INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

An ultrasound study of alveolar and retroflex consonants in Arrernte: stressed and unstressed syllables

Marija Tabain[1], *Richard Beare*[2, 3]

[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

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

Abstract

This study presents ultrasound data from six female speakers of the Central Australian language Arrernte. We focus on the apical stop contrast, alveolar /t/ versus retroflex /ʈ/, which may be considered phonemically marginal. We compare these sounds in stressed and unstressed position. Consistent with previous results on this apical contrast, we show that there are minimal differences between the retroflex and the alveolar at stop offset; however, at stop onset, the retroflex has a higher front portion of the tongue, and often a more forward posterior portion of the tongue. This difference between the alveolar and the retroflex is particularly remarked in unstressed prosodic context. This result confirms our previous EPG and EMA results from two of the speakers in the present study, which showed that the most prototypical retroflex consonant occurs in the unstressed prosodic position.

Index Terms: ultrasound, Australian languages, apical contrast, alveolar, retroflex

1. Introduction

Arrernte is a language of Central Australia, spoken by about 2000 people in and around the administrative township of Alice Springs (*Mparntwe*). Like most Australian languages, it has a very rich inventory of coronal consonants [1, 2, 3]. The coronal places of articulation in Arrernte are dental, alveolar, retroflex and alveo-palatal. The dental and palatal sounds are classified as laminal consonants, and the alveolar and retroflex sounds are classified as apical consonants [4, 5].

In this study we focus on the apical stops – alveolar /t/ and retroflex /ʈ/, which may be considered to form a marginal phonemic contrast, given their very low functional load (i.e. very small number of minimal or near-minimal pairs). Previous studies have shown that there is a tremendous amount of phonetic overlap between these two sound categories [6, 7, 8, 9]. However, recent work has shown that at least part of the contrast is mediated by prosodic context. Acoustic results [10] show that both the alveolar and the retroflex have a higher spectral centre of gravity for the stop burst when preceding a stressed vowel (significantly higher only in the case of the alveolar). The alveolar /t/ in turn has a higher spectral centre of gravity than the retroflex /ʈ/ in this stressed prosodic context. By contrast, spectral centre of gravity is lower for these apical sounds in unstressed prosodic context, with no observable difference between the alveolar and the retroflex (result based on seven speakers).

This acoustic work is complemented by a recent EPG (electropalatography) and EMA (electro-magnetic articulometry) study of these apical consonants [11], based on recordings of two female speakers (who were also included in the above-mentioned acoustic study). The EPG and EMA apical data were likewise examined with regard to prosodic context – that is, whether the following vowel is stressed or unstressed. The articulatory results suggest that the tongue is most retracted for the retroflex in the unstressed (i.e. weak) context, but that both apicals have a more advanced tongue position in stressed (i.e. strong) positions. In general, there is more variability in tongue-palate contact patterns for the alveolar /t/ than for the retroflex /ʈ/. In addition, jaw position is highest for /t/ in stressed position, and lowest for /ʈ/ in unstressed position.

Taken together, these results suggest that the most prototypical retroflex is the one found in the unstressed prosodic context. The EPG data suggest that retroflex /ʈ/ in unstressed position shows a ballistic forward movement of the tongue, with initial closure in the post-alveolar or pre-palatal region, and release further towards the front of the hard palate. By contrast, a prototypical "alveolar" articulation would show closure and release at the same forward point along the hard palate. However, many apical articulations show a pattern somewhat intermediate between these two extremes, with closure and release at a region slightly more posterior to the alveolar zone. Morever, the articulatory and acoustic results together suggest that both the alveolar and the retroflex are released at a similar point along the hard palate (since stop burst occurs at the moment of release of the consonant, and a higher spectral centre of gravity suggests a more forward stop release), with the possibility that the stressed /t/ is released further forward than either the stressed /ʈ/ or unstressed /t/ or /ʈ/.

