journal of Phonetics*, 75, 43–56. <https://doi.org/10.1016/j.wocn.2019.04.003>

- Schaeffler, F. (2005). *Phonological quantity in Swedish dialects: Typological aspects, phonetic variation and diachronic change*. [Doctoral dissertation, Umeå University].
- Schertz, J., & Clare, E. J. (2020). Phonetic cue weighting in perception and production. *WIREs Cognitive Science*, 11(2), e1521. <https://doi.org/10.1002/wcs.1521>
- Schiel, F., Draxler, C., & Harrington, J. (2011, 28–31 January). *Phonemic segmentation and labelling using the MAUS technique*. New Tools and Methods for Very-Large-Scale Phonetics Research, University of Pennsylvania.
- Siddins, J., Harrington, J., Kleber, F., & Reubold, U. (2013). The influence of accentuation and polysyllabicity on compensatory shortening in German. *Proceedings of INTERSPEECH 2013*. <https://doi.org/10.21437/Interspeech.2013-177>
- Solé, M.-J., & Ohala, J. (2010). What is and what is not under the control of the speaker: Intrinsic vowel duration. In C. C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Papers in laboratory phonology 10*. Mouton de Gruyter.
- Stoakes, H., Butcher, A., Fletcher, J., & Tabain, M. (2012). Place or manner as perceptually salient in Yolngu Matha: A closed set 4-AFC listening task in quiet and in noise. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, 225–228.
- Sundkvist, P., & Gao, M. (2015). A regional survey of the relationship between vowel and consonant duration in Shetland Scots. *Folia Linguistica*, 49(1), 57–83. <https://doi.org/10.1515/flin-2015-0002>
- Suomi, K., Meister, E., Ylitalo, R., & Meister, L. (2013). Durational patterns in Northern Estonian and Northern Finnish. *Journal of Phonetics*, 41(1), 1–16. <https://doi.org/10.1016/j.wocn.2012.09.001>
- Tabain, M., Fletcher, J., & Butcher, A. (2014). Lexical stress in Pitjantjatjara. *Journal of Phonetics*, 42, 52–66. <https://doi.org/10.1016/j.wocn.2013.11.005>
- Tabain, M., & Jukes, A. (2016). Makasar. *Journal of the International Phonetic Association*, 46(1), 99–111. <https://doi.org/10.1017/S002510031500033X>
- Tomaschek, F., Truckenbrodt, H., & Hertrich, I. (2011). Processing German vowel quantity: Categorical perception or perceptual magnet effect? *Proceedings of the 17th International Congress of Phonetic Sciences*, 2002–2005.
- Turk, A. (2012). The temporal implementation of prosodic structure. In A. C. Cohn, C. Fougeron, M. K. Huffman, & M. E. L. Renwick (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 242–253). Oxford University Press.
- van Dommelen, W. A. (1999). Auditory accounts of temporal factors in the perception of Norwegian disyllables and speech analogs. *Journal of Phonetics*, 27(1), 107–123. <https://doi.org/10.1006/jpho.1999.0087>
- van Leyden, K. (2004). *Prosodic characteristics of Orkney and Shetland dialects: An experimental approach*. LOT.
- Wang, Y., O'Shannessy, C., Davis, V., Bundgaard-Nielsen, R., Roberts, J., & Foster, D. (2024). Production and perception of stop voicing in Central Australian Aboriginal English: A cross-

- generational study. *Australian Journal of Linguistics*, 44(1), 69–98. <https://doi.org/10.1080/07268602.2024.2365167>
- Waters, B. (1979). A distinctive features approach to Djinang phonology and verb morphology. *Work Papers of SIL-AAB, Series A, Volume 4*. Summer Institute of Linguistics.
- Waters, B. (1989). *Djinang and Djinba—A grammatical and historical perspective*. Pacific Linguistics.
- Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, 93(4), 2152–2159. <https://doi.org/10.1121/1.406678>
- White, L., & Mády, K. (2008). The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. *Proceedings of Speech Prosody 2008*, 363–366. <https://doi.org/10.21437/SpeechProsody.2008-82>
- White, L., & Turk, A. (2010). English words on the Procrustean bed: Polysyllabic shortening reconsidered. *Journal of Phonetics*, 38(3), 459–471. <https://doi.org/10.1016/j.wocn.2010.05.002>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. [Computer software]. Springer-Verlag. <https://ggplot2.tidyverse.org>
- Wilkinson, M. (2012). *Djambarrpuyŋu: A Yolŋu variety of northern Australia*. LINCOM Europa.
- Winkelmann, R., Bombien, L., Scheffers, M., & Jochim, M. (2023). *wrassp: Interface to the ‘ASSP’ Library*. (Version 1.0.4) [Computer software]. <https://CRAN.R-project.org/package=wrassp>
- Wood, R. (1978). Some Yuulngu phonological patterns. In S. A. Wurm (Ed.), *Papers in Pacific linguistics, No. 11* (Vol. Series A, No. 51, pp. 53–117). Pacific Linguistics.
- Yu, A. C. L. (2008). The phonetics of quantity alternation in Washo. *Journal of Phonetics*, 36(3), 508–520. <https://doi.org/10.1016/j.wocn.2007.10.004>
- Yunupingu, H. R. (1996). Looking at language in a Yolngu way. In M. Cooke (Ed.), *Aboriginal languages in contemporary contexts: Yolŋu Matha at Galiwin’ku* (pp. 47–50). Batchelor College.

