Tongue body position differences in the coronal stop consonants of Wubuy

Christian Kroos1, Rikke Bundgaard-Nielsen1, Louis Goldstein2, Catherine Best1,3

1MARCS Auditory Laboratories, University of Western Sydney, Australia
2University of Southern California, USA
3Haskins Laboratories, USA
c.kroos@uws.edu.au, r.bundgaardnielsen@uws.edu.au, louisgoldstein@usc.edu, c.best@uws.edu.au

Abstract

We investigated the role of the posterior tongue body in the articulation of the four-way coronal stop consonant series of the Australian language Wubuy. Significant tongue body posture differences between the dental, alveolar, retroflex and alveo-palatal stops were found. Further analysis revealed that the differences along the anterior-posterior axis were of the same magnitude as the difference between context vowel tongue back position and consonant midpoint position. Vowel tongue back position in the anterior-posterior dimension was predicted with a higher than expected accuracy by the consonant midpoint position, indicating a strong impact of the consonant on the vowel in Wubuy.

Index Terms: Wubuy, coronal consonants, tongue body, articulatory analysis, electromagnetic articulometry (EMA)

1. Introduction

A four-way place-of-articulation contrast among coronal stop consonants appears to be a rare feature among the phonological systems of the languages of the world. A simple articulatory observation seems to lend itself to a persuasive explanation for its scarcity when considering how close the neighbouring places of articulation (dental, alveolar, retroflex, and alveo-palatal) have to be. Only with precise motor control can those distinctions be maintained. They appear to be unforgiving with respect to speech production errors and, on the perceiver’s side, have a low tolerance for noise interference from the environment.

However, as shown specifically for indigenous languages of Australia, the set of coronal stop consonants is not only distinguished by place of articulation, but also by additional articulatory characteristics. For instance, [1] proposes that for alveolar and retroflex stop consonants vocal tract closure is achieved with the tongue tip (apical), and for dental and alveo-palatal stop consonants with the tongue blade (laminal). Further, [2] proposes that tongue tip constriction orientation is up for apicals, and down for laminals, while [3] also proposes differences in tongue tip posture: neutral for alveolar, up and tilted backward for retroflex, up for dental and down for alveo-palatal. [4] found evidence for different tongue postures separating apicais and laminals, specifically proposing two different types of coordination between tongue tip and tongue body.

The potential need for additional mechanism to enhance the difference between the members of the coronal stop series make a decisive role of the tongue body very likely. Wubuy (also known as Nunggubuyu) is spoken in southeastern Arnhem Land, and preserves the full stop series in all vowel contexts [3]. Thus, it lends itself ideally to investigating the above characterised topics on the level of phonetic realisation of the phono-logical contrasts by examining respective tongue body movements.

In this study we addressed the following research questions:

1. Are there systematic differences in the position of the posterior tongue body in the coronal stops of Wubuy?
2. Is the tongue body retracted for retroflex stop consonants (claimed to be a universal non-violable constraint [5])?

2. Method

2.1. Participants and stimuli

Three female native speakers of Wubuy (ages 51-61) were recorded. The speakers were born in the Numulwar area of Arnhem Land (Northern Territories, Australia) and raised there by native speakers. Two participants reported speaking the Aboriginal language Anindilyakwa with some of their relatives in addition to Wubuy. All three speakers had acquired English as a second language in a classroom setting. They acquired English from the respective ages of five, eight and ten. They also had some basic linguistic training and were involved with local school language revitalisation efforts.

One speaker (W1) had primarily spoken English since she was a young woman. As a result she might have become dominant in English despite speaking Wubuy alongside English in her everyday life. Previous analysis found differences between her articulation and that of the other two speakers [6]. Since speech production is highly variable and tongue positions depend among other things on the shape of the hard palate and given the limited number of speakers in the study, it is impossible to determine whether the speaker employed a different articulation strategy or was influenced by her frequent use of English. To be on the methodologically conservative side the speaker was excluded from the current analysis. Note also that the third speaker (W3) is lacking her upper incisors.

Target words were chosen to include one of the four possible coronal consonants, dental /t/, alveolar /t/, retroflex /r/ and alveo-palatal /l/ in a symmetric vowel context of each of the three vowels /a/, /i/, and /u/, resulting in the target consonants in word medial position: / açuti/, / ići/, / açuli/. All target words were real words of Wubuy and were embedded in a carrier phrase selected to minimise co-articulatory interference: /ha'ajamaa ... aja'ába/ (“I say ... now”).