In this study we examine the Arrernte apical stops in stressed and unstressed position using a third articulatory technique, namely ultrasound. Although ultrasound does not give good information about the coronal portion of the tongue [12], it does provide excellent information on overall tongue shape, including the most posterior portion of the tongue (i.e. the tongue root in the pharyngeal region). Preliminary results [13] suggest that the retroflex /ʈ/ has a slightly more forward posterior portion of the tongue, and a slightly higher front part of the tongue, when compared to the alveolar /t/. The purpose of this study is to examine these two apical sounds dependent on prosodic context – stressed or unstressed.

2. Method

2.1 Speakers and Recordings

Six female speakers of Arrernte were recorded to ultrasound using the Telemed Echo Blaster 128 CEXT-1Z, the Articulate Instruments stabilization headset [14], the Articulate Instruments pulse-stretch unit, and the AAA software version 2.16.07 [15]. In addition we used an MBox2 Mini soundcard, a

  http://dx.doi.org/10.21437/Interspeech.2017-578

Sony lapel microphone (electret condenser ECM-44B), and an Articulate Instruments Medical Isolation Transformer. The ultrasound machine, sync pulse, sound card and a software dongle were connected via USB to a Dell Latitude E6420 laptop running Windows software. Typical frame rate was 87 f.p.s., using a 5-8 MHz convex probe set to 7 MHz, a depth of 70 mm and a field of view of 107.7 degrees (70%).

For each speaker, sample palate traces were taken in order to aid with subsequent tracking of the tongue contours. Bite plane was also measured by pressing the tongue up against a ruler held in place by the molars, and all data presented here have been rotated to this bite plane.

Speakers read a list of 92 Arrernte words designed to present the four coronal places of articulation for the oral stop, nasal, lateral and rhotic consonant series. Although some of the stimulus words illustrated homorganic consonant clusters, most of the words contained singleton consonants only. Adjacent vowels were the central vowels /a/ or /ə/, which are the most common vowels in Arrernte (Arrernte has three phonemic vowels, /a ə i/, and a fourth vowel [u] which occurs as a result of rounding on a consonant – rounded consonants were avoided in the list, though were not entirely absent). Where possible, the target consonants were illustrated both in stressed and in unstressed word position. Note that schwa can be stressed in Arrernte.

The words were displayed on the laptop screen. Speakers were asked to say each word three times, or as often as possible within the 5-second recording window set by the Articulate Assistant software. Some speakers were able to produce four or five repetitions in each 5-second window. Each speaker read the list through at least once, and four speakers read through the list a second time. Two speakers presented in the current study were not literate in Arrernte, and were prompted by another (literate) speaker who was present in the room, and/or by the author MT providing an English gloss of the target word. One of these speakers read through the list once, and the other speaker read through the list twice. Note also that the ultrasound machine does not begin recording until about 150 ms into the 5-second audio recording window – as a result, some repetitions were discarded because they were cut off by these limitations.

2.2 Analyses

Acoustic data were labelled using the EMU speech software package [16] version 2.3, and tongue contours were tracked using the AAA software. Tracked tongue data were exported to Simple Signal File Format (SSFF) for compatibility with EMU. Analyses were conducted using EMU/R version 4.4, interfaced with the R statistical package version 3.1.2 [17]. Smoothing spline ANOVAs were calculated using the *gss* package in R [18] using polar co-ordinates [19]. Figures presented were created using the *ggplot2* package [20].

2.3 The current study

For the purposes of the current study, only the alveolar and retroflex stops /t ʈ/ in an intervocalic (i.e. central vowel) context, or in word-initial context, will be examined. The focus of the study is on the realization of these stops in stressed and unstressed position. The apical contrast is in principal neutralized in word-initial position [3, 4, 5], and orthographically this neutralized apical is written as the alveolar 't'. Since metrical stress in Arrernte falls on the first vowel following a consonant [4], this means that many stressed tokens of /t/ are in the potentially neutralized context. However, the realization of word-initial vowels in Arrernte is not consistent: words which orthographically begin with a consonant may be realized with an epenthetic schwa vowel (considered phonologically underlying [21]), and words which orthographically begin with an /a/ may not have that vowel realized (meaning that a following /t/ or /ʈ/ is produced without an initial vowel, even of the shortest duration – in the present database, one such token of /ʈ/ exists).

