Acoustic and durational characteristics of Anindilyakwa vowels

Rosey Billington[1], John Mansfield [2], Hywel Stoakes[2]

[1]Australian National University, [2]The University of Melbourne

rosey.billington@anu.edu.au, john.mansfield@unimelb.edu.au, hstoakes@unimelb.edu.au

Abstract

Anindilyakwa, an Aboriginal language of northern Australia, has a vowel system which is unusual among Aboriginal languages, and the subject of divergent analyses. Previous research notes extensive variation in how certain vowels are produced, with observations that non-low vowels are strongly influenced by their consonant environment. There is also extensive variation in whether certain vowels are produced, leading to suggestions that these vowels are epenthetic. This study presents a first phonetic investigation of the acoustic and durational properties of Anindilyakwa vowels, with a focus on the effects of consonant place of articulation on the realisation of non-low vowels.

Index Terms: vowels, acoustics, duration, epenthesis, Australian languages, coarticulation

1. Introduction

1.1. Anindilyakwa language

The Anindilyakwa language [1] is owned and spoken by the Warnumamalya people of Groote Eylandt in northern Australia. Until recently, a prevailing view was that Anindilyakwa was a language isolate, with striking apparent differences compared to its nearest mainland neighbours. It is now viewed as a Gunwinyguan language, most closely related to Wubuy, but having undergone significant phonological and phonotactic restructuring [1]. While there have been various studies of the sound system of Anindilyakwa [2][3][4][5], there is currently no consensus on the segmental contrasts and ways these interact with specific phonological processes. This is an area of particular interest for closer investigation, drawing on new data types.

1.2. Anindilyakwa phoneme inventory

While there are varying analyses of the Anindilyakwa consonant inventory, it is broadly typical of Australian languages, with a large number of place contrasts, and more sonorants than obstruents [6]. In the inventory shown in Table 1, the 'anterior' place of articulation merges potential apical-alveolar and lamino-dental distinctions, which are not apparent in data for the present study but which have been reported elsewhere as a marginal contrast [5]. A more typologically unusual feature of the inventory is the labialised dorsal consonants /kʷ/ and /ŋʷ/ (and according to some, additional labialised bilabials [4]). Some descriptions also propose a contrastive series of prenasalised stops (/mp, nt, n̪t̪, ɳʈ, ɲc, ŋk, ŋkʷ/ [4][5]) and labial-velar double articulations (/kp, ŋm, ŋp/ [5]), but here we treat these instead as consonant sequences based on their distributions and other phonotactic analyses [7].

For the vowel system, existing proposals include just one core phonemic vowel /a/, plus marginal /e/ and phonetic [i, ə, u] [2]; two phonemic vowels /a, ɨ/ [4]; and four phonemic vowels, either /a, e, i, u/ [3] or /a, ɛ, i, ə/ [5]. Regardless of the phonemic analysis, there is broad agreement that vowel phones in Anindilyakwa include two 'low' vowels [ɛ, a], and three 'non-low' vowels [i, ə, u]. In part, the differing analyses are due to observations that the production of the 'non-low' vowels, in terms of frontness and rounding, is conditioned to at least some extent by the place of articulation of neighbouring consonants. As observed by Heath [2] and subsequent analysts, non-low vowels are generally realised as [u] when adjacent to a labial or labialised dorsal consonant, and [i] when adjacent to an alveo-palatal consonant. Elsewhere, they are generally realised as [ə]. At the same time, various lexical exceptions have been reported, especially in instances where [i] has no conditioning palatal, such as [aɻimpa] 'stingray sp.' [3] and [mipina] 'same' [5], and some analysts report fluctuation between different qualities in the same environments and words [2]. There is also some evidence for conditioning in the low vowels, where [ɛ] frequently appears adjacent to a palatal and is viewed by some as an allophone of [a] [4], but elsewhere found to be contrastive [5]. The apparent predictability but also variation in non-low vowel quality suggests a need for closer examinination of the relationship between non-low vowel phones [i∼ə∼u] and consonant context. Proposals for the Anindilyakwa vowel system differ from the typologically more common 'triangular' 3–5 vowel systems found in many Australian languages, with /i, (e), a, (o), u/ (and often length contrasts) [8]. However, some similar analytical challenges posed by interactions between vowel quality and consonant environment can be found in Arandic languages such as Kaytetye and Central Arrernte, with arguably 'vertical' small vowel systems in which height is the primary parameter of contrast (e.g. [9], [10]).