2.2. Articulatory data acquisition

The Carstens AG500 Articulograph was used to measure articulatory movements of the tongue lips and jaw with the acoustic signal concurrently recorded using two Schoeps CMC6 microphones with a highly directional MK41 capsule. The recordings
tongue back position differences

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Axis</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>W2</td>
<td>ANT-POST</td>
<td>1.46, 16.05</td>
<td>16.05</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>LONG</td>
<td>1.37, 15.06</td>
<td>38.12</td>
<td>0.001</td>
</tr>
<tr>
<td>W3</td>
<td>ANT-POST</td>
<td>1.57, 21.99</td>
<td>55.53</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>LONG</td>
<td>1.90, 26.57</td>
<td>28.76</td>
<td>0.001</td>
</tr>
</tbody>
</table>

2.3. Data post-processing and analysis

The raw voltage amplitudes from the EMA recordings were converted to positional and orientational information using TAPADM [7] and corrected for head motion using custom made head-correction software on the basis of the orientation angles of the sensors [8]. The reference head pose coordinates for the last step were obtained by using the average head pose over the whole session. Visual inspection of the data showed that the reference pose was well aligned with the cardinal axis of the EMA system (since the speakers maintained their gaze at the stimulus monitor).

Kinematically-based gestural temporal landmarks were defined with the MVIEW [9] in Matlab (The MathWorks, Inc.). The onsets and offsets of both the closing and the opening movement were semi-automatically determined by applying a 20% threshold on the tangential peak velocity of the TT sensor (the primary articulator for the coronal consonants). The onset of the opening movement was considered the consonant gesture onset (GONS) and the offset of the closing movements the gesture offset (GOFF) while the offset of the opening movement was considered the consonant nucleus (not to be confused with syllable nucleus in the syntactic sense) onset (NONS) and onset of the closing movement was considered the nucleus offset (NOFF). The temporal midpoint of the consonant (MIDC) was determined as NONS + (NOFF − NONS)/2.

The x, y, and z coordinates of the TB sensors were extracted at GONS, MIDC and GOFF. MVIEW uses a right-handed participant-centred coordinate system where the x axis corresponds to the anterior-posterior axis increasing towards the anterior, y to the lateral axis increasing toward the left hand side of the participant and z to the longitudinal axis increasing upward. We will use the abbreviations ANT-POST, LAT, and LONG respectively. Within our setting and focus on the TB sensor, ANT-POST corresponds to tongue back fronting or retraction and LONG to tongue back height.

3. Results and Discussion

In all of the following we will always analyse the two speakers (W2 and W3) separately as currently no sufficiently accurate normalisation method for speaker vocal tract morphology is available.

3.1. Tongue back position differences

To test whether the means of the ANT-POST and LONG coordinates of the four different consonant types were significantly different from each other at the temporal consonantal midpoint, two repeated measures General Linear Model (GLM) procedures were applied. The independent variable was place of articulation.
ticulation, the dependent variables were the ANT-POST coordinates at MIDC and the LONG coordinates at MIDC, respectively. The means are shown in Figure 1 and 2 and the results of the statistical inference test are summarised in Table 1.

The four consonant types were found to be significantly different in all four tests, indicating that at least for some of the consonant types the posterior part of the tongue body is placed at a specific location distinguishing it from the others. This occurred despite that fact the tongue tip is the active articulator and the places of contact of the tongue tip with the alveolar ridge (teeth) are in the close vicinity of each other for the four consonant types.

Post-hoc comparisons were conducted to determine whether all or only some of the consonant types contributed to the overall significant outcome of the GLM. Most likely the lamina consonants would show different mean locations along the LONG axis as they are articulated with the tongue tip pointing downward [3, 4]. Also, as pointed out in the Introduction, the alveolar consonant and the retroflex could be expected to differ along ANT-POST with the retroflex being more posterior.

As can be seen from Table 2 for speaker W2 and Table 3 for speaker W3 most comparisons revealed significantly different mean positions for the tongue back sensors along both dimensions. With regard to ANT-POST only the comparison dental-alveolar failed to reach significance for both speakers (although it missed significance only marginally for W2) and the dental-alveolar for W3 (again only marginally). Along the LONG axis the dental-retroflex and the alveolar-retroflex contrast failed to reach significance for both speakers and the dental-alveolar for speaker W3. Accordingly, all consonant types are distinguished in TB position by both speakers along at least one spatial dimension with the exception of the dental-alveolar contrast for speaker W3. Note also that in accordance with previously observed differences in the overall tongue movement mechanism of laminals and apicals a strong difference between those two categories was found.

Thus the posterior tongue body appears to be precisely controlled to maintain the difference between the coronal consonants although not the primary articulator. This is surprising as for all of those comparisons the three vowel contexts were pooled and coarticulatory effects should have statistically weakened the observed consonant effects by introducing substantial variance that is not modelled. The tongue body cannot be positioned freely in vowels without acoustic consequences. This finding supports the notion of the “place-of-articulation imperative” proposed in [10].

As for the question of whether the tongue is retracted in retroflex consonants, the mean differences with respect to ANT-POST in Table 2 and 3 reveal that this indeed is the case relative to the alveolar and the dental stop. Note that positive numbers indicate that the first mentioned consonant is more fronted and, obviously, negative numbers the opposite (for the LONG axis positive numbers mean that the first mentioned consonant is higher). However, the magnitude is comparatively small as is evidenced by the fact that the dental and the alveo-palatal exhibits a more fronted tongue back relative to the retroflex than the alveolar. Thus it is questionable whether the observed difference could be indeed counted as tongue back retraction.

