Interspeech 2018
2-6 September 2018, Hyderabad

Formant measures of vowels adjacent to alveolar and retroflex consonants in Arrernte: stressed and unstressed position

Marija Tabain[1], *Richard Beare*[2, 3] *Andy Butcher*[4]

[1] Department of Languages and Linguistics, La Trobe University, Melbourne, Australia
[2] Department of Medicine, Monash University, Melbourne, Australia
[3] Murdoch Children's Research Institute, Melbourne, Australia
[4] Flinders University, Adelaide, Australia

m.tabain@latrobe.edu.au, richard.beare@ieee.org, andy.butcher@flinders.edu.au

Abstract

This study presents formant data for six speakers of Arrernte, a language of central Australia. The focus of the study is the (marginal) phonemic contrast between two sets of apical consonants: alveolar and retroflex. The apical contrast is studied for the stop, nasal and lateral manners of articulation: /t ʈ/, /n ɳ/ and /l ɭ/. The apical consonants are examined both in strong prosodic context (preceding a stressed vowel) and in weak prosodic context (preceding an unstressed vowel). Formant data are sampled 10 ms before the onset of the consonant, and 10 ms after the offset of the consonant. Results show no differences in F2 or F4 in the various conditions studied, and results for F1 show differences between obstruents and sonorants. F3 is lower at consonant onset than consonant offset for retroflex stops in the weak prosodic context, and to a lesser extent for retroflex stops in the strong prosodic context; it is also lower for laterals in the weak prosodic context. Other effects on F3 suggest that the apical contrast is most clearly realized for the stop manner of articulation.

Index Terms: Australian languages, apical contrast, alveolar, retroflex, stress, prosody, manner contrasts

1. Introduction

Arrernte is an Arandic language of Central Australia, spoken in and around the township of Alice Springs [1, 2, 3, 4]. In linguistic terms it is well-known for its extensive inventory of coronal consonants, as well as its minimal vowel inventory [5, 6, 7, 8]. The coronal consonants include dental, alveolar, retroflex and alveo-palatal places of articulation, at each of the stop, nasal and lateral manners of articulation. The vowel inventory contains the central vowels /a/ and /ə/, and the lexically less frequent vowel /i/.

The focus of the present study is on the alveolar versus retroflex contrast in Arrernte: /t ʈ/, /n ɳ/ and /l ɭ/. Both the alveolar and retroflex places of articulation are classed as apical, yet the contrast has been shown to be marginal, with much overlap between the two phonemic categories [9, 10]. Articulatory (i.e. electropalatography and electro-magnetic articulography) data suggest that the most prototypical retroflex articulation occurs in a weak prosodic context – that is, following the stressed vowel of Arrernte (the stressed vowel occurs after the first consonant in the word, regardless of whether the word begins with a consonant or a vowel [11]). The prototypical retroflex articulation involves a closure in the post-alveolar or pre-palatal region, followed by a ballistic forward movement of the tongue during consonant closure: this articulation is most consistently found in the weak prosodic position (i.e. following the stressed vowel). However, other intermediate articulations are observed as well: closure may be at a point between an alveolar place of articulation and a post-alveolar place of articulation; and there may or may not be a forward movement of the tongue during closure. By contrast, the most prototypical alveolar articulation – with both closure and release of the consonant occurring at the alveolar place of articulation – is more likely to occur in the strong prosodic position, i.e. preceding the stressed vowel.

Although the above description suggests a complementary distribution of the alveolar versus retroflex contrast, there are in fact several minimal pairs in the language (such as *ateme* /atəm/ 'split/burst' versus *arteme* /aʈəm/ 'build/cover'), for each of the manners of articulation, suggesting that the contrast is indeed phonemic. However, even minimal pairs have been shown to display extensive variation in articulation between repetitions.