Table 1: *Anindilyakwa consonant inventory.*

	lab.	ant.	retro.	alv.-pal.	dors.	lab. dors.
stop	p	t̪ (t)	ʈ	c	k	kʷ
nas.	m	(n̪) n	ɳ	ɲ	ŋ	ŋʷ
lat.		l̪ (l)	(ɭ)	ʎ		
trill		r				
appr.	w		ɻ	j		

1.3. Predictability of vowel occurrence

Previous researchers also note that beyond vowel quality, there is variation in whether or not non-low vowels are produced in certain contexts, and Heath [2] argues that the phones [i, ə, u] only occur as 'brief' interconsonantal epenthetic vowels, and are largely predictable in where they occur. Vowel epenthesis, which broadly relates to the surface insertion of vocalic segments and manifests in many ways crosslinguistically

[1]ISO 639-3: aio; glottocode: anin1240

[11], is uncommon among Australian languages.[2] A recent information-theoretic analysis investigating the predictability of vowel occurrence across different consonant environments in Anindilyakwa, based on both orthographic representations in a wordlist [12] and on the corpus of segmented production data used in the current study, finds that there is indeed a high level of predictability in the occurrence of 'non-low' vowels compared to 'low' vowels depending on the manner and place of adjacent consonants, for example that putative epenthesis is rare in sequences of dorsal and labial consonants and in homorganic nasal and stop sequences, but extremely common in other environments, such as in consonant sequences of equal or increasing sonority. However, patterns in the presence/absence of non-low vowels are more variable in the production data than in the wordlist data, in that the same lexeme may be produced in different ways, and in that non-low vowels may be omitted in contexts where they appear in the orthographic wordlist [7].

2. Research aims

The present study builds on the investigation of the presence compared to absence of non-low vowels in Anindilyakwa speech production data, and examines the phonetic characteristics of these vowels. In particular, we use formant measurements to investigate the claim that non-low vowel quality is largely influenced by the place-of-articulation of neighbouring consonants. Secondly, we use durational measurements to test the claim that non-low vowels are 'brief' in comparison to low vowels. The aim is to develop the understanding of how production patterns for Anindilyakwa vowels accord with impressionistic descriptions of vowel realisation in the language, and lay the groundwork for targeted research on coarticulation and natural speech processes.

3. Method

3.1. Participants

We present data collected on Groote Eylandt with seven Anindilyakwa speakers, five women and two men. The speakers' ages range from approximately 25 to 80 years old. They all speak Anindilyakwa as their main daily language, and all are multilingual in Kriol, English and other regional languages (especially Wubuy and Yolngu Matha).

3.2. Materials and procedures

Due to the varying phonological analyses (and orthographic conventions) for Anindilyakwa as well as the reported variation in vowel quality and presence, existing materials are highly inconsistent in the representation of lexemes, presenting a challenge for stimuli design. Therefore, data collection for this exploratory phonetic study (and associated study of vowel presence/absence [7]) focuses on eliciting naturalistic utterances likely to represent a broad range of segmental combinations. A set of target nouns was prepared, and elicited using picture prompts and in some cases spoken English prompts. Speakers were audio-recorded producing the target nouns in utterance frames they found meaningful, in accordance with community preferences for the conduct of this study. The same prompts were used for each speaker, meaning there is some comparability across the target nouns, but as the speakers were not required to use specific sentence frames, there is substantial diversity in how they chose to respond to the prompts. The number of utterances collected for each speaker ranged from 53–111, with the exception of one female speaker who produced 24 utterances.

