



Open Library of Humanities

An investigation of the dynamics of vowel nasalization in Arabana using machine learning of acoustic features

Christopher Carignan*, Department of Speech, Hearing and Phonetic Sciences, University College London, United Kingdom; The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Australia, c.carignan@ucl.ac.uk

Juqiang Chen*, School of Foreign Languages, Shanghai Jiao Tong University, Shanghai, China, juqiang.c@sjtu.edu.cn

Mark Harvey, School of Humanities, Creative Industries, and Social Sciences, University of Newcastle, Australia, Mark.Harvey@newcastle.edu.au

Clara Stockigt, Department of Linguistics, University of Adelaide, Australia, clara.stockigt@adelaide.edu.au

Jane Simpson, College of Arts & Social Sciences, Australian National University, Australia, jane.simpson@anu.edu.au

Sydney Strangways, Department of Speech, Hearing and Phonetic Sciences, University College London, United Kingdom

*Corresponding authors.

This paper presents exploratory research on temporally dynamic patterns of vowel nasalization from two speakers of Arabana. To derive a dynamic measure of nasality, we use gradient tree boosting algorithms to statistically learn the mapping between acoustics and vowel nasality in a speaker-specific manner. Three primary findings emerge: (1) NVN contexts exhibit nasalization throughout the entirety of the vowel interval, and we propose that a similar co-articulatory realization previously acted to resist diachronic change in this environment; (2) anticipatory vowel nasalization is nearly as extensive as carryover vowel nasalization, which is contrary to previous claims; and (3) the degree of vowel nasalization in word-initial contexts is relatively high, even in the #_C environment, suggesting that the ongoing sound change $*\eta a > \#a$ has involved the loss of the oral constriction associated with η but not the complete loss of the velum gesture.



1. Introduction

This paper presents an exploratory study of temporally dynamic patterns of vowel nasalization in Arabana, a language of northern South Australia with 15 self-identified speakers according to the 2016 Census (Australian Bureau of Statistics, 2016). Our primary research goal is to document the synchronic realization of vowel nasality in Arabana, through investigating time-varying patterns of a statistically derived acoustic metric of vowel nasalization in six different phonetic environments as produced by two speakers. We use these synchronic observations to focus additionally on which aspects of vowel nasalization may or may not have played a significant role in the phonetics and phonology of sound change in Arabana. We consider two sound changes in this study: Pre-stopping and initial dropping.

1.1. Pre-stopping

Synchronically, many Australian languages show a pattern called *pre-stopping*, where a phonological sequence consisting of a tonic vowel and a nasal or lateral /^hV(N,L)/ is realized phonetically as [^hVC_[stop](N,L)] (Dixon, 2002, p. 597; Fletcher and Butcher, 2014, pp. 109–110). While non-contrastive phonetic pre-stopping is common among Australian languages, phonologically contrastive pre-stopping is rare, i.e., languages where /^hV(N,L)/, /^hVC_[stop](N,L)/, and /^hVC_[stop]/ all contrast phonologically with one another are infrequent among Australian languages. Diachronic analyses agree that, in those cases where contrastive pre-stopping does exist, it developed from non-contrastive pre-stopping (Hercus, 1972, 1994, p. 37; Koch, 1997, pp. 276–278; McEntee and Butcher, 2021, pp. 39–44; Simpson and Hercus, 2004, pp. 188–189; Sommer, 1969, pp. 54–55).

Arabana is of particular interest for research on pre-stopping, because it is one of the few Australian languages in which pre-stopping is phonologically contrastive (Harvey et al., 2019); examples illustrating the contrast between pre-stopped and plain nasals in Arabana are provided in **Table 1**. The change from phonetically conditioned pre-stopping to phonologically contrastive pre-stopped forms is an exhausted sound change in Arabana and is no longer synchronically active in the language; the complex and unpredictable distribution of pre-stopped forms suggests that the change has indeed been inactive for some time. It should be noted that **there is no evidence for non-contrastive pre-stopping synchronically in Arabana**: Synchronic /#(C)VN/ forms do not have optional pre-stopped realizations, e.g., /amaɲi/ ‘grandmother’ can only be realized as [amaɲi] and does not have [apmaɲi] as an alternative realization. The change operated extensively across the lexicon, but there is one environment where contrastive pre-stopped forms did not develop. This is in the environment */NVN/, where there is a consistent absence of contrastive pre-stopping, i.e., */NVN/ did not develop into /^hNVC_[stop]N/ but continued as /^hNVN/. Consistent absence of contrastive pre-stopped forms in the /^hNVN/ environment

holds for all Australian languages with contrastive pre-stopping: Adnyamathanha (McEntee and Butcher, 2021, pp. 32–33), Arabana (Hercus, 1994, pp. 38–39), Arandic languages (Koch, 1997, pp. 276–278), Olkola–Oykangand (Sommer, 1969, pp. 54–55).

Plain nasal		Pre-stopped cluster	
kaman̩tali	‘separately’	kapmari	‘sibling-in-law’
jaŋi-ŋta	‘speak-Present’	waŋni-ŋta	‘cook-Present’
anari	‘this way’	watna	‘yamstick’

Table 1: Plain and pre-stopped nasal contrast examples in Arabana.

Koch (1997) suggests that the most probable explanation for the absence of contrastive pre-stopping in the /^lNVN/ environment is that the vowel was fully nasalized when the change was taking place elsewhere in the lexicon. Nasalization throughout the vowel is antithetical to a pre-stopped [^lNVC_[stop]N] form, which requires a raised velum during the oral stop consonant closure in order to generate an increase of intra-oral air pressure. Due to this aerodynamic conflict, phonetic pre-stopping would not have occurred in this environment and, subsequently, did not develop into a contrastive pre-stopped form. By contrast, in an environment such as /^l(C)VN/ that did develop contrastive pre-stopping, the hypothesis is that the vowel was realized as oral throughout and, therefore, velum height posed no aerodynamic conflict for a pre-stopped realization.

A crucial factor in this hypothesis is, therefore, the height of the velum at the right edge of the vowel interval in a /(C/N)VN/ environment, where a potential pre-stopped realization might arise. In other words, the extent and degree of *anticipatory* nasalization in Pre-Arabana would determine whether or not an aerodynamic conflict might influence the development of this particular sound change, whereas carryover nasalization would have posed no particular aerodynamic issue.¹ Most of the literature on vowel nasalization in Australian languages reports evidence for significant carryover nasalization but very limited evidence for anticipatory nasalization (Butcher, 1999; Butcher and Loakes, 2008, among others), although there are some potential discrepancies. Using dynamic nasal airflow data, Stoakes et al. (2020) show that anticipatory vowel nasalization is restricted in Bininj Kunwok both in timing and in magnitude, but much less so for pre-/ŋ/ vowels compared to vowels preceding other nasal consonants. Tabain et al. (2020) report, however, that there may be acoustic evidence for anticipatory nasalization

¹ In this respect, it may be noted that laterals did undergo pre-stopping when preceded by a nasal in Arabana, e.g., ‘nose’ *mi_la > mi_ɬla.

in Arrernte which may not be obvious to human labellers, and that due to these atypical acoustic characteristics the “impression of a lack of vowel nasalization before nasal consonants may not be entirely true.” (p. 2748). Indeed, if nasal co-articulation was primarily a carryover phenomenon in Pre-Arabana and anticipatory nasalization was limited, then pre-stopping in the */NVN/ environment would not necessarily have been impeded since the degree of nasalization would have been relatively low toward the end of the vowel interval, i.e., precisely the location where a pre-stopped realization might occur. However, the fact that pre-stopping did not occur in the */NVN/ environment suggests that the vowel may have been characterized by a high degree of nasalization throughout.

1.2. Initial dropping

Initial dropping, where the initial consonant of words is lost, is a common sound change in Australian languages (Dixon, 2002, pp. 589–602). The predominant diachronic pattern in Australian languages was that only singleton consonants occurred in word- and root-initial position historically, whereas both vowels and consonant clusters were largely absent word- and root-initially (Dixon, 2002, pp. 553–555). The diachronic reconstruction for Pre-Arabana follows the predominant pattern: The historical stage immediately preceding the synchronic Arabana that we report here contained only consonants word- and root-initially (Hercus, 1979). However, synchronic Arabana differs from Pre-Arabana in that it does have vowel-initial roots and words, and these synchronic vowel-initial forms derived from their Pre-Arabana counterparts through the process of initial dropping.

There have been **two distinct initial dropping changes** in Arabana. The first change is the reduction of homorganic semivowel-vowel sequences: */#ji/ > /#i/ and */#wu/ > /#u/. This sound change has operated exhaustively across the lexicon, i.e., there are no roots with initial /#ji/ or /#wu/ in synchronic Arabana, there are only roots with initial /#i/ or /#u/. In a vocabulary of 2142 words (Hercus, 1979), 100 (4.7%) have an initial /i/ and 54 (2.5%) have an initial /u/. Synchronic roots with initial /a/ did not arise via this same process, however, i.e., there were no changes such as */#ja/ > /#a/ or */#wa/ > /#a/. Rather, initial /a/ forms came about diachronically through the second initial dropping change, the reduction of a nasal-vowel sequence: */#ŋa/ > /#a/. This second change differs from the first in that it has not operated exhaustively across the lexicon: Synchronically, there is a contrast between */#ŋa/ and /#a/ in the lexicon with 93 (4.3%) words having an initial /ŋa/ and 119 (5.5%) words having initial /a/ (ibid). Some forms have undergone the */#ŋa/ > /#a/ change since Arabana was first recorded in the late 19th century (Helms, 1896; Todd, 1886). For example, ‘yes’ is recorded as /ŋaraji/ in the earliest sources, but as /araji/ in later sources (Hercus, 1994, p. 32). The */#ŋa/ > /#a/ change was therefore an actively spreading process when the speakers reported in the current paper acquired Arabana as a first language.

There is no quantitative data on the phonetics of initial dropping as an active process from any Australian language. The unexamined assumption is that initial dropping involves the loss of all temporal and gestural content relating to the initial segment; in the case of the second initial dropping change outlined above, this would imply a loss of all gestural features related to the nasal onset /ŋ/. In the current study, we examine forms with an initial /a/ and show that these vowels are produced with nasal co-articulation patterns that are between those associated with post-oral and post-nasal vowel contexts. In other words, /#a/ forms show nasal co-articulation patterns that are intermediate between /#Ca/ and /#Na/ forms. We interpret this as an indication of a partial maintenance of the effect of co-articulation (i.e., carryover nasalization on the vowel) despite a complete loss of the temporal interval associated with the original source of the co-articulation (i.e., the nasal consonant). In this connection, it may be noted that the reductions of the homorganic semivowel-vowel sequences, $*\#j i > \#i$ and $*\#w u > \#u$, can be analysed as involving loss of temporal content without loss of gestural content.²

The overall distribution of nasalization within vowels in Arabana is therefore a particular question of interest, since data related to synchronic productions of vowels in nasal environments may help clarify these remaining uncertainties about the absence of pre-stopped /'NVSN/ forms and the gestural nature of initial dropping involving $*\#ŋa/ > \#a/$. However, there is no quantitative data on synchronic vowel nasalization in Arabana, data which could provide a database for the reconstruction of vowel nasalization and assessment of its potential interactions with sound change. The current paper addresses this shortcoming.