The above picture is complicated by the fact that the apical contrast is neutralized in word-initial position, as is typical of such a contrast in the world's languages [12, 13, 14] – but see [15]. Moreover, one salient aspect of retroflex production in Arrernte is the possible pre-palatalization of the (usually stressed) retroflex, such that the word *arteme* may be pronounced [ajʈəm] instead of [aʈəm], for instance. Whether a retroflex /ʈ ɳ ɭ/ should be used in these palatalized transcriptions, or an alveolar /t n l/ instead, is not entirely clear, and relates to the broader issues outlined above. This realization appears to be a clear speech variant, and speakers freely alternate between a palatalized and a non-palatalized production even in citation repetitions of the same word.

In acoustic terms, subtle contrasts have been found between the alveolar and retroflex phonemes. The spectral centre of gravity of the stop burst following the release of the consonant has been shown to be slightly lower for retroflexes than for alveolars, presumably reflecting a larger front resonating cavity for the retroflex at the moment of stop release (i.e. a more posterior release for the retroflex as compared to the alveolar [16]). When analyzed according to prosodic context [17], there is some evidence that the spectral centre of gravity is higher for the alveolar in a strong prosodic context than in a weak prosodic context, suggesting a more forward release of the alveolar preceding a stressed vowel.

As regards the other manners of articulation, retroflex laterals have been shown to have a lower F3 at consonant midpoint than alveolar laterals in Arrernte [18]. There is some

10.21437/Interspeech.2018-1126

evidence that F4 may also be a little lower for retroflexes, and F2 a little higher and closer to F3. For nasals, by contrast, there are no significant differences in nasal formants or antiformants between alveolars and retroflexes [19], although there is some evidence of a higher F2 for Arrernte retroflexes. (Interestingly, although there was minimal difference in spectral centre of gravity between alveolars and retroflexes for the lateral and nasal manners of articulation, there was instead a significant difference in the second spectral moment, (variance) with retroflex nasals and laterals having lower variance than alveolar nasals and laterals – this spectral moments result for Arrernte was contra results for other Central Australian languages).

Finally, in terms of Arrernte vowel formants, there is no difference in F2 between alveolars and retroflexes in either the preceding vowel, nor in the following vowel [20]. However, there is evidence of effects on F1, with a tendency for F1 to be lower when adjacent to retroflexes (although this is not consistent across speakers). By contrast, F3 is consistently lower in the vowel preceding retroflexes, and it may also be slightly lower in the vowel following retroflexes.

The purpose of this study is to determine whether the articulatory patterns described above are in any way reflected in the acoustic formant data. The articulatory data suggested that the most prototypical retroflex articulation was in the weak prosodic position, while the most prototypical alveolar articulation was in the strong prosodic position. Given that these articulatory differences involve differences both in point of consonant closure and consonant release, we examine the vowel formants near the start and endpoints of the alveolar and retroflex consonants. We treat the different manners of articulation separately, since articulatory data suggest that the contrast between alveolars and retroflexes may be most clear for the nasal manner of articulation, and least clear for the lateral manner of articulation.

2. Method

2.1. Speakers and Recordings

Data are presented for 6 speakers of Arrernte, five female and one male. Four of the speakers (including the male speaker) were recorded to reel-to-reel tape at the Institute for Aboriginal Development in Alice Springs in 1990. The other two speakers were recorded direct to computer in a professional-grade recording studio at Macquarie University in Sydney in 2004, under the supervision of a professional recording technician. The digital recordings were acquired in WAV format at a sample rate of 44.1 kHz with 32 bits per sample; the analog recordings were converted to WAV format at a sample rate of 22.05 kHz with 32 bits per sample. All of the speakers who participated in these recordings were senior figures in language maintenance in their community.

Stimuli consisted of single words which were repeated by the speaker three times in a row, without carrier phrase. The list of words was designed to illustrate the sounds of each language in different positions in the Word (i.e. word-initial, word-medial, and, where permitted, word-final position), and in different vowel contexts. Dis-fluent tokens were discarded.