3.3. Data processing and analysis

Utterances were orthographically transcribed and used to create an EMU-SDMS hierarchical database [13] [14] [15], following conversion to IPA and automated phone segmentation via the Australian Aboriginal Language model [16] in WebMAUS [17]. Segmentation was manually checked and corrected across the data. The database contains a total of 473 utterances and 5593 vowel tokens (see Table 2). Low vowel tokens (N=3674) were labelled [ɐ] and [ɛ] by the authors according to perceived vowel quality, recalling that previous phonological analyses agree that there is a contrastive open central vowel, and most likely a marginal mid front vowel. Non-low vowel tokens (N=1919) were labelled [ɪ, ə, ʊ] according to perceived vowel quality, recalling that previous analyses, while varied, posit these differing vowel qualities as largely arising from consonantal effects.

Table 2: *Number of vowel tokens in dataset.*

height	quality	# tokens
'low'	[ɐ]	3165
	[ɛ]	509
'non-low'	[ɪ]	471
	[ə]	843
	[ʊ]	605

4. Results

4.1. Formant frequency

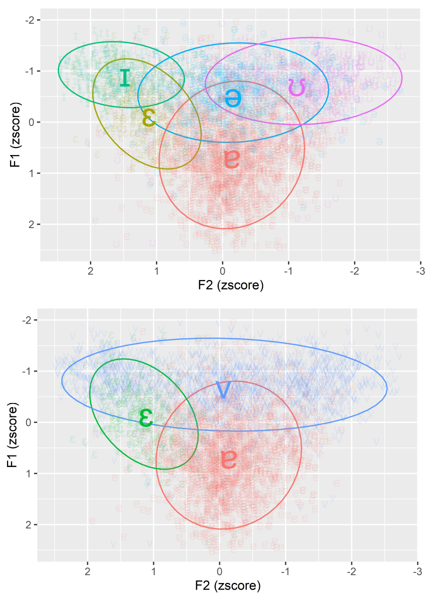

Figure 1: *F1 and F2 midpoints (Lobanov-normalised), by labelled vowel quality (top) and with non-low vowels grouped together (bottom).*

Normalised first and second frequency measures, based on vowel midpoints, are shown in Figure 1 (excluding all tokens with zero values for F1, F2 or F3). In the top panel, vowels are plotted according to labelled vowel quality, in line with

[2]The prevalence of word-final /a/ in Anindilyakwa is also argued to be related to a separate epenthesis process [5].

the five vowel phones which have been reported by all previous researchers, regardless of their phonemic analysis. While a lot of variation is apparent, distributions broadly correspond to vowel realisations which are open central [ɐ], mid front [ɛ], close front [ɪ], mid (or close) [ə/ɨ], and close back [ʊ]. With all non-low vowels grouped together as ‘v’ (bottom panel), in line with the most minimalist analyses of all non-low vowels as either epenthetic [2] or allophones of a single phoneme [4], we can see that their distributions are concentrated in the upper part of the vowel space, and that variation within this grouping, as well as across the three groupings in the top panel, is largely in the front-back dimension. For [ɐ], mean F1 is 726Hz (s.d. 199) and mean F2 is 1697Hz (s.d. 235) for the female speakers, and mean F1 is 625Hz (s.d. 165) and mean F2 is 1435Hz (s.d. 259) for the male speakers. For [ɛ], mean F1 is 540Hz (s.d. 95) and mean F2 is 2193Hz (s.d. 167) for the female speakers, and mean F1 is 516Hz (s.d. 140) and mean F2 is 1892Hz (s.d. 230) for the male speakers. For all non-low vowel phones grouped together, mean F1 is 421Hz (s.d. 85), and mean F2 is 1717Hz, with a large standard deviation of 478, for the female speakers, and mean F1 is 442Hz (s.d. 143) and mean F2 is 1504Hz, with a similarly large standard deviation of 498, for the male speakers.

Table 3: *Mean (and s.d.) F1 and F2 at midpoints for non-low vowel tokens produced by female (f) and male (m) speakers, by consonant context (*=only one token for (m) in this context).*