2. Arabana phonological structures

Arabana has a three vowel system /i, a, u/, which is unequally distributed across the lexicon with /a/ forming a majority class. In a vocabulary of 2142 words, the vowel distribution is as follows (Hercus, 1972): /a/ 4214 (57%), /i/ 1732 (24%), /u/ 1419 (19%). Our research on Arabana required the use of real word forms, which limits the choice of items that can be analyzed due to this unequal distribution. Moreover, given the large range of environments under consideration—'CVC, 'NVN, 'NVC, 'CVN, #'VN, #'VC—it was only for the /a/ vowel that sufficient data could be collected across the entire range of environments. Given these limitations, along with a particular focus on $*\#ŋa/ > \#a/$ initial dropping, we have chosen to analyze nasalization in only the vowel /a/ for the purposes of this study. With regard to consonants in Arabana, the inventory is set out in **Table 2**.

² We can compare this with another Australian language, Iwaidja, in which consonant lenition has been shown to involve the loss of gestural content without a complete loss of temporal content (Shaw et al., 2020).

	Labial	Dental	Alveolar	Retroflex	Palatal	Velar
Stop	p	t̪	t	ɽ	c	k
Nasal	m	n̪	n	ɽ̃	ɲ	ŋ
Lateral		ɭ̪	ɭ	ɭ̃	λ	
Tap			ɾ			
Trill			r			
Approximant	w			ɻ	j	

Table 2: Arabana consonantal inventory.

In terms of morphological, prosodic, and word structures, Arabana conforms to the following general Australian patterns (Baker, 2014):

- Arabana is a suffixing language, and morphology is largely agglutinative
- Lexical morphemes are minimally disyllabic
- There is no minimum size constraint on grammatical morphemes
- Footing is trochaic
- Feet are aligned with the left edges of polysyllabic morphemes
- The head foot is the left-most foot
- The tonic vowel is always the first vowel in the word. The first vowel may be word-initial or it may be preceded by a C onset.

Arabana differs from general Australian patterns in two ways. The first difference is that vowel-initial lexical roots are not a marginal phenomenon: Out of a vocabulary of 2142 words, 275 (13%) are vowel-initial (Hercus, 1972). The second difference is in the inventory of heterosyllabic clusters.³ Arabana has an inventory of 29 heterosyllabic clusters, and in accordance with general Australian patterns, 24 (83%) conform to the Syllable Contact Law (SCL) with the sonority of C1 being greater than the sonority of C2 (Harvey et al., 2019, p. 447). There are five exceptional heterosyllabic clusters which violate the SCL: /p.m/, /t̪.n̪/, /t̪.n̪/, /t̪.n̪/, /t̪.ɭ/. These clusters are precisely those that are found in the contrastive pre-stopped forms of synchronic Arabana, which arose historically from pre-stopped allophones of plain nasals and laterals.

3. Methods

Detailed information about the methodological choices made for this study, as well as the justifications for these choices, are available in the Supplementary Materials.

³ Arabana, like most Australian languages, does not permit tautosyllabic clusters.

3.1. Datasets

The analysis reported here is based on two datasets. Dataset 1 is experimental materials recorded from Sydney Strangways, a first-language speaker of Arabana born in 1932. These materials were recorded in four field trips between July 2017 and September 2020. Dataset 2 is audio data from Laurie Stuart, another first-language speaker of Arabana born in 1913, now deceased. These audio materials were not recorded under experimental conditions. Rather, they were originally recorded to accompany a set of Arabana teaching materials already in print (Wilson, 2004). In order to provide comparison data with Dataset 1, we cross-referenced both data sets for the same vowel environments. As the audio materials from Dataset 2 were produced for pedagogical purposes, however, the prosodic environment of the data varied from token to token. Given the difference in age between the two speakers, Laurie acquired Arabana as a first language within a larger cohort of speakers than Sydney. Laurie also acquired Arabana at an earlier stage in the active spread of the change $*/\# \eta a/ > / \# a/$ across the lexicon.

3.1.1. Stimuli

We searched the Arabana lexicon for words with the following phonetic environments of the tonic vowel: 'CVC, 'NVN, 'NVC, 'CVN, $\#VC$, and $\#VN$. We shortlisted target words based on their ease-of-picturability and then trialled photos and illustrations for each target word. To prepare the visual elicitation materials, we first randomised the order of target word list, and created PDF slideshows where each slide contained a single image illustrating the target word as shown below in **Figure 1**. For each slideshow, we also produced a run sheet containing a table of the word list, their unique identifiers, and the English definition. The word list and associated metadata are provided in the Supplementary Materials.



Figure 1: Example elicitation materials for two Arabana words, $/aṇaku/$ “I don’t know” (left, $\#VN$ environment) and $/ṇamaruṇa/$ “bereaved person”, (right, $\#NVN$ environment).

3.1.2. Recording

Sydney was recorded by Margaret Carew on one field trip in July 2017 and by the fourth author in three field trips between April 2018 and September 2020. In each field trip, the recording site was a quiet location. For the 2017 field trip, a Tascam DR-100 audio recorder was used, with a Rode NT4 stereo microphone. The sample rate was set at 48 kHz with a bit depth at 24 bits per sample. For the 2018–2020 field trips, recordings were made with a Zoom H5 recorder using its internal microphone.

Prior to a recording session, the speaker reviewed the run sheet, accepting some of the words and rejecting a number of words. Rejections were mostly on the basis of incorrect language affiliation, e.g., identifying a number of words as belonging to other language varieties such as Wangkangurru or Diyari. He proposed correct Arabana replacement words for most of these items, matching these to the English meanings provided. In some cases, the speaker corrected the word to make it conform to correct Arabana. Throughout these discussions he made a number of such substitutions and additions, and also commented on meaning, clarified definitions, discussed semantic nuances, and in some cases provided various inflected forms. The result of these discussions was a revised list of target words that closely matched the items in the run sheet.

Once the review was complete, the speaker was recorded producing a number of tokens in response to the visual stimuli (range: 4–8 tokens, median: 6 tokens). Audio from the recording sessions were then annotated in ELAN (Lausberg and Sloetjes, 2009); annotation was carried out for the target word, along with replacement and key commentary provided by the speaker. Information on all target words is provided in the Supplementary Materials.

3.2. Annotation procedure

For each token in both data sets, we annotated the first vowel along with its neighboring consonants. Word-initial vowel environments were included in order to observe patterns related to initial dropping, yielding the six vowel environments previously mentioned: /CVC/, /NVN/, /NVC/, /CVN/, /#VC/, and /#VN/.⁴ Annotation was carried out via visual inspection of the audio waveform and broadband spectrogram in Praat (Boersma and Weenink, 2021). The boundaries between nasals and vowels were determined by waveform amplitude and appearance of formant structures particularly in higher frequencies. Nasals have reduced waveform amplitude and different formant positions and attenuated formant structures compared with surrounding vowels. Stops were identified through a break in formant structures, i.e., closure, followed by turbulence in the waveform and a spike in high frequencies in the spectrogram, i.e., bursts. Visual examples of annotation boundaries for vowels, nasals, and stops are shown in **Figure 2**.

⁴ It is important to note that contrastively pre-stopped forms, by definition, are those in which the vowel is followed by an oral stop. Thus, the environments /#VC/, /CVC/, and /NVC/ in our annotation convention include items which were similarly composed of the sequence /(C/N)VC/ in Pre-Arabana, as well as those which arose diachronically from phonetically pre-stopped /(C/N)VN/ forms.

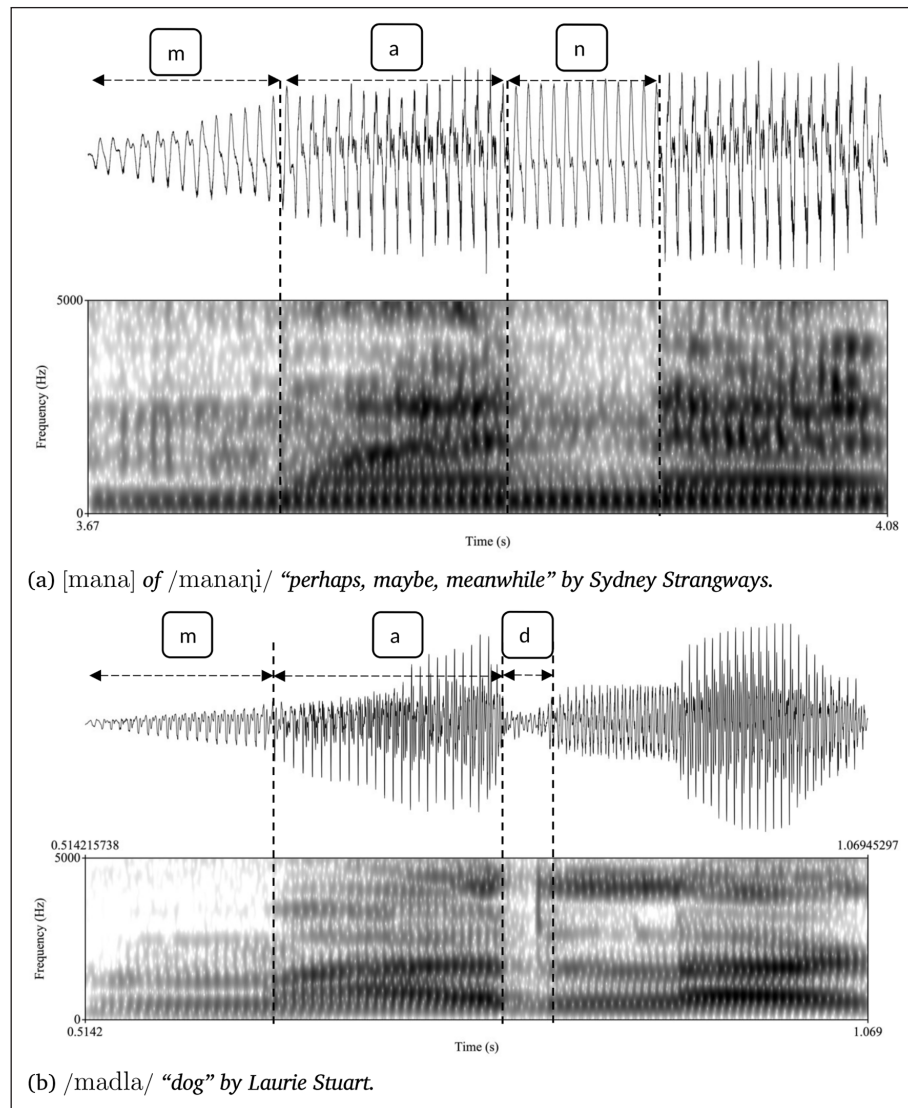


Figure 2: Partial waveforms and spectrograms of two Arabana words, 1) /manani/ “perhaps, maybe, meanwhile” by Sydney Strangways, and 2) /madla/ “dog” by Laurie Stuart, annotated for acoustic landmarks of nasals, vowels and stops.