2.2. Analyses

Intervocalic stop, nasal and lateral tokens were extracted for analysis. Only central vowel contexts were used – that is /a/ and /ə/. Tokens were coded according to whether the following vowel was lexically stressed or not: if the following vowel was stressed, the token was labelled "Strong", and if the following vowel was unstressed, the token was labelled "Weak".

All analyses were conducted using the EmuR speech analysis software [21, 22], interfaced with the R statistical software version 3.4.3 [23, 24]. Vowel formants were extracted based on the Snack pitch and formant tool in the legacy Emu speech software. Formants were sampled 10 ms before the end of the preceding vowel, and 10 ms after the start of the following vowel. This strategy was chosen because there are sudden changes in cavity affiliation of resonances at the boundary between a vowel and a nasal consonant, or between a vowel and a lateral consonant (c.f. [25]). By sampling 10 ms before/after the consonant boundary, it is expected that only vowel formants will be sampled, rather than consonant formants. In addition, a formant sample point 10 ms into the following vowel is towards the end of the stop burst/aspiration for apical stops in Arrernte (with burst/duration values for this class of consonants at around 10-15 ms – [17]); this later sampling strategy is thus more likely to provide an accurate estimate of the vowel formant following release of an oral stop. It is of course the case that by sampling at a point that is not perfectly located at the acoustic start or endpoint of the vowel, the most extreme value of any formant movement will not be captured – this is a recognized limitation of the strategy designed to minimize formant-tracking errors.

It should be noted that as part of data exploration for this study, we converted the formant values to Bark, and also calculated the difference between F2 and F3 as a measure of spectral prominence in the mid-frequency region of the spectrum. However, the Bark formant values showed the same pattern as the corresponding formant values in Hertz; and the F3-F2 values showed the same pattern as the F3 values on their own (as will be seen below). For this reason, only the Hertz values of formants 1-4 will be presented below.

2.3. Number of Tokens

Table 1. *Number of tokens.*

	Alveolar Strong	Alveolar Weak	Retroflex Strong	Retroflex Weak
Stop	162	72	174	399
Nasal	246	114	105	369
Lateral	210	207	78	159
TOTAL	**618**	**393**	**357**	**927**

Table 1 shows the number of stop, nasal and lateral tokens extracted from the database of six speakers. The tokens are divided according to whether they are alveolar or retroflex; strong or weak. There are thus four categories: Alveolar Strong, Alveolar Weak, Retroflex Strong, and Retroflex Weak. It can be seen that by far the largest number of tokens overall is Retroflex Weak. However, while the Retroflex Weak tokens dominate the stop and nasal manners of articulation, this is not true of the lateral manner of articulation, where alveolar tokens (strong and weak) are more common. The next largest category overall is Alveolar Strong (although there are fewer tokens in this category for the stop manner of articulation compared to the nasal and lateral manners). Finally, the smallest numbers of tokens are in the Alveolar Weak and Retroflex Strong categories – in particular, there are fewer than 100 tokens of Alveolar Weak stops, and fewer than 100 tokens of Retroflex Strong laterals. These patterns likely reflect lexical frequencies in the language, although there are no electronic databases of Arrernte which can confirm whether or not this is true.

3. Results

Table 2 shows duration of consonant closure (in milliseconds) for the apical consonants in this study: Alveolar Strong, Alveolar Weak, Retroflex Strong and Retroflex Weak. Data are collapsed across speakers, but presented separately for each manner of articulation.