context	N	F1 (f)	F2 (f)	F1 (m)	F2 (m)
LBD_LBD	3	434 (30)	913 (293)	580 (-)	838 (-) *
LAB_LBD	22	411 (83)	907 (231)	464 (92)	1408 (785)
DOR_LBD	17	429 (69)	865 (179)	463 (82)	689 (124)
ANT_LBD	138	414 (58)	1264 (309)	483 (145)	1103 (445)
PAL_LBD	65	399 (45)	1627 (257)	363 (61)	1392 (313)
LBD_LAB	14	405 (56)	948 (205)	473 (-)	2208 (-) *
LAB_LAB	5	434 (23)	1087 (11)	378 (130)	1374 (407)
DOR_LAB	54	420 (98)	1320 (305)	522 (248)	1333 (485)
ANT_LAB	150	402 (64)	1650 (305)	476 (147)	1381 (410)
PAL_LAB	59	381 (59)	2056 (329)	347 (72)	1774 (370)
LAB_DOR	20	407 (128)	1139 (360)	354 (-)	2293 (-) *
DOR_DOR	3	487 (177)	1906 (132)	-	-
ANT_DOR	238	441 (97)	1832 (358)	487 (136)	1674 (444)
PAL_DOR	91	389 (54)	2411 (250)	404 (84)	2120 (224)
LBD_ANT	127	448 (93)	1178 (226)	451 (162)	1057 (342)
LAB_ANT	161	423 (75)	1447 (299)	403 (94)	1364 (434)
DOR_ANT	166	473 (101)	1738 (245)	577 (195)	1510 (373)
ANT_ANT	142	435 (100)	1835 (190)	435 (89)	1355 (294)
PAL_ANT	98	429 (65)	2204 (259)	384 (92)	1928 (275)
LBD_PAL	53	363 (58)	1624 (342)	336 (90)	1268 (256)
LAB_PAL	66	374 (38)	2189 (321)	352 (68)	1837 (380)
DOR_PAL	13	394 (52)	2292 (553)	410 (85)	2136 (97)
ANT_PAL	52	386 (53)	2202 (293)	361 (51)	2051 (148)
PAL_PAL	41	352 (67)	2406 (283)	317 (58)	2219 (110)

The realisation of non-low vowels in the front-back dimension is examined in more detail in Figure 2. All non-low vowel tokens are plotted according to consonant context, here the place of articulation of the preceding and following consonant.[3] As can be seen in Table 3, some homorganic consonant contexts correspond to very few tokens; these are shown here for completeness, but acoustic measures for these contexts are naturally not very meaningful at this stage. The differing token numbers relate to the varying presence/absence of non-low vowels depending on consonant manner and place, as discussed in Section 1.3. As can be seen, F1 and F2 patterns for non-low vowels are highly gradient; different consonantal contexts do not produce discrete groups, but rather a continuum of realisations ranging from higher F2 values in palatal environments towards the left, and lower F2 values in labio-dorsal environments towards the right (see Table 3). Statistical tests using linear mixed-effects models via `lme4` [18], with random intercepts for speaker and word, indicate that there is a significant effect of consonant environment on F2 for the non-low vowels ($p<0.001$), whether the fixed effect is the C1_C2 context, or just the following C, or just the preceding C. In Tukey-adjusted post-hoc pairwise comparisons, there are significant differences for the majority of place of articulation comparison environments. For F1, the effect of consonant environment is also significant ($p<0.001$), whether treated as C1_C2 context, following C, or preceding C, and post-hoc tests indicate significant differences particularly when the following consonant is palatal, and when the preceding consonant is palatal or labial/labialised dorsal compared to anterior or dorsal.

Inspection of lexemes with non-low vowels shows that while the overall effects of consonant context are more or less as expected based on previous descriptions, for example with [ɪ] occurring after a palatal glide and [ʊ] after a bilabial stop in [jɪpʊɻɐt̪ɐ] ‘wallaby’, there are also exceptions. Some lexical exceptions appear to be acoustically variable: for example, tokens of the word ‘good’ are variably produced as [ɛnɪŋɐpɐ], with [ɪ] in a non-palatal context, or as [ɛnəŋɐpɐ], with [ə] between the palatal and dorsal nasals. But there also appear to be more robust lexical exceptions: for the word ‘back’, tokens are quite consistently produced as [məɻɪrpɐ], with close front [ɪ] as the non-low vowel between the trill and retroflex approximant, suggesting that an unconditioned /ɪ/ vowel is part of the lexical representation of this word.