3.3. Analysis

In the absence of physiological measures of nasality, we must infer the degree of nasalization acoustically. A traditional approach to estimating the degree of nasalization from vowel acoustics is to choose (generally) a single acoustic correlate in an implicit, top-down manner. Acoustic correlates that are commonly used as metrics of vowel nasality are A1-P0, A1-P1, or formant bandwidths (in particular, F1 bandwidth), due to the general tendency for these features to pattern with changes in nasality (Fujimura, 1961; Maeda, 1993; Feng and Castelli, 1996; Chen, 1997; Stevens, 2000). One issue with this approach is the assumption that the relationship

between the chosen acoustic correlate and the degree of nasalization is similar across languages and speakers. However, single-metric acoustic analyses of vowel nasalization have been shown to be inconsistent in their ability to estimate the degree of nasalization in a way that is robust to linguistic, inter-speaker, and intra-speaker variation (Styler, 2017), due in part to the complex acoustic interactions between the oral and nasal cavities, which can be idiosyncratic and difficult to predict accurately and consistently (Carignan, 2018). An alternative approach is to ascertain the mapping between acoustic features and the degree of nasalization in an explicit, bottom-up manner through speaker-specific statistical learning of acoustic patterns that characterize the presence of nasality in the signal. An application of this approach is described in Carignan (2021), where the generalized method is referred to as Nasalization from Acoustic Features (NAF). In that study, speaker-specific machine learning of the acoustic realization of nasalization was carried out using principal components regression. In the current study, we extend the NAF method to use gradient-boosted decision trees to learn speaker-specific patterns of the relationship between acoustic variation and the height of the velum.

3.3.1. Acoustic features of nasality

The NAF method follows a typical “shotgun” or “kitchen sink” approach to machine learning of speech, in which a wide variety of acoustic features is implemented as a set of predictor variables; some of the chosen features are assumed specifically to be acoustic correlates of vowel nasalization and others are not. By including both types of acoustic features, the aim is to capture the presence of nasality through feature engineering based on domain-specific knowledge, while at the same time incorporating a degree of flexibility in the pattern learning process of the models.

Eighteen acoustic features of nasality were measured using the Nasality Automeasure Praat (Boersma and Weenink, 2021) script created by Will Styler:⁵ The frequencies, amplitudes, and bandwidths of F1-F3; P0 and P1 amplitude; P0 prominence; A1-P0 and A1-P1, as well as their formant-compensated analogs (Styler, 2017); A3-P0; and H1-H2. The script was run in “Full-Auto” mode with defaults for all parameters. Five additional measurements of broad spectral changes were added to these features: The first four spectral moments (center of gravity, variance, skew, kurtosis) were measured using the *emuR* R package (Winkelmann et al., 2021), and a measure of nasal murmur—quantified as the ratio of low frequency (0–320 Hz) amplitude to high frequency (320–5360 Hz) amplitude (Pruthi and Espy-Wilson, 2004)—was calculated from the decibel-transformed amplitude spectrum using the *signal* R package (Ligges et al., 2021). In addition to these 23 phonetically-informed acoustic features, 14 Mel-frequency cepstral coefficients (MFCCs) were calculated using the *tuneR* R package (Ligges et al., 2022), representing a set of

⁵ Available at: https://github.com/stylerw/styler_praat_scripts/tree/master/nasality_automeasure. Details about the script and its implementations of the acoustic measures of nasality can be found in Styler (2015).

phonetically-uninformed features. Finally, the delta coefficients (i.e., sample-wise differences) of these 37 features were computed to capture abrupt temporal change in the acoustics, resulting in 74 total features to be used in modeling. Each of these 74 features were measured at 11 equidistant points throughout the target vowel, thereby normalizing time within the vowel interval.

3.3.2. Data sampling

An important consideration in most machine learning procedures is ensuring that the data used for model training is equally distributed across relevant classes, in order to avoid over-fitting of the model to only a subset of the data. For the purposes of the current study, the relevant classes are the six V environments. However, there is an imbalance in the number of observations across these classes for both speakers: A range of 171–450 observations ($SD = 123.8$) for Laurie and a range of 1071–5373 observations ($SD = 1631.7$) for Sydney. Training a model on these data runs the risk of the model primarily learning patterns in the majority class (C_N for Laurie, C_C for Sydney) while under-representing patterns in the other classes, in particular the minority class (N_N for Laurie, C_N for Sydney). For large data sets, randomly sampling the data so that all classes have a number of observations that is equal to the minority class is acceptable, since enough data still remains to allow for a properly trained model. For small data sets, such as those from extremely low-resource languages like Arabana, as much of the available data that can be retained should be retained. Therefore, instead of simply throwing away data through random sampling, re-sampling of the acoustic feature sets was carried out in the current study.

In order to re-sample the acoustic features for both speakers we used the procedure available at <https://github.com/ChristopherCarignan/multivariate-resampling>. This procedure re-samples the data set to a target number of observations, N , either under-sampling or over-sampling as appropriate. The re-sampling was carried out separately for each V environment in a given speaker's data set. For environments that contain more samples than N , N samples were randomly selected from the total set (i.e., under-sampling). For environments that contain fewer samples than N , new observations were generated (i.e., over-sampling) by creating weighted averages of feature vectors of randomly selected nearest neighbors in a multidimensional space. Over-sampling the data in this way avoids duplication of existing data samples while maintaining the general properties of the original data, including the relative distributions of each of the individual features. Further details about the re-sampling used, as well as an exploration of the effects of the over-sampling procedure on the distribution of the acoustic features, are available in the Supplementary Materials.

N was chosen according to different criteria for the two speakers, given the difference in the sizes of their respective data sets: For Sydney (whose data set was larger) N was set to the median number of samples across the six V environments (1485), and for Laurie (whose data set was

smaller) N was set to the largest number of samples across the six V environments (450), i.e., the C_N environment. Thus, after the re-sampling procedure was carried out, the data set for Sydney contained exactly 1485 vectors of the 74 acoustic features in each of the six V environments, and the data set for Laurie contained exactly 450 vectors of the 74 acoustic features in each of the six V environments. The re-sampling procedure was carried out after data associated with the boundaries of the vowel interval were excluded, as explained in the following section.

3.3.3. Data selection

One of the propositions of the NAF method is that the statistical learning of orality and nasality should be carried out using acoustic features associated with the velum in its closed (i.e., oral) and open (i.e., nasal) positions. For the current study, we accomplished this by selecting features at time points adjacent to oral consonants and nasal consonants, following the assumption that articulatory and aerodynamic requirements constrain the height of the velum at these time points. We use the time points shown in **Figure 3** to illustrate feature selection in the target vowel of the word /kaŋi/ “too much, excessively.” For this token, the time points labelled “2” and “10” (i.e., 10% and 90% of the vowel interval, respectively) were used to register oral features (point 2) and nasal features (point 10). Using this approach, oral features were selected at 10% of the vowel interval in the C_C and C_N environments and at 90% of the vowel interval in the C_C, N_C, and #_C environments, while nasal features were selected at 10% of the vowel

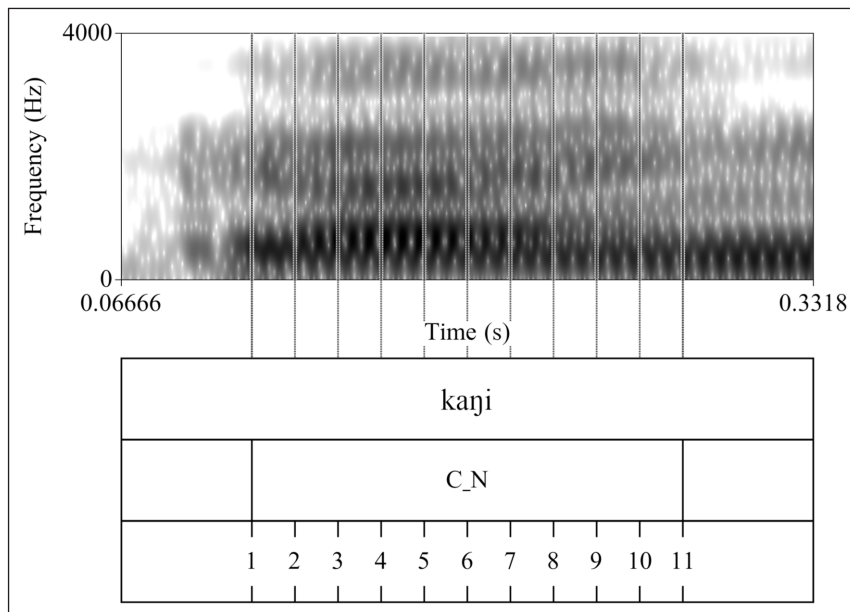


Figure 3: An example of the 11 time points from which acoustic measures were taken across each target vowel interval. Shown is a spectrogram of [a] of /kaŋi/ “too much, excessively,” with the 11 time points overlaid.

interval in the N_N and N_C environments and at 90% of the vowel interval in the N_N, C_N, and #_N environments. 80% of these samples were randomly selected for model training. The remaining 20% of these samples, the features at time point 2 in the word-initial environments, and the features at time points 3–9 in all items, were used for model predictions (i.e., generation of the final NAF measurements). Time points 1 and 11 (i.e., the absolute boundaries of the vowel interval) were excluded from all stages of the analysis with the exception of calculating delta features, in order to avoid the most extreme effects of co-articulation on the acoustic signal.