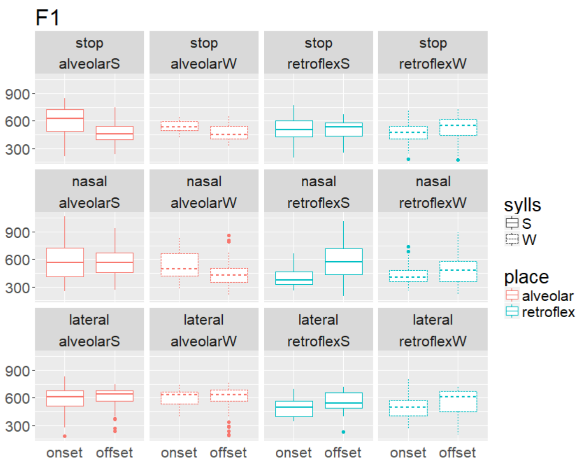
It can be seen that broadly speaking, duration values are very similar for the three different manners of articulation. Durations are longer for Strong tokens than for Weak tokens, with values tending around 180 ms for Alveolar Strong, and around 150-160 ms for Retroflex Strong. By contrast, Alveolar Weak durations tend around 90-100 ms, and Retroflex Weak durations tend around 100-120 ms. It can thus be seen that by sampling formants at 10 ms into the vowel, this is at most sampling at the first or last 10% of vowel duration (and in most instances at less than 10%).

These duration values are perhaps surprisingly long compared to languages like English, where stop closure duration in weak syllables in particular can be very short. This likely reflects the importance of the consonant in Arrernte phonology (c.f. [8]).

Table 2. *Consonant closure duration (in milliseconds).*

	Alveolar Strong	Alveolar Weak	Retroflex Strong	Retroflex Weak
Stop	177	85	165	95
Nasal	177	98	156	110
Lateral	179	96	146	119
Mean	**178.1**	**94.9**	**158.6**	**105.5**

Figure 1 shows boxplots of formant values averaged across all six speakers in this study. Data are sampled at the onset and at the offset. Starting first with F1 and the alveolars, it can be seen that there is no difference between onset and offset values for the nasal and lateral manners of articulation. However, for the stop manner of articulation, there is some evidence that F1 is higher at consonant onset than at consonant offset. An examination of individual speaker plots showed that for 3/6 speakers, F1 was higher at onset than at offset for the Alveolar Strong category, and for 4/6 speakers for the alveolar Weak category.