4.2. Duration

Normalised duration values can be seen in Figure 3, excluding pre-pausal vowels (which may exhibit final lengthening). Distributions indicate that the non-low vowels are typically much shorter than the low vowels. The mean duration for the low vowel [ɐ] is 103ms (s.d. 45), and for [ɛ] the mean duration is 96ms (s.d. 35). For the non-low vowel qualities grouped together, as shown in the right panel, the mean duration is 52 ms (s.d. 24), meaning that the low vowels are on average approximately twice as long as the non-low vowels. Means for the non-low vowels are similar if separated according to labelled vowel quality, as shown in the left panel; for [ɪ], 61ms (s.d. 28), for [ə], 48ms (s.d. 21), for [ʊ], 53ms (s.d. 21). Statistical tests using linear mixed-effects models with random intercepts for speaker and word indicate that there is a significant effect of vowel quality on duration (whether the fixed effect of vowel quality is non-low vs. low vowels or the five labelled vowel qualities), and Tukey-adjusted post-hoc pairwise comparisons confirm the differences between [ɐ, ɛ] and [ɪ, ə, ʊ] ($p<0.001$).

5. Discussion and conclusions

Our findings for first and second formant frequency accord with impressionistic observations that vowel production in Anindilyakwa broadly corresponds to vowel qualities [ɐ, ɛ, ɪ, ə, ʊ] [2][3][4][5]. At the same time, substantial variation is apparent, with highly gradient patterns for non-low vowels depending on the consonantal context. While substantial variation in vowel realisation, particularly due to consonantal effects,

[3]Our data included only small numbers of retroflex consonants, which have been grouped as anterior here.

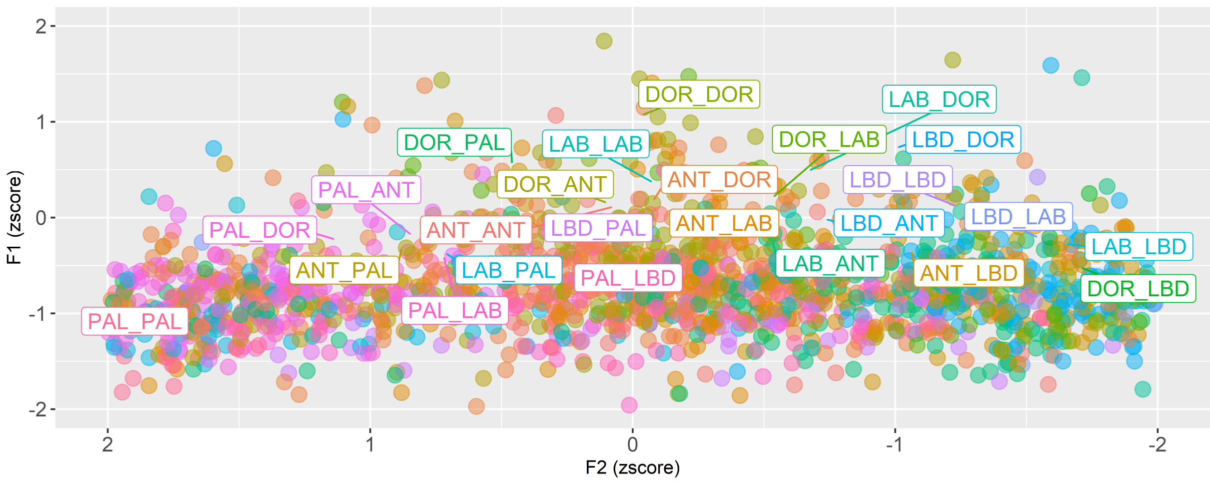

Figure 2: *F1 and F2 midpoints (Lobanov-normalised) for non-low vowel tokens, by consonant context.* LBD=*labialised dorsal,* LAB=*labial,* DOR=*dorsal,* ANT=*anterior*, PAL=palatal.