Table 1 gives the number of tokens of /t/ and /ʈ/ produced by speakers in an intervocalic or word-initial context (numbers in brackets denote the number of tokens that are word-initial – for example, of 29 Stressed /t/ tokens for speaker Carmel, 19 are word-initial). However, only data from cells in which there are at least 10 tokens are included for analysis in the present study, for reasons of statistical validity (these cells are in bold). Note that speakers Janet and Sabella were also the speakers in the above-mentioned EPG and EMA study.

It can be seen that there are many tokens of Stressed /t/, and many tokens of Unstressed /ʈ/. However, there are relatively fewer tokens of Unstressed /t/ and Stressed /ʈ/. This may reflect a lexical tendency across the language. However, it should be noted that any Stressed /ʈ/ tokens that included pre-palatalization were not included in this study. Pre-palatalization of retroflex /ʈ/ is a common, apparently clear-speech, variant of this sound in stressed position – a word such as *artepe* 'back' /aʈəp/ may be pronounced [ɛjʈəpə] (where the first schwa vowel is stressed). Since pre-palatalization has a profound effect on overall tongue shape, such productions were not included in the current set of comparisons. As an indication of how common this clear-speech variant is, the two speakers who apparently produced no tokens of Stressed /ʈ/ based on Table 1, in fact produced 10 (Carmel) and 22 (Sabella) tokens of Stressed and Pre-Palatalized /ʈ/. (In addition, speaker Janet produced 8 pre-palatalized tokens, speaker Lorrayne 15 such tokens, and speaker Mia 16 such tokens. Speaker Phyllis did not produce any pre-palatalized /ʈ/ tokens). Note also that the lack of unstressed tokens for speaker Phyllis is due to the fact that this speaker tended not to produce a word-final vowel.

Table 1. *Number of tokens.*

SPEAKER	Carmel	Janet	Lorrayne	Mia	Phyllis	Sabella
/t/ Stressed	**29 (19)**	**51 (35)**	**75 (42)**	**51 (34)**	**27 (14)**	**53 (33)**
/t/ Unstressed	**11**	**11**	**11**	**15**	-	4
/ʈ/ Stressed	-	9	**11**	3	**11 (1)**	-
/ʈ/ Unstressed	**16**	**33**	**37**	**28**	-	**22**

Tongue spline data were sampled at the acoustic onset, and acoustic offset of the consonant. It is recognized that it is difficult to mark the onset of the stop consonant in a word-initial position based on ultrasound data – however, a comparison of duration values for stops in word-initial position and intervocalic position, with duration values for nasals and laterals in word-initial position and intervocalic position, suggested that our word-initial stop duration values were a reasonable estimate.

3. Results

Since our study is a 2-by-2 design (alveolar versus retroflex, and stressed versus unstressed), we will present figures contrasting: the alveolar in stressed and unstressed position; the retroflex in stressed and unstressed position; the alveolar and the retroflex together in stressed position; and the alveolar and the retroflex together in unstressed position. Data will only be included if sufficient numbers of tokens were present in each data cell for a given comparison.

Figure 1 presents SS-ANOVA plots of the onset and offset of the alveolar stop /t/ for the four speakers for whom there were at least 10 tokens of this sound in both stressed and unstressed position. The stressed tokens are shown in red, and the unstressed tokens are shown in blue.