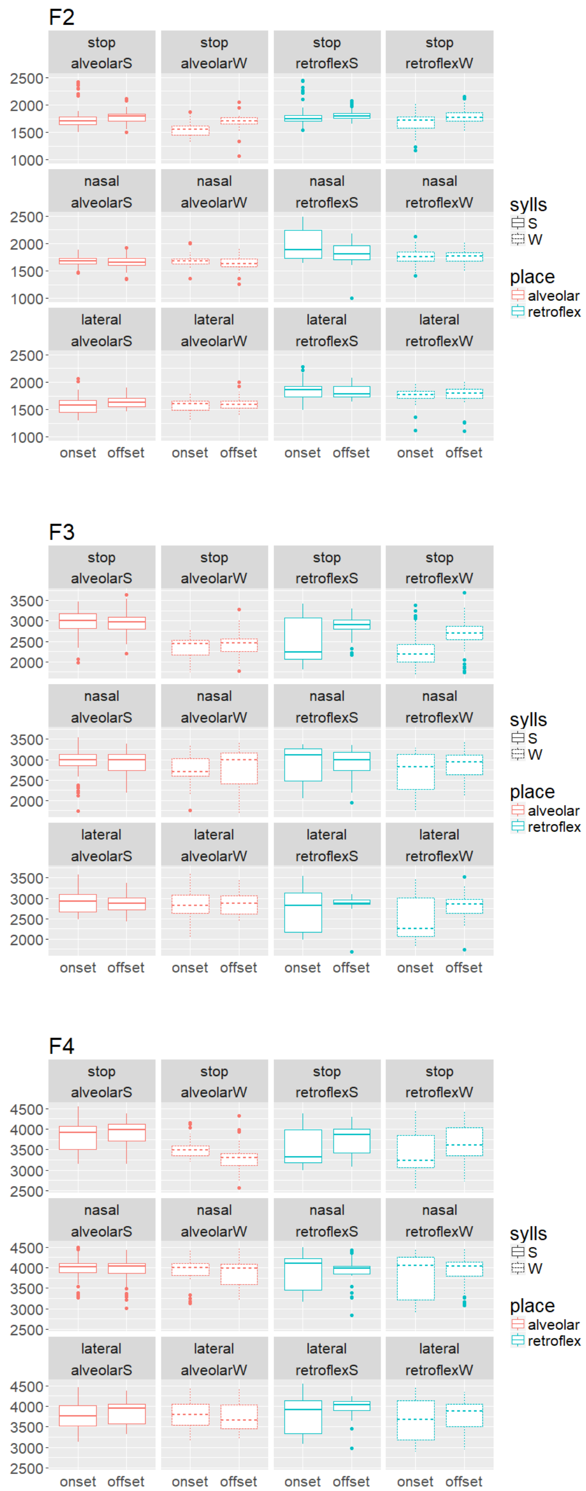


Figure 2. *Formants 1-4 averaged across all six speakers.*

The reverse pattern holds for the retroflex consonants, with F1 values being *lower* at consonant onset than at offset – this is clearer for the nasal and lateral manners of articulation (where either 4/6 or 5/6 speakers showed this pattern for the four lateral and nasal retroflex categories shown on the plot) than for the stop manner of articulation (where 3/6 speakers showed this pattern for the two stop retroflex categories shown on the plot).

Turning now to F2, there appear to be no significant differences between onset and offset for any of the four apical categories, for any of the manners of articulation, with the exception of the Alveolar Weak stop (where F2 at onset is *lower* than at offset). Similar observations can be made about F4 as well, with no significant differences between onset and offset for any of the four apical categories, for any of the manners of articulation, with the exception of the Alveolar Weak stop (where F2 at onset is *higher* than at offset).

Finally we consider F3. The boxplots suggest that there are no significant differences between onset and offset in F3 for any of the alveolar categories for any manner of articulation. By contrast, F3 is clearly lower at onset than at offset for the Retroflex Weak stop place of articulation. There is a trend towards the same pattern for the Retroflex Weak lateral, and also for the Retroflex Strong stop. An examination of individual speaker data showed that all six speakers had a lower F3 at onset than at offset for the Retroflex Weak stop, while 4/6 speakers had a lower F3 at onset than at offset for the Retroflex Strong stop. In addition, 3/6 speakers had a lower F3 at onset than at offset for the Retroflex Weak nasal and the Retroflex weak lateral. Individual speaker patterns were more sporadic for the Retroflex Strong nasal and for the Retroflex Strong lateral. These F3 results suggest that the Retroflex Weak category is the most likely to show lowered F3 at onset, although the effect is clearest for the stop manner of articulation.

It should also be pointed out that the F3 values for the Alveolar Weak stop category are particularly low (below 2500 Hz) at both onset and offset, with values comparable to F3 values at the onset of the Retroflex stop categories. These results suggest that the Alveolar Weak stop has some acoustic characteristics of retroflexion in terms of its F3 patterning.

4. Discussion

The clearest effects of the apical contrast were observed on F1 and F3. Considering firstly F3, evidence for the most prototypical retroflex articulation – with a clearly lower F3 at onset than at offset – is seen for the Retroflex Weak stop. There is also some evidence of such an articulation in the Retroflex Weak lateral, and also the Retroflex Strong stop. By contrast, the Alveolar Weak stop has a lower F3 at both onset and offset, suggesting a more posterior placement of this articulation at both onset and offset, without a forward movement of the tongue during closure. This F3 pattern aligns with one of the articulatory strategies described above, namely an intermediate point of articulation, with closure and release at the same point along the palate. Such intermediate articulations and their acoustic correlates are reflective of the marginal phonemic status of these apical sounds. It may however be noted that the Alveolar Strong stop has the highest F3 values, a result which is consistent with the spectral centre of gravity results mentioned in the Introduction section.