In total, for Laurie’s data, 402 samples were used for model training (80% of 604 total samples at time points 2 and 10, with subsequent exclusion of word-initial environments) and 2298 samples were used to generate NAF predictions (all remaining data, including word-initial environments and samples at time points 3–10). For Sydney’s data, 1356 samples were used for model training (80% of 2032 total samples at time points 2 and 10, with subsequent exclusion of word-initial environments) and 7554 samples were used to generate NAF predictions (all remaining data, including word-initial environments and samples at time points 3–10).

3.3.4. NAF predictions using XGBoost

Gradient boosting is a technique for supervised machine learning that builds a prediction network as an ensemble of separate weak learner models. In the case where these weak learners are decision trees, the algorithm is known as a gradient-boosted decision tree model. XGBoost is an open-source implementation of gradient tree boosting that is designed specifically for speed and performance, which has recently shown to match or even surpass deep neural networks in many applied machine learning competitions, e.g., Kaggle, especially when using data with a relative small number of variables (such as the acoustic feature set used here). In the current study, the *XGBoost* R package (Chen et al., 2022) was used to build separate gradient-boosted decision tree models for the two Arabana speakers.

As recommended in Carignan (2021), the NAF model was trained using oral and nasal observations labelled with numerical values 0 and 1, respectively, and the model was specified to minimize linear regression error, rather than as a classification problem. In this way, values on a 0–1 scale can be generated directly as response predictions from the trained model and these values can be interpreted as a linear mapping along the oral–nasal dimension: Values that are halfway between those associated with oral (0) and nasal (1) correspond to a half-degree of nasalization (0.5), and so forth, while allowing response values to surpass these bounds if appropriate, i.e., negative values are permitted for observations predicted to be *especially oral* and values > 1 are permitted for observations predicted to be *especially nasal*.

A separate XGBoost model was trained for each speaker. Default values for all hyperparameters were used, with the exception of *max_depth* (the complexity of the ensemble; used to control over-fitting), *eta* (the learning rate of each decision tree; used as a form of shrinkage),

gamma (the threshold of loss reduction required to make a decision split), and *subsample* (the proportion of the training samples used to train each decision tree). The values of these four hyper-parameters were tuned using 5-fold cross-validation (CV) of the training data in a full grid search of the hyper-parameter space. The hyper-parameter values that resulted in the lowest average CV error were used to build a subsequent 5-fold CV model to determine the optimal number of iterations to run using the tuned parameters. A final model was then built using all of the training data, the tuned hyper-parameters, and the optimal number of iterations. This model was used to predict response values for the remaining data (see Section 3.3.3), which will henceforth be referred to as “NAF values.”⁶

The resulting NAF values will be assessed both qualitatively and quantitatively in Section 4. For qualitative assessment, smoothed averages were created to observe the change in the degree of nasalization over the normalized time course of the vowel interval (10%–90%) for each of the six environments. For quantitative assessment, the NAF values were used as the dependent variable in a linear mixed-effects (LME) model with the vowel environment as a fixed effect and random intercepts by speaker;⁷ the LME model was built using the *lme4* R package (Bates et al., 2022). Tukey pair-wise contrasts were computed using the *multcomp* R package (Hothorn et al., 2022), with the α level adjusted using the Bonferroni method for maximally conservative reduction of Type I error inflation due to performing multiple comparisons.

4. Results

4.1. Qualitative results

The data smooths in **Figures 4–6** were created using the *ggplot2* R package (Wickham et al., 2021) with locally weighted smoothing (“LOESS” smoothing). These smooths display the respective means and 95% confidence intervals bands (i.e., standard error of the mean) of the six vowel environments, each of which is denoted by a separate color and line style. The pattern for C_C is as expected: The NAF values are low throughout the entire vowel interval, suggesting that the velum remains raised throughout the vowel. The pattern for N_N is precisely the opposite: The NAF values are high throughout the entire vowel interval, suggesting that the velum remains lowered throughout the whole vowel. The patterns for the N_C and C_N environments are practically mirror images of one another: The degree of nasalization (as inferred from the NAF values) decreases in a linear manner throughout the vowel interval in the N_C environment

⁶ It may be helpful to remind the reader at this point that the 10% and 90% vowel interval samples used in the model training were not also used in generating the response predictions. This is important for interpreting the resulting NAF values, which are indeed unbiased generalizations of the trained model at all time points since none of these observations were ever used in training the model in the first place.

⁷ A model that incorporated a full random effect structure was unable to converge, even using the *bobyqa* optimizer and increased iterations; see the Supplementary Materials for details.

and increases in a linear manner throughout the vowel interval in the C_N environment. This cline-like linear change is generally characteristic of phonetic (rather than phonological) vowel nasalization (Cohn, 1990), and the pattern for the C_N environment suggests not only that anticipatory nasalization does indeed occur in Arabana but that it is fairly substantial both in magnitude and in temporal extent, reaching 50% magnitude halfway through the vowel interval.

The patterns for the two word-initial vowel conditions are perhaps the most surprising. On the one hand, the degree of nasalization reaches a relatively high level at the end of the vowel interval in the #_N environment and a relatively low level at the end of the vowel interval in the #_C environment, suggesting that velum height is indeed conditioned by anticipatory co-articulatory effects in these environments. On the other hand, rather than the beginning of the vowel being oral in word-initial position (as might be expected for what is ostensibly a phonologically oral vowel), the degree of nasalization is already at a moderate level from the very start of the vowel interval in both the #_N and #_C environments, suggesting a moderate degree of vowel nasalization word-initially. In other words, the time-varying pattern of vowel nasalization for #_N is intermediate between those of C_N and N_N, while the pattern for #_C is intermediate between those of C_C and N_C. Since the 10% time points in these environments were not included in the model training, it is possible that these patterns involving the left edge of the vowel interval are simply the result of under-fitting of the model for these cases. We will explore these word-initial environments in greater detail as a post-hoc exploration in Section 4.3.

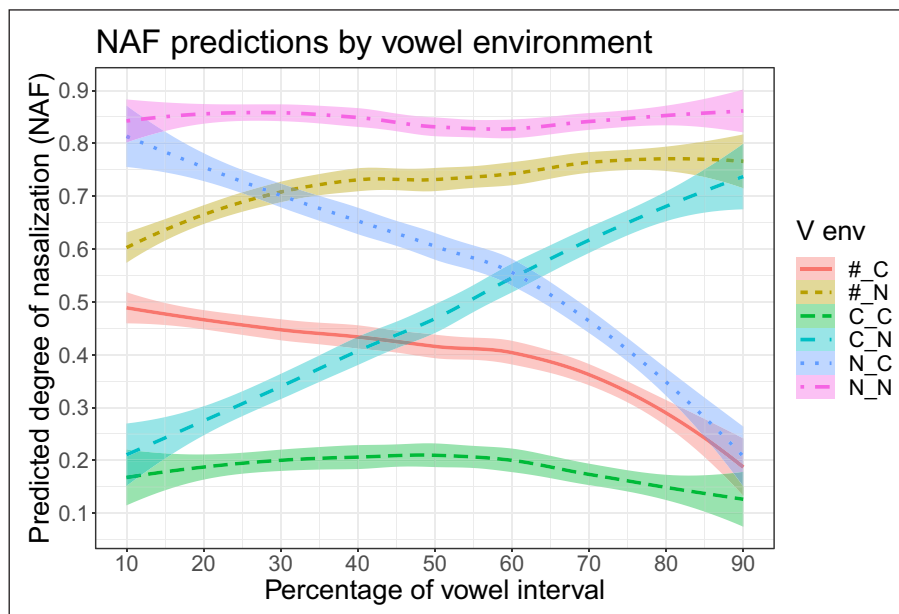


Figure 4: Predicted degree of nasalization (NAF values) over normalized time from 10% to 90% of the vowel interval. The six different phonetic environments of interest are denoted by both color and line type. Means of each phonetic environment are displayed along with 95% confidence interval bands.

4.1.1. Individual speaker results

In this section, we look at the individual results for the two speakers to better understand the degree of inter-speaker variability. **Figure 5** displays the NAF results for Sydney, and **Figure 6** display the NAF results for Laurie. In comparing these individual results to the aggregate results in **Figure 4** it is apparent that the aggregated results are more similar to the patterns produced by Sydney than to those produced by Laurie. This is to be expected, since the aggregated results are averages and Sydney contributes 3.3× more data to the aggregate than Laurie. Moreover, the larger degree of category overlap in **Figure 6** compared to **Figure 5** is likely a result of Laurie's relatively small data set and, subsequently, some degree of under-fitting in this speaker's XGBoost model. Nevertheless, the within-speaker patterns for Laurie are consistent with Sydney's patterns with the exception of one notable difference: Whereas the degree of nasalization in the #_N environment begins at a moderate level and increases throughout the vowel interval for Sydney's productions, Laurie exhibits a consistently high degree of nasalization throughout the entire vowel interval. In this manner, Laurie's realization of vowel nasality in the #_N environment is identical to that of the N_N environment. However, by and large, the time-varying patterns of nasalization are remarkably similar between the two speakers.

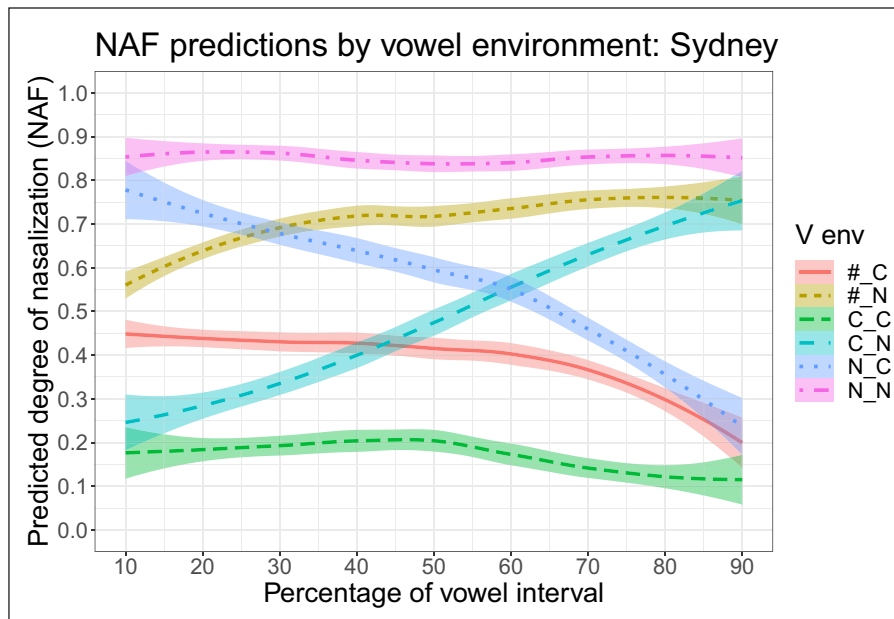


Figure 5: Predicted degree of nasalization (NAF values) over normalized time from 10% to 90% of the vowel interval, for speaker Sydney. The six different phonetic environments of interest are denoted by both color and line type. Means of each phonetic environment are displayed along with 95% confidence interval bands.