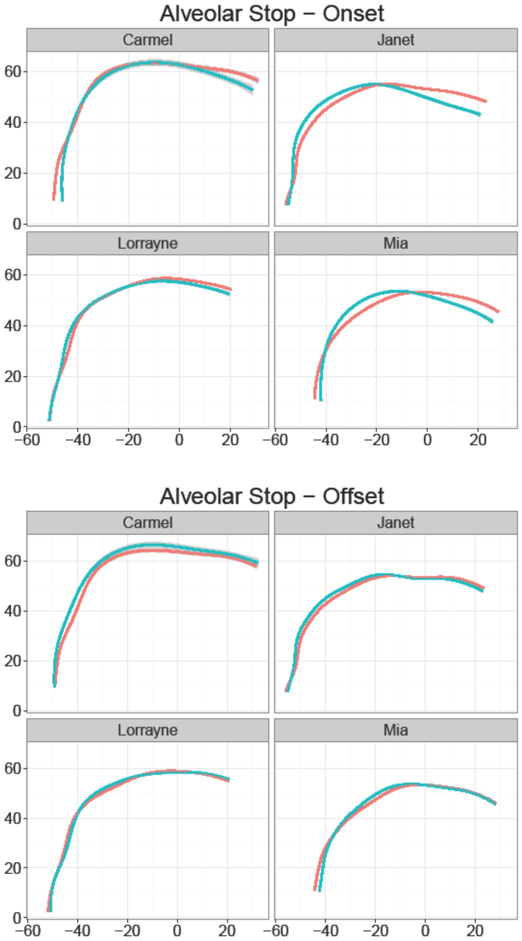

Figure 1: *SS-ANOVA plots of tongue contours for the* *__alveolar__* *stop /t/ taken at stop onset (top panel) and stop offset (bottom panel) for four speakers. Stressed tokens are in red, and unstressed tokens are in blue. In this and in subsequent figures, the front part of the tongue is to the right of each panel.*

It can be seen that there is very little difference between the stressed and unstressed tokens for speakers Carmel and Lorrayne, at both stop onset and offset, although the front part of the tongue is a little higher at stop onset in stressed position. Speakers Janet and Mia likewise show very little difference between stressed and unstressed contexts at stop offset, but a clearly higher front portion of the tongue at stop onset. In addition, these speakers have a slightly more forward back portion of the tongue in stressed position at stop onset.

Figure 2 presents SS-ANOVA plots of the onset and offset of the retroflex stop /ʈ/ for the one speaker for whom there were at least 10 tokens of this sound in both stressed and unstressed position.

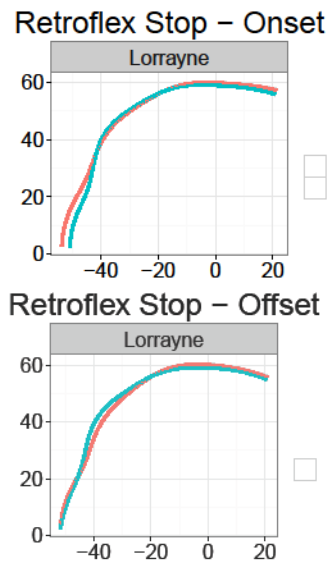

Figure 2: *SS-ANOVA plots of tongue contours for the* *__retroflex__* *stop /ʈ/ taken at stop onset (top panel) and stop offset (bottom panel) for one speaker. Stressed tokens are in red, and unstressed tokens are in blue.*

It can be seen that there is very little difference between the stressed and unstressed retroflex production, at least for this one speaker. In general, the front part of the tongue may be a little higher in stressed position than in unstressed position, while the posterior portion of the tongue varies from stop onset to offset.

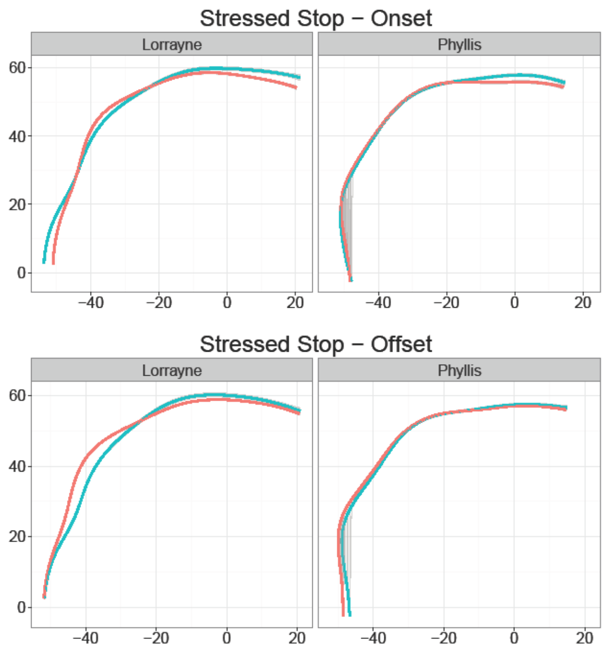

Figure 3: *SS-ANOVA plots of tongue contours for the* *__stressed__* *alveolar and retroflex stops taken at stop onset (top panel) and stop offset (bottom panel) for two speakers. Alveolar tokens are in red, and retroflex tokens are in blue.*