Importantly, there is no evidence of differences in F3 for the nasal consonants. This is despite the fact that vowel formants were sampled 10 ms from the marked boundary between consonant and vowels; and also despite extensive acoustic and articulatory evidence that Australian languages align velo-pharyngeal closing/opening very tightly with supra-laryngeal closing/opening, precisely in order to minimize the vowel nasalization which would compromise acoustic cues to place of articulation [26]. The F3 results in the present study are also contra observed articulatory data (from two of the speakers included in the present study), which suggest that the nasal manner of articulation is most amenable to a prototypical retroflex articulation, with initial consonant closure at a clearly post-alveolar or pre-palatal place of articulation. This result clearly requires further consideration, although it does suggest that stops allow for the clearest acoustic cues to consonant place of articulation.

Turning now to F1, there is evidence of some interplay between the alveolar versus retroflex contrast, and the manner contrast obstruent (stop) versus sonorant (nasal and lateral). For the stops, there is evidence of a *higher* F1 at consonant onset for the *alveolars* (in both strong and weak prosodic contexts), whereas for the nasals and laterals, there is evidence of a *lower* F1 at onset for *retroflexes* (in both strong and weak prosodic contexts). If these differences in F1 are to be interpreted as allophonic variation in vowel context, they suggest a higher vowel quality preceding retroflex sonorants, but following alveolar stops. However, such reports are not found in the literature and are not consistent with the description of pre-palatalization mentioned in the Introduction, which applies across all manners of articulation. Moreover, the present data were sampled not at the vowel midpoint, but close to the consonant edge. The F1 results therefore more likely reflect a transition state between the central vowel and the consonant closure or release, with F1 treated as a Helmholtz resonance of the oral cavity when the active articulator approaches closure. Typically F1 is lower for laminal consonants, due to a longer and/or narrower constriction between the tongue and the palate (or a longer resonator overall). However, a lower F1 at the onset of a retroflex, which is an apical articulation, suggests a longer Helmholtz resonator overall, which is not in line with a more posterior placement of the consonant closure (assuming the pharyngeal cavity is the main resonator in this case). Instead, it appears that the area of the constriction and/or of the main cavity needs to be considered. If the area of the constriction is smaller, and/or if the area of the main cavity is larger, the Helmholtz resonance is lower. It is possible that the very small amount of tongue-palate contact typical of an apical post-alveolar closure leads to a relatively larger resonating cavity behind the constriction, thereby lowering F1 before a retroflex consonant. It is also possible that the area of the constriction between the tongue tip and the hard palate is very small as the tongue tip is slowly placed in the post-alveolar region, in preparation for a ballistic forward movement of the tongue (as opposed to contacting the hard palate at high speech as is typical of articulations, such as alveolars, which have a virtual target beyond the hard palate [27]). Such considerations clearly require further articulatory-to-acoustic modelling research.

5. Acknowledgements

We would like to thank the Australian Research Council for funding this research. We would also like to thank Gavan Breen for his support. Finally, we would like to thank our speakers – Janet Turner, MM Turner, Raphael Turner, Sabella Turner, Therese Ryder and Veronica Dobson – for their commitment to speech research. *Kele*.

6. References

[1] G. Breen and V. Dobson, "Central Arrernte," *Journal of the International Phonetic Association,* vol. 35, pp. 249-254, 2005.

[2] J. Henderson, *Topics in Eastern and Central Arrernte grammar*. Lincom Europa: Germany, 2013.

[3] G. Breen and R. Pensalfini, "Arrernte: a language with no syllable onsets," *Linguistic Inquiry,* vol. 30, pp. 1-25, 1999.

[4] D. Wilkins, *Mpwarnte Arrernte*. PhD thesis: Australian National University, 1989.

[5] R. Dixon, *The Languages of Australia*. Cambridge: Cambridge University Press, 1980.

[6] R. Dixon, *Australian Languages: Their Nature and Development.* Cambridge: Cambridge University Press, 2002.