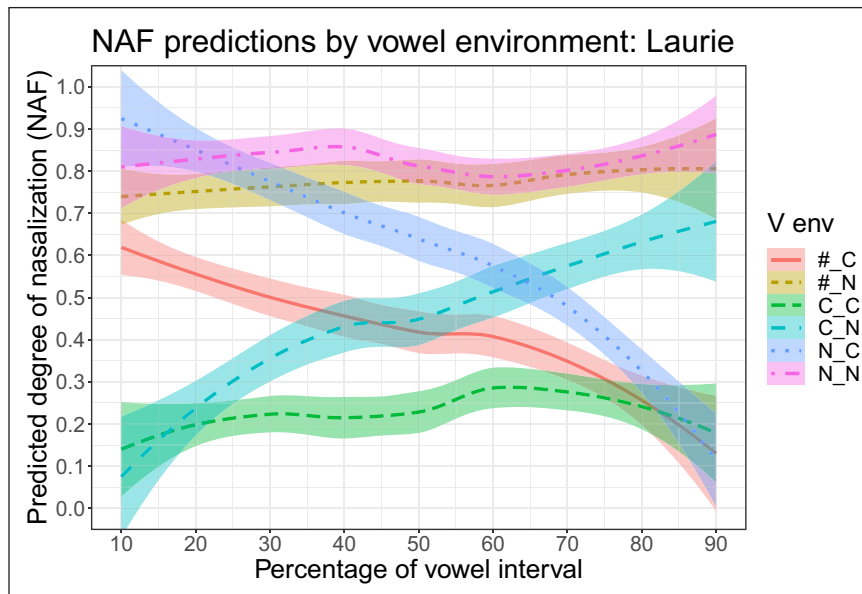


Figure 6: Predicted degree of nasalization (NAF values) over normalized time from 10% to 90% of the vowel interval, for speaker Laurie. The six different phonetic environments of interest are denoted by both color and line type. Means of each phonetic environment are displayed along with 95% confidence interval bands.

4.2. Quantitative results

The distributions of all NAF values within the target vowels—i.e., all values combined, independent of the time course of the vowel interval—are displayed as both probability density functions and horizontal box plots in **Figure 7**, with the separate vowel environments denoted by color. The distributions for the N_N and C_C environments are as expected: A left-skewed distribution centered on 1 for the N_N environment, suggesting that most of the samples display a high degree of nasalization but with a range of samples displaying moderately less; and a right-skewed distribution centered on 0 for the C_C environment, suggesting that most of the samples display a low degree of nasalization but with a range of samples displaying moderately more. The results for the N_C and C_N environments are also unsurprising. The NAF values span the entire range for both environments, due to the time-varying patterns observed above: In the N_C environment the NAF values start high, end low, and pass through the entire range of values in between, and in the C_N environment the NAF values start low, end high, and pass through the entire range of values in between. There is, however, a difference between the two environments with regard to the average of the range of NAF values: The average is slightly higher in the N_C compared to the C_N environment. Finally, the results for the word-initial vowel environments mirror the time-varying qualitative results: In word-initial position, vowels generally display a moderate degree of nasalization. However, there is a large range of NAF values in the word-initial position, especially for the #_C environment.

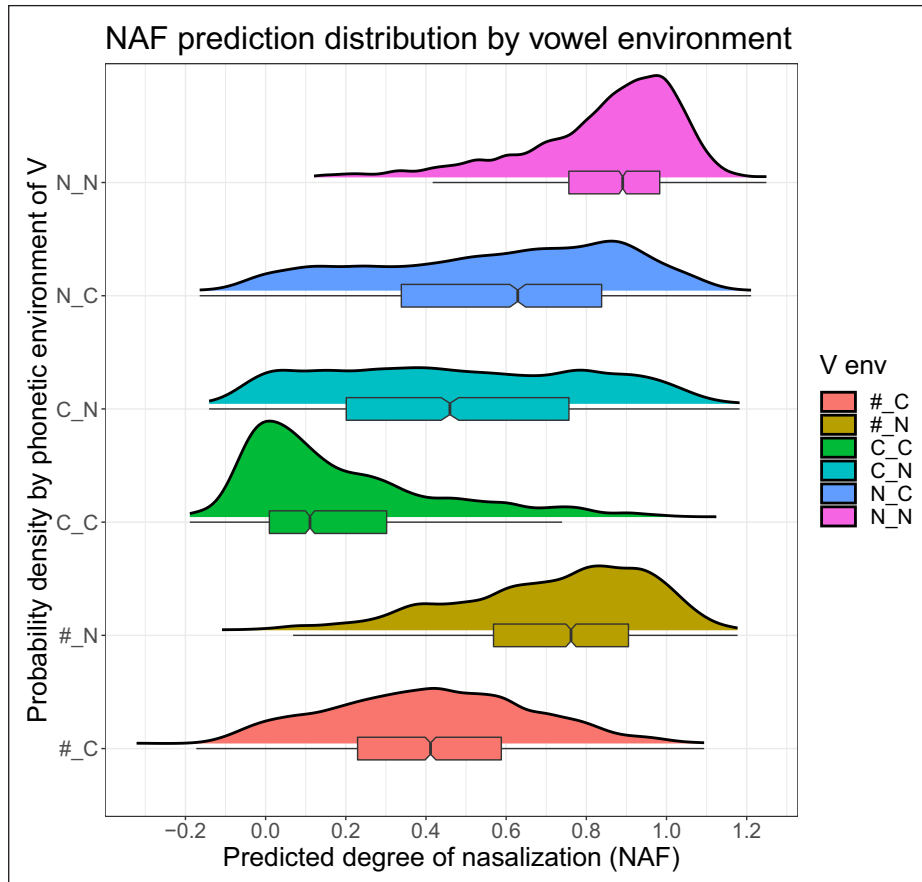


Figure 7: Probability densities and corresponding box plots for the predicted degree of nasalization (NAF values) in target vowels of the six different phonetic environments of interest, denoted by color.

The results of the Tukey contrast tests for the LME model are shown in **Table 3**. Significant differences at the Bonferroni-adjusted α level are marked with asterisks. All of the pair-wise differences are significant according to the Tukey contrast tests, and thus the ranking of the average degree of nasalization in the six vowel environments is as follows: $C_C < \#_C < C_N < N_C < \#_N < N_N$. There are two particularly important aspects of these results that we would like to note. First, the quantitative results support the qualitative observations that $\#_C$ patterns between C_C and N_C , while $\#_N$ patterns between C_N and N_N . Second, the greater degree of nasalization for N_C compared to C_N supports the claim of a greater degree of carryover compared to anticipatory vowel nasalization in Australian languages (Butcher, 1999; Butcher and Loakes, 2008; Stoakes et al., 2020), although the difference between the two environments in our data is small—indeed, this pair-wise difference yields the second smallest estimate magnitude of all 15 contrast tests. This suggests that, even though carryover nasalization is greater than anticipatory nasalization in Arabana, anticipatory nasalization is nonetheless fairly substantial.

Linear Hypotheses:	Estimate	Std. Error	z value	Pr ($> z $)
#_N – #_C == 0	0.307530	0.008772	35.057	< 2e-16 ***
C_C – #_C == 0	-0.222951	0.009029	-24.693	< 2e-16 ***
C_N – #_C == 0	0.064788	0.009006	7.194	9.46e-12 ***
N_C – #_C == 0	0.169078	0.008993	18.801	< 2e-16 ***
N_N – #_C == 0	0.437206	0.008981	48.681	< 2e-16 ***
C_C – #_N == 0	-0.530481	0.009030	-58.746	< 2e-16 ***
C_N – #_N == 0	-0.242742	0.009007	-26.949	< 2e-16 ***
N_C – #_N == 0	-0.138452	0.008994	-15.394	< 2e-16 ***
N_N – #_N == 0	0.129676	0.008982	14.437	< 2e-16 ***
C_N – C_C == 0	0.287739	0.009257	31.082	< 2e-16 ***
N_C – C_C == 0	0.392029	0.009244	42.407	< 2e-16 ***
N_N – C_C == 0	0.660157	0.009233	71.500	< 2e-16 ***
N_C – C_N == 0	0.104290	0.009222	11.308	< 2e-16 ***
N_N – C_N == 0	0.372418	0.009211	40.433	< 2e-16 ***
N_N – N_C == 0	0.268127	0.009198	29.151	< 2e-16 ***

Table 3: Tukey pair-wise contrast tests for the linear mixed effects model created to test for the effect of phonetic environment on the predicted degree of nasalization (NAF values). p-values are Bonferroni-adjusted, and significant differences at the adjusted α level are marked with asterisks: * 0.05, ** 0.01, *** 0.001.

4.3. Post-hoc analysis: Word initial vowel nasalization

The qualitative assessment in Section 4.1 suggested that word-initial vowels are moderately nasalized by both of the Arabana speakers, even in the ostensibly oral #_C environment, and the quantitative assessment in Section 4.2 confirmed these patterns. As previously noted, since the 10% time points in these word-initial environments were not included in the model training, it is possible that these patterns associated with the left edge of the vowel are simply the result of model under-fitting. In other words, since the model was never explicitly trained on acoustic features at the beginning of the vowel interval in word-initial contexts, it may be the case that the

model predictions in these contexts are simply ambiguous (i.e., neither oral nor nasal, according to the model), and therefore fall in the middle of the 0–1 scale. In this section, we investigate the patterns of nasalization that may be inferred by two of the features, both of which are commonly used as single-metric acoustic correlates of vowel nasalization: (formant-compensated) A1-P0 and F1 bandwidth.

When the nasal cavity is acoustically coupled to the oropharyngeal cavity during the production of vowel nasalization, additional poles (spectral resonances) associated with the side-branching nasal cavity are introduced to the combined acoustic transfer function (Maeda, 1993; Stevens, 2000). Chen (1997) proposed two measures to capture the relationship between the amplitudes of oral and nasal poles, with A1-P0—i.e., the difference between the amplitude of the most prominent F1 harmonic and the amplitude of the harmonic estimated to correspond to the lowest-frequency nasal pole—being the most robust measure. Correction functions based on the frequencies and bandwidths of nearby formants were also proposed to help make the measure even more robust. Thus, (formant-corrected) A1-P0 has often been used as an acoustic correlate of nasalization, with a decrease indicating an increase in the degree of nasalization.