For the next two sets of figures, we present alveolar and retroflex data on the same plot. In these plots, the alveolar stop is in red, and the retroflex stop is in blue. Figure 3 shows these stops in stressed position. It can be seen that there is comparatively little difference between the stressed alveolar and the stressed retroflex. The front part of the tongue is a little higher for the retroflex than for the alveolar at stop onset, and the posterior portion of the tongue may be a little more forward for speaker Lorrayne at stop offset.

Finally, we consider the alveolar and retroflex stops in unstressed position. It can be seen in Figure 4 that there are comparatively fewer differences between alveolar and the retroflex at stop offset, compared to stop onset. At stop onset it is clear that the front part of the tongue is higher for the retroflex than for the alveolar, for all four speakers (and perhaps for three out of four speakers, the posterior portion of the tongue is a little more forward). By contrast at stop offset, the differences are minimal, with the front part of the tongue being only slightly higher for two speakers (Janet and Mia).

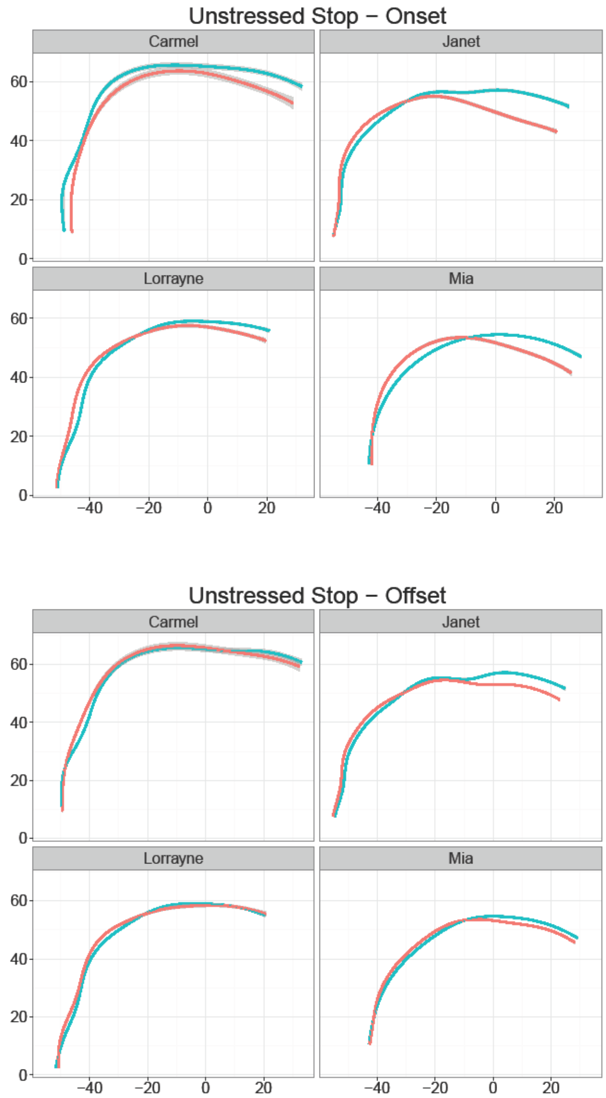

Figure 4: *SS-ANOVA plots of tongue contours for the* *__unstressed__* *alveolar and retroflex stops taken at stop onset (top panel) and stop offset (bottom panel) for four speakers. Alveolar tokens are in red, and retroflex tokens are in blue.*

4. Discussion

Across the SS-ANOVA data presented above, it is clear that differences are greater between the alveolar and the retroflex, or between stressed and unstressed contexts, when examined at stop onset as opposed to stop offset. This is consistent with a well-known result in the literature on apical contrasts, namely that both the retroflex consonant and the alveolar consonant are released at the alveolar region [22, 23]. This leads to a perceptual asymmetry, in that the cues to the contrast are weaker at the right edge of the consonant – in the case of a stop consonant, at the stop burst release. Consequently, the cues at the left edge – i.e. a stop onset – are stronger, leading to the cross-linguistic observation that the apical contrast is weak in initial position, and hence often neutralized. As a result, retroflex consonants are often preceded by a vowel.