[7] N. Evans, "Current issues in the phonology of Australian languages," in John Goldsmith (Ed.), *The Handbook of Phonological Theory* pp. 723–761). Cambridge, MA, Oxford: Blackwell, 1995.

[8] A. Butcher, "Australian Aboriginal Languages: Consonant-Salient Phonologies and the 'Place-of-Articulation Imperative'," in J. Harrington and M. Tabain (eds). *Speech Production: Models, Phonetic Processes, and Techniques.* New York, USA: Psychology Press, pp. 187-210, 2006.

[9] M. Tabain, "An EPG study of the alveolar vs. retroflex apical contrast in Central Arrernte," *Journal of Phonetics,* vol**.** 37, pp. 486-501, 2009.

[10] M. Tabain and R. Beare, "An EPG and EMA study of apicals in stressed and unstressed position in Arrernte," *18th International Congress of the Phonetic Sciences,* Glasgow: Scotland, 2015.

[11] N. Topinzi, Nina and A. Nevins, "Moraic onsets in Arrernte," *Phonology* vol. 34, pp. 615-650, 2017.

[12] H. Simonsen, I. Moen and S. Cowen, "Norwegian retroflex stops in a cross-linguistic perspective," *Journal of Phonetics*, vol. 36, pp. 385–405, 2008.

[13] D. Steriade, "Directional asymmetries in place assimilation: A perceptual account," in E. Hume and K. Johnson (Eds.) *The role of speech perception in phonology* (pp. 219–250). San Diego: Academic Press, 2001.

[14] A. Kochetov and N. Sreedevi, "Manner-specific tongue shape differences in the production of Kannada coronal consonants," paper presented at the Spring 2016 meeting of the Acoustical Society of America, 2016.

[15] R. Bundgaard-Nielsen, B. Baker, C. Kroos, M. Harvey and C. Best. "Vowel acoustics reliably differentiate three coronal stops of Wubuy across prosodic contexts", *Laboratory Phonology* vol. 29, pp. 133–161, 2012.

[16] M. Tabain, "Jaw movement and coronal stop spectra in Central Arrernte," *Journal of Phonetics,* vol. 40, pp. 551-567, 2012.

[17] M. Tabain, G. Breen, A. Butcher, A. Jukes, and R. Beare, "Stress effects on stop bursts in five languages," *Laboratory Phonology* vol. 7(1): 16, pp. 1–23, 2016.

[18] M. Tabain, A. Butcher, G. Breen and R. Beare. "An acoustic study of multiple lateral consonants in three Central Australian languages," *Journal of the Acoustical Society of America,* vol. 139, pp. 361-372, 2016.

[19] M. Tabain, A. Butcher, G. Breen and R. Beare. "An acoustic study of nasal consonants in three Central Australian languages," *Journal of the Acoustical Society of America,* vol. 139, pp. 890-903, 2016.

[20] M. Tabain and G. Breen "Central vowels in Central Arrernte: a spectrographic study of a small vowel system", *Journal of Phonetics,* vol. 39, pp. 68-84, 2011.

[21] R. Winkelmann, J. Harrington and K. Jänsch.. EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech and Language,* vol. 45, pp. 392-410, 2017.

[22] J. Harrington, *The Phonetic Analysis of Speech Corpora*. Blackwell, 2010.

[23] R Core Team "R: A language and environment for statistical computing. R Foundation for Statistical Computing," Vienna, Austria. URL http://www.R-project.org/, 2014.

[24] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York, 2009.

[25] K. Stevens, *Acoustic phonetics.* Cambridge MA: MIT Press, 1998.

[26] H. Stoakes, *An acoustic and aerodynamic analysis of consonant articulation in Bininj Gun-wok.* PhD thesis, University of Melbourne, 2014.

[27] P. Perrier, Y. Payan, M. Zandipour and J. Perkell, "Influences of tongue biomechanics on speech movements during the production of velar stop consonants: a modelling study," *Journal of the Acoustical Society of America,* vol. 114, pp. 1582–1599, 2003.