### 3.2. Movement magnitude

To get a better idea of the magnitude of those differences we computed the difference in TB along ANT-POST between MIDC (the consonant midpoint) and the consonant gesture onset (GONS) and offset (GOFF), respectively. The resulting distances reflect the difference between the TB position for the vowel and the coronal consonant and therefore are expected to be much larger than the difference among the consonants. Figure 3 shows the mean differences. Surprisingly, they are if anything smaller than the within consonant differences, pointing toward very little movement of the tongue back sensor along the ANT-POST axis during the vowel-to-consonant and consonant-to-vowel transition. As mentioned above, the posterior part of the tongue needs to fulfill certain requirements in order to produce the different vowels used as context in this study as they vary along the front-back dimension (and also in tongue height).

### Table 2: Speaker W2

<table>
<thead>
<tr>
<th>Consonant</th>
<th>t - t</th>
<th>t - c</th>
<th>t - c</th>
<th>t - c</th>
<th>t - c</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANT-POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>3.66</td>
<td>7.05</td>
<td>-2.01</td>
<td>3.39</td>
<td>-5.68</td>
</tr>
<tr>
<td>&lt; α</td>
<td>0.002</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

### Table 3: Speaker W3

<table>
<thead>
<tr>
<th>Consonant</th>
<th>t - t</th>
<th>t - c</th>
<th>t - c</th>
<th>t - c</th>
<th>t - c</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANT-POST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difference (mm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>-2.01</td>
<td>-0.21</td>
<td>-5.16</td>
<td>1.80</td>
<td>-3.15</td>
</tr>
<tr>
<td>&lt; α</td>
<td>0.000</td>
<td>*</td>
<td>0.000</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Table 4: **Regression analysis.** The predictor is the TB position currently underway. 

<table>
<thead>
<tr>
<th>Time</th>
<th>( t - t )</th>
<th>( t - c )</th>
<th>( t - c )</th>
<th>( t - c )</th>
<th>( t - c )</th>
<th>( t - c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>W2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GONS</td>
<td>0.75</td>
<td>0.56</td>
<td>0.59</td>
<td>8.29</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>GOFF</td>
<td>0.82</td>
<td>0.68</td>
<td>0.57</td>
<td>10.71</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>W3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GONS</td>
<td>0.64</td>
<td>0.41</td>
<td>0.32</td>
<td>6.39</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>GOFF</td>
<td>0.46</td>
<td>0.22</td>
<td>0.35</td>
<td>3.99</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>

### 3.3. Impact on vowel

Based on the findings described above, it can be hypothesised that the place-of-articulation imperative leads to a considerable influence on the vowel dictating some aspect of the tone positioning. As a preliminary assessment of the suspected impact we conducted a regression analysis with the TB ANT-POST MIDC coordinate as the predictor and the TB ANT-POST CONS coordinate and GOFF coordinate, respectively, as the dependent variables. To put it simply, we attempted to predict the tongue back position of the preceding and the following vowel from the consonant. Obviously, this should fail and result in very low correlations and regression coefficients close to zero except for a small part of the variance in the vowels due to coarticulation effects. Table 4 summarises the results.

Strikingly, a large amount of variance can be recovered with the predictor variable going as high as 68% in the case of speaker W2 and the following vowel. The lowest value is still 22% for speaker W3 and also the following vowel. The finding supports the hypothesis that tongue back vowel position is heavily influenced by a preceeding or following coronal stop consonant. In accordance with previous empirical findings and consequences that can be derived from theoretical considerations (place-of-articulation imperative) one would expect to find a small vowel system where the few existing vowels are acoustically spread out over a large area of the formant space. In a way, vowel precision appears to be sacrificed to maintain the four-way coronal contrast by embedding articulatory information about the consonant much deeper/further into the vowel than the typically observed transitions.

Clearly, further research into the topic is required and currently underway.

Table 4: **Regression analysis.** The predictor is the TB position along the ANT-POST axis at MIDC, the dependent variable is the same sensor location at time points GONS and GOFF, respectively. Note that the model also included a constant offset coefficient which is not shown here.

### 4. Acknowledgements

The authors would like to thank the Wubuy participants and their families for their participation. We are also grateful to two anonymous reviewers for helpful comments and suggestions. This research was supported by grants from the University of Western Sydney (MARCS Auditory Labs), University of Newcastle, and the U.S. National Institutes of Health NIH (grant DC000403). The research was approved by the University of Western Sydney Human Research Ethics Committee.

The authors wish to especially thank Brett Baker and Mark Harvey for the help with the stimulus generation and their shared insights into Wubuy phonology as well as their support with regard to all aspects of the project relating to our speakers.

### 5. References