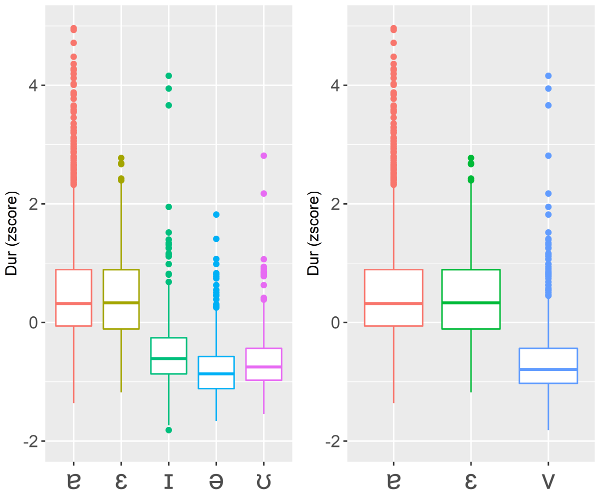

Figure 3: *Duration (Lobanov-normalised), by labelled vowel quality (left) and with non-low vowels grouped together (right).*

is common for Australian languages with small vowel systems [8], in Anindilyakwa this has posed particular analytical challenges, and the arguably epenthetic status of the non-low vowel phones is a notable typological difference. However, although Anindilyakwa has undergone dramatic phonological restructuring [1], the resulting system may still exhibit a common pattern proposed for Australian languages, the 'place of articulation imperative' [6], whereby various phonetic and phonological patterns are optimised for maintaining perceptual cues to consonant place distinctions; the non-low vocalic segments in Anindilyakwa may well have such a role (regardless of whether these vowels are interpreted as epenthetic [2] or allophones of one phoneme [4]). Some lexical items in this dataset showed [ɪ] occurring in environments other than adjacent to alveo-palatal consonants, as also reported elsewhere [3] [5], aligning with impressions of a marginal contrast.

The duration results for the low vs. non-low vowels reinforce impressions that these two sets of vowel phones pattern differently in the Anindilyakwa sound system. While open vowels, crosslinguistically, tend to be somewhat longer than close vowels, here the Anindilyakwa low vowels are twice as long as the non-low vowels, a larger difference than might be expected on the basis of biomechanical factors rather than when duration is a cue to vowel distinctions [19] [20]. These findings also match impressions reported by [2] that the non-low vowels are 'very brief' and the low vowels are 'considerably longer'; in fact, Heath estimates they are 'normally at least twice the duration'. Van Egmond [5] similarly notes that while there is no contrastive length, the low vowels are characteristically longer than the non-low vowels. Shorter durations are not uncommon for epenthetic vocalic segments [11]. The difference between the low and non-low vowels here is also larger than the reported duration difference between contrastive long and short open vowels in closely-related Wubuy, where /aː/ is 1.28 times longer than /a/ [21].

A matter that remains to be investigated is the role of prosody, and how word-level prosodic patterns interact with vowel quality and duration as well as vowel presence/absence. Previous discussions of stress in Anindilyakwa claim a predominant pattern of penultimate primary stress [4] [5], but also indicate that stress may be quantity-sensitive, in that the low vowels, while not analysed as contrastively long, attract stress due to their longer durations, and closed syllables likewise attract stress [5]. Where these factors compete, stress may vary. However, [5] notes that these are preliminary analyses based on isolated word forms. While word prosody has not yet been closely investigated for the present data, exploratory examinations indicate that pitch peaks are typically located towards the left edges of words, sometimes on syllables containing non-low vowels, suggesting that the analysis of word prosody needs to be revisited. Given that non-low vowels which arguably function as epenthetic are suggested to historically derive from centralisation of full vowels *i, *u and *a in coarticulatory and unstressed environments [1], further investigation of prosodic patterns may shed light on both synchronic and diachronic processes.

This exploratory investigation of Anindilyakwa vowel quality, building on analyses of non-low vowel presence vs. absence, bolsters the evidence that both the occurrence and the quality of non-low vowels is to a large extent predictable, based on the surrounding consonant environment. Future work will benefit both from analyses of controlled production data for vowel phones, in known epenthesis environments and with different combinations of consonant place of articulation, as well as analyses of spontaneous speech data, in order to better understand natural speech processes and individual speaker behaviour in Australian languages with small vowel systems.