In the present study, in both stressed and unstressed contexts, the retroflex has a higher front part of the tongue at stop onset than does the alveolar. However, the contrast is clearly greater in unstressed context than in stressed context (Figure 4 versus Figure 3). This higher front part of the tongue is often, although not always, accompanied by a more forward posterior portion of the tongue. Such a more forward tongue root for the retroflex has also been observed in the Dravidian language Kannada [24], and this more forward tongue root is a little more evident in the unstressed context. Taken together, these results provide some confirmation for the previous EPG and EMA results, which suggest that the most prototypical retroflex articulation occurs in the unstressed prosodic context.

Finally, it must be acknowledged that there is evidence that the alveolar stop has a slightly higher front part of the tongue in stressed context, as opposed to unstressed context. This result is as always more evident at stop onset than at stop offset, where differences are minimal, and suggests a slightly more retroflex articulation for the stressed alveolar compared to the unstressed alveolar, at least at stop onset. This is perhaps the opposite of what might be expected, given that acoustic results (i.e. a higher centre of gravity for the stop burst) suggested that the release of the stressed alveolar stop is a little further forward than for the unstressed alveolar. Since our stressed alveolar tokens included word-initial stops, it is possible that this reflects the neutralization of the apical contrast in word-initial position. Further clarification on this point is required, and this will be the topic of future work. Overall, however, our study has highlighted the complexity of these sounds for phonetic research.

5. Acknowledgements

We would like to thank the Australian Research Council for funding this research via a Future Fellowship to Marija Tabain. We would also like to thank Gavan Breen for his support, and Casey Tait for labelling the acoustic data. Finally, we would like to thank our speakers – Carmel Ryan, Janet Turner, Lorrayne Gorey, Mia Mulladad, Phyllis Stevens and Sabella Turner – for their commitment to speech research. *Kele.*

6. References

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

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

[3] 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.

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

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

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

[7] M. Tabain, "A preliminary study of jaw movement in Arrernte consonant production," *Journal of the International Phonetic Association, vol.* 39, pages 33-51, 2009.

[8] M. Tabain, "EPG data from Central Arrernte: a comparison of the new Articulate palate with the standard Reading palate," *Journal of the International Phonetic Association*, vol. 41, pages 343-367, 2011.

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

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

[11] M. Tabain & 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.

[12] M. Stone, "A guide to analysing tongue motion from ultrasound images," *Clinical Linguistics and Phonetics,* vol. 19 pages 455-502, 2005.

[13] M. Tabain & R. Beare, "A preliminary ultrasound study of nasal and lateral consonants in Arrernte," *Interspeech 2016.* San Francisco: USA, pages 2423-2427, 2016.

[14] Articulate Instruments Ltd., *Ultrasound Stabilisation Headset Users' Manual: Revision 1.4*. Edinburgh, UK: Articulate Instruments Ltd., 2008.

[15] Articulate Instruments Ltd., *Articulate Assistant Advanced User Guide: Version 2.14*. Edinburgh, UK: Articulate Instruments Ltd., 2012.

[16] Jonathan Harrington, *The Phonetic Analysis of Speech Corpora*. Blackwell, 2010.

[17] 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.

[18] Chong Gu, "Smoothing Spline ANOVA Models: R Package gss," *Journal of Statistical Software*, vol. 58(5), pages 1-25. URL http://www.jstatsoft.org/v58/i05/, 2014.

[19] J. Mielke, "An ultrasound study of Canadian French rhotic vowels with polar smoothing spline comparisons," *Journal of the Acoustical Society of America* vol. 137, pages 2858-2869, 2015.

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

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

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

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

[24] 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.