The increased surface area of the acoustically-coupled vocal tract and the soft tissues of the nasal cavity absorb acoustic energy more than occurs in the oral cavity alone (i.e., during oral vowel production), resulting in a global reduction in formant amplitude and widening of formant bandwidths (Stevens, 2000, p. 193). The widening of formant bandwidths is predicted to be most evident in lower frequencies, due to the relatively close proximity of oral and nasal poles. Thus, F1 bandwidth has often been used as an acoustic correlate of nasalization, with an increase indicating an increase in the degree of nasalization.

In order to directly compare the results for A1-P0 and F1 bandwidth, we use here P0-A1 instead of A1-P0; thus, an increase in either P0-A1 or F1 bandwidth is inferred as an increase in the degree of nasalization. **Figure 8** displays the results for formant-compensated P0-A1 and **Figure 9** displays the results for F1 bandwidth. Both of these measures suggest that the word-initial environment is characterized by a high degree of nasalization, even in the #_C environment (which is presupposed to be both phonologically and phonetically oral). In fact, based on these two measures alone, the degree of nasalization in these word-initial contexts is suggested to be even higher than when immediately adjacent to a nasal consonant, i.e., in N_N and N_C. Although this latter observation would be phonetically puzzling if indeed accurate, these results do at least support the findings from the primary NAF analysis that word initial contexts are nasalized in Arabana. A potential explanation for this pattern is discussed in Section 5.2.

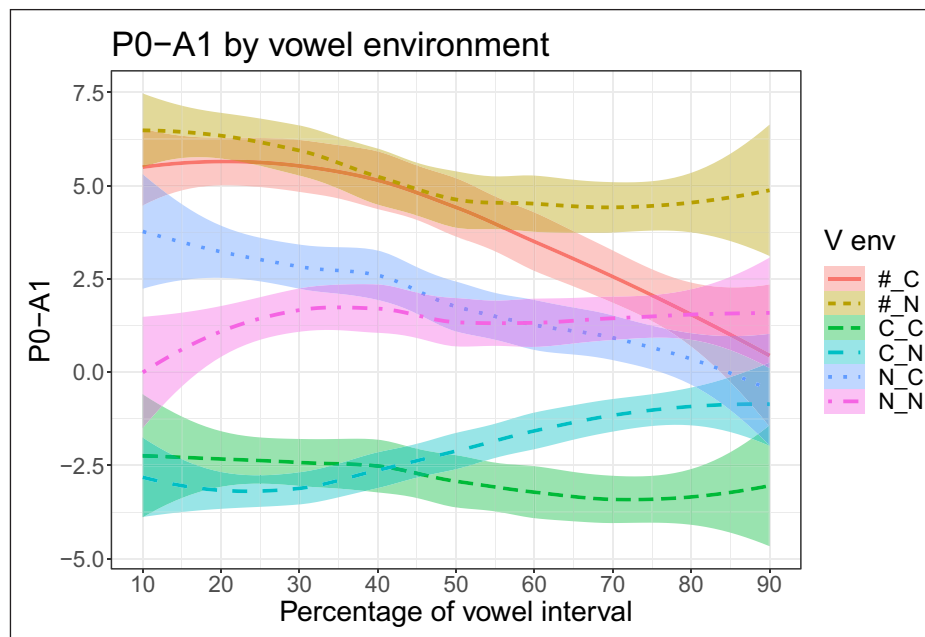


Figure 8: P0-A1 values over normalized time from 10% to 90% of the vowel interval. The six different phonetic environments of interest are denoted by both color and line type. Means of each phonetic environment are displayed along with 95% confidence interval bands.

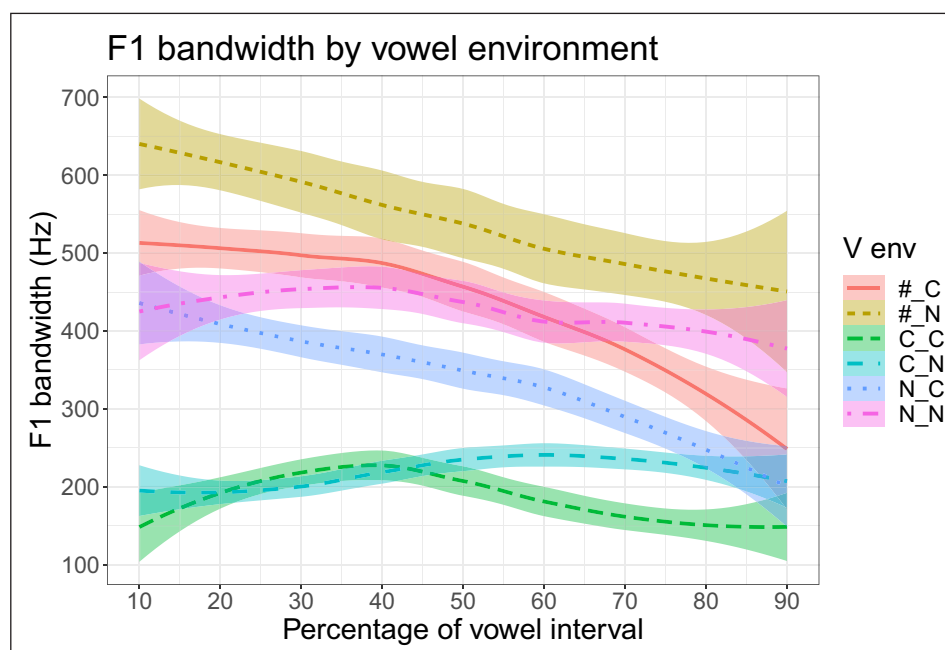


Figure 9: F1 bandwidth values over normalized time from 10% to 90% of the vowel interval. The six different phonetic environments of interest are denoted by both color and line type. Means of each phonetic environment are displayed along with 95% confidence interval bands.

5. Discussion

5.1. Pre-stopping and anticipatory vowel nasalization

In addition to evidence for substantial carryover nasalization, which is typical for Australian languages, we have also observed evidence for substantial anticipatory nasalization in Arabana, even though the overall degree of nasalization in the vowel is marginally less than observed for carryover nasalization. As noted in Section 1.1, a pre-stopped realization and anticipatory nasalization are antithetical to one another: The velum cannot be both high (for a pre-stopped production) and low (for nasalization) at the same time, i.e., at the right edge of the vowel. In order to resolve this aerodynamic conflict, either nasalization is lost and a stop can arise, or nasalization is maintained and a stop does not arise.⁸ Pre-stopping is a completed sound change process in Arabana, and our results show evidences of both of these outcomes: Vowels preceding oral consonants do not exhibit anticipatory nasalization (regardless of the historical origin of the consonant), and vowels preceding nasal consonants do exhibit anticipatory nasalization.

How, then, can we account for the one environment where pre-stopped forms did not arise historically in Arabana and where nasalization at the right edge of the vowel is synchronically preserved, i.e., /NVN/? Our results suggest that, synchronically, anticipatory nasalization is present in all pre-nasal contexts: /NVN/, /CVN/, and /#VN/. Each of these thus poses a potential aerodynamic conflict for the realization of a pre-stopped form, since the degree of nasalization is high at the right edge of the vowel. If we reconstruct similar patterns for Pre-Arabana, then the question remains: Why would pre-stopping occur in the /CVN/ and /#VN/ contexts, but not in /NVN/? We propose that the answer is to be found in the suggestion made by Koch (1997): The vowel was fully nasalized when the sound change was taking place.

Our results indicate a high degree of nasalization by both speakers throughout the entire vowel duration in the /NVN/ environment. This pattern would be at odds with co-articulatory planning that is prevalent in one direction (carryover) but restricted in another (anticipatory), as has been claimed for other Australian languages (Butcher, 1999; Butcher and Loakes, 2008; Stoakes et al., 2020). Our results suggest that anticipatory nasalization is *not* restricted in Arabana, however: Both carryover co-articulation and anticipatory co-articulation are substantial in both magnitude and temporal extent. When carryover and anticipatory planning are both active within a single vowel segment (such as in the /NVN/ environment) then consistent nasalization throughout the vowel is not at odds with co-articulatory planning but is, rather, expected. Thus, our findings of both extensive anticipatory nasalization in all pre-nasal contexts, as well as consistent nasalization throughout the vowel interval in specifically the /NVN/ context, are congruent with

⁸ This is, of course, a simplification of two outcomes of diachronic resolution to this aerodynamic constraint. Synchronically, the competing aerodynamic requirements of stop production and nasalization manifest in more gradient effects on articulatory coordination, such as a reduction in the duration and magnitude of the velum gesture, e.g., as observed in German VNC contexts (Carignan et al., 2021).

one another. We therefore reconstruct */NVN/ in Pre-Arabana as having the same realization as we have observed here, and we propose that it is precisely the consistent nasalization throughout the vowel interval that resisted the development of pre-stopping in this particular environment.

5.2. Initial dropping and partial gestural maintenance

Our results show evidence of maintenance of vowel nasalization in environments where deletion of vowel nasalization might reasonably be predicted. With the sound change */#ŋa/ > #/a/, it is predicted that the loss of the timing and gestural coordination targets for the initial */ŋ/ would be accompanied by the loss of the carryover nasalization from this consonant onto the following vowel, i.e., the loss of the nasal consonant would involve the complete loss of its gestural content. Therefore, on general grounds, the prediction is that the degree of vowel nasalization in the pair #CVN and #VN should pattern similarly, while the degree of vowel nasalization in the pair #CVC and #VC should also pattern similarly. Rather, according to the NAF metric used this study, #VC patterns intermediately between #CVC and #NVC for both speakers, while #VN patterns intermediately between #CVN and #NVN for Sydney and similarly to #NVN for Laurie.

These results suggest that the sound change */#ŋa/ > #/a/ has involved the loss of the oral constriction associated with /ŋ/—as indicated by the loss of its temporal slot in speech production—but not a complete loss of the velum gesture. The end result is thus a partial maintenance of the effect of co-articulation (i.e., carryover nasalization on the vowel) even though the original source of the co-articulation (i.e., the nasal consonant) has been lost. As discussed in Section 1.2, the change */#ŋa/ > #/a/ was actively spreading through the lexicon when the speakers reported here acquired Arabana as a first language. This may be an important factor in the maintenance of some velic gestural content, even though the lingual gestural content of /#ŋ/ has been lost. The fact that Laurie acquired Arabana 19 years earlier than Sydney and therefore earlier in the active spread of the initial dropping change may also be relevant to the fact that Laurie's #VN realizations are similar to his #NVN realizations whereas Sydney's #VN realizations are intermediate between his #CVN and #NVN realizations.