6. Acknowledgements

We wish to thank all the Anindilyakwa speakers who taught us about their language, and especially the seven who spoke sentences for this study: Ramesh Lalara, Kathleen Mamarika, Katelynn Bara, Edith Mamarika, Judy Lalara, Coleen Mamarika and Joel Marawili. We also thank staff at the Groote Eylandt Language Centre – James Bednall, Brighde Collins and Carolyn Fletcher – who provided generous support and hospitality to the second author in his field visits. We gratefully acknowledge funding support from the University of Melbourne Research Unit for Indigenous Language, and the Australian Research Council grant DE180100872.

7. References

[1] van Egmond, M.-E. and Baker, B. "The genetic position of Anindilyakwa", Australian Journal of Linguistics, 40(4):492–527, 2020.

[2] Heath, J. "Draft grammatical sketch of Anindhilyagwa".

[3] Stokes, J. "Anindilyakwa phonology from phoneme to syllable". In Waters, B., editor, *Work Papers of SIL-AAIB, Series A Volume 5*, pages 139–181. SIL, Darwin, 1981.

[4] Leeding, V. J. Anindilyakwa phonology and morphology. PhD thesis, University of Sydney, 1989.

[5] van Egmond, M.-E. Enindhilyakwa phonology, morphosyntax and genetic position. PhD thesis, University of Sydney, 2012.

[6] Butcher, A. "Australian Aboriginal languages: Consonant-salient phonologies and the 'place of articulation imperative"'. In Harrington, J. M. and Tabain, M., editors, *Speech production: Models, phonetic processes and techniques*, pages 187–210. Psychology Press, New York, 2006.

[7] Mansfield, J., Billington, R., and Stoakes, H. "Vowel predictability and omission in Anindilyakwa". PsyArXiv, pre-print.

[8] Fletcher, J. and Butcher, A. "Sound patterns of Australian languages". In Koch, H. and Nordlinger, R., editors, *The languages and linguistics of Australia: A comprehensive guide*, pages 91–138. De Gruyter Mouton, Berlin, 2014.

[9] Breen, G. "The wonders of Arandic phonology". In Simpson, J., Nash, D., Laughren, M., and Alpher, B., editors, *Forty years on: Ken Hale and Australian languages*, pages 45–69. Pacific linguistics, Canberra, 2001.

[10] Tabain, M. and Breen, G. "Central vowels in Central Arrernte: A spectrographic study of a small vowel system", Journal of Phonetics, 39(1):68–84, 2011.

[11] Hall, N. "Vowel epenthesis". In Oostendorp, M., van, Ewen, C. J., Hume, E., and Rice, K., editors, *The Blackwell companion to phonology*, volume 3, pages 1576–1596. Wiley, 2011.

[12] Waddy, J. A. "Draft Anindilyakwa dictionary". 1989.

[13] Winkelmann, R., Harrington, J., and Jänsch, K. "EMU-SDMS: Advanced speech database management and analysis in R", Computer Speech & Language, 45:392–410, 2017.

[14] Winkelmann, R., Jänsch, K., Cassidy, S., and Harrington, J. emuR: Main package of the EMU Speech Database Management System, 2021. R package, Version 2.3.0.

[15] R Core Team. R: A language and environment for statistical computing, 2022. Version 4.2.0.

[16] Stoakes, H. and Schiel, F. A Pan-Australian model for MAUS, 2017.

[17] Kisler, T., Reichel, U., and Schiel, F. "Multilingual processing of speech via web services", Computer Speech & Language, 45:326–347, 2017.

[18] Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. "Fitting linear mixed-effects models using lme4", Journal of Statistical Software, 67(1):1–48, 2015.

[19] Lindblom, B. "Vowel duration and a model of lip mandible coordination", Speech Transmission Laboratory Quarterly Progress and Status Report, 8(4):1–29, 1967.

[20] Solé, M.-J. and Ohala, J. "What is and what is not under the control of the speaker: Intrinsic vowel duration". In Fougeron, C., Kühnert, B., D'Imperio, M., and Vallée, N., editors, *Papers in Laboratory Phonology 10*, pages 607–655. De Gruyter Mouton, Berlin, 2010.

[21] Bundgaard-Nielsen, R. and Baker, B. "The vowel inventory of Roper Kriol". In The Scottish Consortium for ICPhS 2015, editor, *Proceedings of the 18th International Congress of Phonetic Sciences*, pages 1–4 (paper 0885). University of Glasgow, Glasgow, 2015.