We would like to note that these patterns occur in an environment where a prosodic boundary might have some affect on the degree of nasalization. Given that the prosodic context of the target words was not controlled for in the current study, these results for the word-initial environment should be approached with an appropriate degree of circumspection. This being said, domain-initial effects have generally been shown to result in the reduction of vowel nasality, not in its enhancement (Jang et al., 2018; Cho et al., 2017).

5.3. Theoretical implications of the study

The results from this study support the notion that individual speech sounds are composed of an ensemble, or constellation of autonomous gestures—as posited by, e.g., Articulatory Phonology

(Browman and Goldstein, 1986; Saltzman and Munhall, 1989). The findings for initial dropping of $^{*}/\# \eta a/ > / \# a/$ suggest that the separate lingual and velar gestures involved in the production of $[\eta]$ were decoupled at some diachronic stage, leaving only the velar gesture as a synchronic vestige of the nasal consonant when the lingual gesture was lost. Once decoupled, the temporal interval of the velar gesture would no longer be time-locked to the timing slot of $/\eta/$ in $^{*}/\# \eta a/$. Rather, the independent velar gesture, which was once part of the gestural constellation of $/\eta/$, would then be “free” to shift to the temporal interval of the vowel, i.e., vowel nasalization. Decoupling of gestural sub-components of speech segments has been argued to play a key role in mechanisms of sound change, since such decoupling might permit independent temporal realignment (Beddor, 2009) and/or kinematic alteration (Carignan et al., 2021) of one of the decoupled gestures.

This has further implications for possible mechanisms of sound change. The development of contrastive nasal vowels arising from loss of a following nasal consonant is typologically common, e.g., Latin *vinum* [winum] ‘wine’ > Old French *vin* [vīn] > Modern French *vin* [vɛ̃]. In these cases, the preceding vowel can be reinterpreted as nasal when the oral constriction of a nasal coda is reduced or lost. Our results suggest that any mechanism that is responsible for the emergence of contrastive vowel nasality from regressive coarticulation and subsequent loss of consonant nasality, might also function in the same manner for the emergence of contrastive vowel nasality from perseverative coarticulation and subsequent loss of consonant nasality. Of course, the development of vowel nasalization in these environments is not inevitable; rather, vowel nasality has been shown to be diachronically malleable (Sampson, 1999), even over relatively short timescales (Zellou and Tamminga, 2014). The between-speaker differences that we observe for the word-initial environments might provide a reason for such malleability. The differing patterns for Laurie’s and Sydney’s word-initial productions suggest that there may be some independence associated with the temporal and magnitudinal characteristics of the decoupled velar gesture: Although both speakers exhibit extensive temporal nasalization in word-initial environments, there is a marked difference between the two speakers with regard to the magnitude of nasalization. While Laurie exhibits high magnitude of nasalization word-initially, Sydney—who is only 19 years Laurie’s junior—exhibits much less so. Thus, a diachronic weakening of the magnitude of the velar gesture, even while the temporal extent of the velar gesture is maintained, may be one possible reason why contrastive vowel nasality has not developed in Arabana, despite partial maintenance of vowel nasalization in the sound change $^{*}/\# \eta a/ > / \# a/$.

5.4. Limitations of the study

One limitation of the current study is the assumption we make about our acoustic metric of nasality, the NAF prediction values. Single acoustic metrics of nasality such as A1-P0 or F1 bandwidth assume that the relationship between the metric and the degree of nasalization holds true for all situations: All languages, all speakers, all phonetic contexts, all recordings, etc. If this relationship fails for any number of reasons, then the reliability of these metrics in accurately

capturing the degree of nasalization diminishes. Moreover, if some other, independent phonetic phenomenon affects these metrics in a systematic manner, then the resulting measurement may be artificially inflated due to the confound. For example, breathy voicing, like vowel nasality, is characterized by lower A1-P0 values and higher F1 bandwidth values (Chen, 1997; Simpson, 2012; Styler, 2015). The acoustic similarity between vowel nasality and breathy voicing may indeed be the cause for their perceptual similarity (Imatomi, 2005; Ohala and Amador, 1981) and also their co-occurrence in synchronic (Carignan, 2017; Garellek et al., 2016) and diachronic (Ohala, 1975) patterns. Ultimately, this means that phonetic effects that are independent of nasality (such as breathy voicing) can influence single-metric approaches in ways that can lead to incorrect interpretations about the degree of nasalization present in the acoustic signal.

Here, we make a different set of assumptions about the metric of nasality used in the study. We do not make the assumption that the relationship between a single given acoustic feature and the degree of nasalization holds for all situations; indeed, a key assertion of the NAF approach is that there is no single metric for which this is true, and so **a better approach is to statistically learn the relationship between a conglomerate of different features and the degree of nasalization for each use case, e.g., each individual speaker.** However, the assumptions we make instead regard the physiological state of the velum: We assume that time points immediately adjacent to oral consonants will coincide with a raised velum and time points immediately adjacent to nasal consonants will coincide with a lowered velum. When assigning the numeric value “0” to ostensibly oral contexts and “1” to ostensibly nasal contexts, we are drawing an interpretation that numeric values on this 0–1 scale correspond to lesser or greater degrees of nasality. It is important to remember what these values actually represent, however: Numeric predictions that observations are more or less similar to values that characterize the training data. Thus, we assume that NAF values near 0 correspond to “oral” when, more accurately, they correspond to “observations with acoustic features that are characteristic of a vowel adjacent to an oral consonant.” Likewise, we assume that NAF values near 1 correspond to “nasal” when, more accurately, they correspond to “observations with acoustic features that are characteristic of a vowel adjacent to a nasal consonant.”

Another limitation of the study is the small number of speakers. Exploratory research with only two speakers is merely a case study, and it is difficult to generalize from such a case study to the language as a whole. However, when working with critically endangered languages such as Arabana, it is often impossible to obtain data from more speakers or even to obtain more data from the same speakers, e.g., when one of the speakers has died since the original data were collected. It is in extreme cases like these when methodologies such as machine learning and data re-sampling allow researchers to make full use of the limited data available to them. We are heartened that, when using such methodologies, similar patterns emerge for both of the Arabana speakers included in this study. These similarities lend support to generalizing from our limited data to the Arabana language more broadly.

A final limitation of the study is the lack of physiological data (e.g., nasal airflow, nasometry, or physical estimates of velum height) to serve as ground truth for comparison with the NAF predictions, especially as a basis for validating the temporal patterns that we observe. All things being equal, more direct estimates of nasality (such as aerodynamic measures) are preferable to less direct estimates of nasality (such as acoustic measures). However, all things were not equal in the current study, namely: (1) physiological equipment were not available for use when the data collection originally took place, and (2) some of the data were collected from a speaker who has since passed away. However, in the original proof of concept of the NAF method (Carignan, 2021), nasometric measures were used as ground truth to compare with NAF predictions. The results indicated not only that the NAF method provides an accurate estimate of the *magnitude* of nasalization (an average correlation coefficient of 0.92 [SD = 0.05] with the ground truth), but also that the NAF method provides an accurate estimate of the *timing* of nasalization (a 95% Bayesian credible interval that contained the ground truth). Those findings support the use of the NAF method as a way of estimating both relative magnitude and timing of vowel nasalization in cases where physiological measurements are not practical or even possible, e.g., in the current study.

6. Conclusion

This paper has presented temporally dynamic patterns of acoustically derived measures of vowel nasalization from two speakers of Arabana, using a variation of the Nasalization from Acoustic Features method (Carignan, 2021), in which we have used gradient tree boosting algorithms to statistically learn the mapping between acoustics and vowel nasality in a speaker-specific manner. By interpreting the predictions from these models as a metric of nasality, we have observed the following ranking of the average degree of nasalization in the six vowel environments that we investigated here: C_C < #_C < C_N < N_C < #_N < N_N. Three primary findings have emerged from this exploratory research. First, NVN contexts display nasalization throughout the entirety of the vowel interval; we reconstruct a similar realization for Pre-Arabana and propose that it was consistent nasalization throughout the vowel which acted to resist the diachronic development of pre-stopping in this particular context. Second, although the average degree of anticipatory vowel nasalization is less than the average degree of carryover vowel nasalization, the overall difference between the two types of co-articulatory nasalization is relatively small, and the temporal extent is considerable for both, which is contrary to previous claims. Third, the degree of vowel nasalization in word-initial contexts is relatively high, even in the #_C environment (which is presupposed to be both phonologically and phonetically oral), suggesting that the sound change $*\#_{\eta}a > \#a$ has involved the loss of the oral constriction associated with η but not a complete loss of the velum gesture, resulting in partial maintenance of nasal co-articulation on the vowel despite the loss of the original source of the co-articulation, the nasal consonant itself.

Acknowledgements

This research was supported by two Language Documentation grants from the ARC Centre of Excellence for the Dynamics of Language (LDG992020 “Local vs Long-distance processes: Nasals and Nasalization in Arabana”; CE140100041 “Metrical Prominence and Pre-stopping in Arabana”) and by a Development Grant from the Arts & Humanities Research Council (AH/V002082/1 “Speakers, Listeners, Languages: Patterns Of Variability And Contrast In Spoken Language Dynamics”). We acknowledge with respect and gratitude the work of †Laurie Stuart and †Luise Hercus in recording Arabana. We thank Greg Wilson for access to his recordings of Arabana and for discussion of Arabana language, and Margaret Carew for her assistance in providing recorded materials and in facilitating further recording. We acknowledge that the Arabana language is the property of Arabana people.

Additional file

The additional file for this article can be found as follows:

- **Supplementary Materials.** “An investigation of the dynamics of vowel nasalization in Arabana using machine learning of acoustic features”. DOI: <https://doi.org/10.16995/labphon.9152.s1>

Competing interests

The authors have no competing interests to declare.

References

- Australian Bureau of Statistics (2016). 2016 Outer Regional Australia (SA).
- Baker, B. (2014). Word structure in Australian languages. In H. Koch & R. Nordlinger (Eds.), *The languages and linguistics of Australia: a comprehensive guide*, number 3 in *The world of linguistics* (pp. 139–213). Berlin & Boston: Walter de Gruyter. DOI: <https://doi.org/10.1515/9783110279771.139>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., Green, P., Fox, J., Bauer, A., & Krivitsky, P. N. (2022). *lme4: Linear Mixed-Effects Models using ‘Eigen’ and S4*. Computer software program available from <https://cran.r-project.org/package=lme4>
- Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, 85(4), 785–821. DOI: <https://doi.org/10.1353/lan.0.0165>
- Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer. Computer software program available from <http://www.praat.org/>
- Browman, C. P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3, 219–252. DOI: <https://doi.org/10.1017/S0952675700000658>

- Butcher, A. (1999). What speakers of Australian Aboriginal languages do with their velums and why: the phonetics of the oral/nasal contrast. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. Bailey (Eds.), *XIVth International Congress of the Phonetic Sciences* (pp. 479–482). University of California.
- Butcher, A., & Loakes, D. (2008). Enhancing the left edge: the phonetics of prestopped sonorants in Australian languages. *Journal of the Acoustical Society of America*, 124, 2527. DOI: <https://doi.org/10.1121/1.4782973>
- Carignan, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *Journal of Phonetics*, 63, 87–105. DOI: <https://doi.org/10.1016/j.wocn.2017.04.005>
- Carignan, C. (2018). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *Journal of the Acoustical Society of America*, 143(5), 2588–2601. DOI: <https://doi.org/10.1121/1.5034760>
- Carignan, C. (2021). A practical method of estimating the time-varying degree of vowel nasalization from acoustic features. *Journal of the Acoustical Society of America*, 149(2), 911–922. DOI: <https://doi.org/10.1121/10.0002925>
- Carignan, C., Coretta, S., Frahm, J., Harrington, J., Hoole, P., Joseph, A., Kunay, E., & Voit, D. (2021). Planting the seed for sound change: Evidence from real-time MRI of velum kinematics in German. *Language*, 97(2), 333–364. DOI: <https://doi.org/10.1353/lan.2021.0020>
- Chen, M. Y. (1997). Acoustic correlates of English and French nasalized vowels. *Journal of the Acoustical Society of America*, 102, 2360–2370. DOI: <https://doi.org/10.1121/1.419620>
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., & Yuan, J. (2022). *xgboost: Extreme Gradient Boosting*. Computer software program available from <https://cran.rproject.org/package=xgboost>
- Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics*, 64, 71–89. DOI: <https://doi.org/10.1016/j.wocn.2016.12.003>
- Cohn, A. C. (1990). *Phonetic and Phonological Rules of Nasalization*. PhD thesis, University of California, Los Angeles. Published as *UCLA Working Papers in Linguistics*, 76.
- Dixon, R. (2002). *Australian languages: their nature and development*. Cambridge Language Surveys. Cambridge University Press, Cambridge. DOI: <https://doi.org/10.1017/CBO9780511486869>
- Feng, G., & Castelli, E. (1996). Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization. *Journal of the Acoustical Society of America*, 99(6), 3694–3706. DOI: <https://doi.org/10.1121/1.414967>
- 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*, number 3 in *The world of linguistics* (pp. 91–138). Berlin & Boston: Walter de Gruyter. DOI: <https://doi.org/10.1515/9783110279771.91>
- Fujimura, O. (1961). Bilabial stop and nasal consonants: A motion picture study and its acoustical implications. *Journal of Speech and Hearing Research*, 4, 233–247. DOI: <https://doi.org/10.1044/jshr.0403.233>

- Garellek, M., Ritchart, A., & Kuang, J. (2016). Breathy voice during nasality: A crosslinguistic study. *Journal of Phonetics*, 59, 110–121. DOI: <https://doi.org/10.1016/j.wocn.2016.09.001>
- Harvey, M., San, N., Carew, M., Strangways, S., Simpson, J., & Stockigt, C. (2019). Prestopping in Arabana. *Australian Journal of Linguistics*, 39(4), 419–462. DOI: <https://doi.org/10.1080/07268602.2019.1643290>
- Helms, R. (1896). Anthropology – report of the Elder scientific expedition 1891. *Transactions of the Royal Society of South Australia*, 16(3), 237–332.
- Hercus, L. (1972). The pre-stopped nasal and lateral consonants of Arabana-Wangkangurru. *Anthropological Linguistics*, 14(8), 293–305.
- Hercus, L. (1979). In the margins of an Arabana-Wangkangurru dictionary: the loss of initial consonants. In S. A. Wurm (Ed.), *Australian linguistic studies*, number 54 in Series C (pp. 621–651). Canberra: Pacific Linguistics.
- Hercus, L. (1994). *A grammar of the Arabana-Wangkangurru language Lake Eyre Basin, South Australia*, volume 128 of Series C. Canberra: Pacific Linguistics.
- Hothorn, T., Bretz, F., Heiberger, R. M., Schuetzenmeister, A., & Scheibe, S. (2022). *multcomp: Simultaneous Inference in General Parametric Models*. Computer software program available from <https://cran.r-project.org/package=multcomp>
- Imatomi, S. (2005). Effects of breathy voice source on ratings of hypernasality. *The Cleft Palate – Craniofacial Journal*, 42(6), 641–648. DOI: <https://doi.org/10.1597/03-146.1>
- Jang, J., Kim, S., & Cho, T. (2018). Focus and boundary effects on coarticulatory vowel nasalization in Korean with implications for cross-linguistic similarities and differences. *The Journal of the Acoustical Society of America*, 144(1), EL33–EL39. DOI: <https://doi.org/10.1121/1.5044641>
- Koch, H. (1997). Pama-Nyungan reflexes in the Arandic languages. In D. Tryon & M. Walsh (Eds.), *Boundary rider: essays in honour of Geoffrey O'Grady* (pp. 271–302). Canberra: Pacific Linguistics.
- Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods*, 41(3), 841–849. DOI: <https://doi.org/10.3758/BRM.41.3.841>
- Ligges, U., Krey, S., Mersmann, O., Schnackenberg, S., Guenard, G., Ellis, D. P. W., Technologies, U., Preusser, A., Thieler, A., Mielke, J., Weihs, C., Ripley, B. D., & Heymann, M. (2022). *tuneR: Analysis of Music and Speech*. Computer software program available from <https://cran.r-project.org/package=tuneR>
- Ligges, U., Short, T., Kienzle, P., Schnackenberg, S., Billinghamurst, D., Borchers, H.-W., Carezia, A., Dupuis, P., Eaton, J. W., Farhi, E., Habel, K., Hornik, K., Krey, S., Lash, B., Leisch, F., Mersmann, O., Neis, P., Ruohio, J., III, J. O. S., Stewart, D., & Weingessel, A. (2021). *signal: Signal Processing*. Computer software program available from <https://cran.r-project.org/package=signal>
- Maeda, S. (1993). Acoustics of vowel nasalization and articulatory shifts in French nasal vowels. In M. K. Huffman and R. A. Krakow (Eds.), *Nasals, Nasalization, and the Velum*, volume 5 of *Phonetics and Phonology* (pp. 147–170). San Diego: Academic Press. DOI: <https://doi.org/10.1016/B978-0-12-360380-7.50010-7>

- McEntee, J. C., & Butcher, A. R. (2021). *What happens when you lick too many rocks: the complexities of Adnyamathanha phonology*. Bellevue Heights.
- Ohala, J. J. (1975). Phonetic explanations for nasal sound patterns. In C. A. Ferguson, L. M. Hyman & J. J. Ohala (Eds.), *Nasálfest: Papers from a Symposium on Nasals and Nasalization* (pp. 289–316). Palo Alto, CA: Stanford University Language Universals Project.
- Ohala, J. J., & Amador, M. (1981). Spontaneous nasalization. *The Journal of the Acoustical Society of America*, 69, S54. Abstract. DOI: <https://doi.org/10.1121/1.386212>
- Pruthi, T., & Espy-Wilson, C. Y. (2004). Acoustic parameters for automatic detection of nasal manner. *Speech Communication*, 43, 225–239. DOI: <https://doi.org/10.1016/j.specom.2004.06.001>
- Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333–382. DOI: https://doi.org/10.1207/s15326969eco0104_2
- Sampson, R. (1999). *Nasal vowel evolution in Romance*. Oxford: Oxford University Press.
- Shaw, J. A., Carignan, C., Agostini, T. G., Mailhammer, R., Harvey, M., & Derrick, D. (2020). Phonological contrast and phonetic variation: The case of velars in iwaidja. *Language*, 96(3), 578–617. DOI: <https://doi.org/10.1353/lan.2020.0042>
- Simpson, A. (2012). The first and second harmonics should not be used to measure breathiness in male and female voices. *Journal of Phonetics*, 40(3), 477–490. DOI: <https://doi.org/10.1016/j.wocn.2012.02.001>
- Simpson, J., & Hercus, L. (2004). Thura-Yura as a subgroup. In C. Bower & H. Koch (Eds.), *Australian languages: classification and the comparative method*, number 249 in Current issues in linguistic theory (pp. 179–206). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/cilt.249.12sim>
- Sommer, B. (1969). *Kunjen phonology: synchronic and diachronic*. Number 11 in Series B. Canberra: Pacific Linguistics.
- Stevens, K. N. (2000). *Acoustic phonetics*. Cambridge, MA: MIT Press. DOI: <https://doi.org/10.7551/mitpress/1072.001.0001>
- Stoakes, H. M., Fletcher, J. M., & Butcher, A. R. (2020). Nasal coarticulation in Bininj Kunwok: An aerodynamic analysis. *Journal of the International Phonetic Association*, 50(3), 305–332. DOI: <https://doi.org/10.1017/S0025100318000282>
- Styler, W. (2015). *On the acoustical and perceptual features of vowel nasality*. PhD thesis, University of Colorado.
- Styler, W. (2017). On the acoustical features of vowel nasality in English and French. *Journal of the Acoustical Society of America*, 142(4), 2469–2482. DOI: <https://doi.org/10.1121/1.5008854>
- Tabain, M., Butcher, A., Breen, G., & Beare, R. (2020). A formant study of the alveolar versus retroflex contrast in three Central Australian languages: Stop, nasal, and lateral manners of articulation. *Journal of the Acoustical Society of America*, 147(4), 2745–2765. DOI: <https://doi.org/10.1121/10.0001012>
- Todd, C. (1886). Peake telegraph station. In E. Curr (Ed.), *The Australian race* (vol. 2, pp. 10–11). Government Printer.

- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., & RStudio (2021). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. Computer software program available from <https://cran.rproject.org/package=ggplot2>
- Wilson, G. (2004). *Arabana, years R to 10: an Arabana teaching framework for reception to year ten: language revitalisation and second language learning*. Adelaide: Department of Education and Children's Services.
- Winkelmann, R., Jaensch, K., Cassidy, S., & Harrington, J. (2021). *emuR: Main Package of the EMU Speech Database Management System*. Computer software program available from <https://cran.r-project.org/package=emuR>
- Zellou, G., & Tamminga, M. (2014). Nasal coarticulation changes over time in Philadelphia English. *Journal of Phonetics*, 47, 18–35. DOI: <https://doi.org/10.1016/j.wocn.2014.09.002>

