A preliminary spectral analysis of palatal and velar stop bursts in Pitjantjatjara.

Marija Tabain1, Richard Beare2, Andrew Butcher3

1 La Trobe University, Melbourne, Australia
2 Monash University, and Murdoch Children's Research Institute, Melbourne, Australia
3 Flinders University, Adelaide, Australia
m.tabain@latrobe.edu.au, richard.beare@ieee.org, andy.butcher@flinders.edu.au

Abstract

This study examines the contrast between (alveo-)palatal stop bursts and velar stop bursts (/c/ vs. /k/), with particular focus on how the contrast between the two is enhanced in Word-initial vs. Word-medial position. Data are presented from nine speakers of Pitjantjatjara, a language of Central Australia. Analyses show that although there are formant differences between palatal and velar stop bursts, the formant contrast is not enhanced in Word-initial position, with the exception of a lower F3 for /k/ preceding the vowel /a/. By contrast, spectral tilt and the 4th spectral moment (kurtosis) are particularly effective at enhancing the contrast between /c/ and /k/ preceding the vowels /a/ and /i/ (with /c/ having less steep tilt values and lower skewness values than /k/); and the 4th spectral moment (kurtosis) is particularly effective at enhancing the same contrast preceding the vowel /u/ (with /c/ having higher kurtosis values than /k/). These results suggest that Word-initial position in this language is marked not only by pitch movement and extra duration, but also by spectral properties of the stop burst.

Index Terms: palatal stops, velar stops, stop bursts, spectral analysis, Australian languages.

1. Introduction

This paper continues our previous work on the spectral properties of stop bursts in the Australian language Pitjantjatjara1 [1], by specifically considering the contrast between the lamino-alveo-palatal /c/ and the velar /k/.

The contrast between palatals and velars is particularly problematic, with broadly two types of palatals defined [8] – one that might be heard as sounding a little more like [k] (i.e. more canonically palatal), and one that might be heard as sounding a little more like [d] (i.e. more alveo-palatal). This ambiguity in the term "palatal" has led Recasens [9] to propose a re-organization of the International Phonetic Alphabet in order to account for the two main different types of palatals. However, in both types of palatal consonants, a large part of the tongue body is recruited in the articulation, and this is also the case with velar consonants. In addition, the presence of the palatal incline behind the alveolar ridge serves as an important obstacle to airflow at the moment of both velar and palatal release, in many cases resulting in articulation. This perceptual similarity between velar or palatal release and the affricate [f] leads to affrication of /k/ being a very common sound change historically, particularly in front vowel contexts [10,11].

In the present paper, we explore how palatal (or more precisely, alveo-palatal) and velar stop consonants might be made perceptually more distinctive from each other. We do this by examining the stop burst in Word-initial vs. Word-medial position. It is known that domain-initial position is typically a position for articulatory strengthening [12]; the prosodic importance of Word-initial position is all the greater in Pitjantjatjara, since word-initial position is the location of what might be termed "primary stress" in the language, with no phonetic evidence of secondary stress elsewhere in the word [13]. The first syllable of the word in Pitjantjatjara is marked by greater duration, and is the site of a Word boundary tone involving significant pitch movement. It also appears to have greater RMS energy, but there is no evidence for effects of spectral tilt on the vowel. This prosodically prominent position (domain-initial, stressed, site of a pitch tone) would suggest that phonemes at this location are maximally distinct from other, similar phonemes.

Before concluding this section, we should point out that the palatal in Pitjantjatjara, as in many Australian languages, is often affricated, particularly in prosodically strong positions, such as Word-initial. Indeed, to English ears, /c/ in these languages often sounds like [ʧ]. In addition, it should be noted that there is no laryngeal contrast in the stop series in Pitjantjatjara – the stops are typically described as voiceless unaspirated. However, there is a noticeable amount of aspiration following velar and palatal stops, as is typically the case in comparison to more forward places of articulation.

2. Method

In total nine speakers of Pitjantjatjara were recorded for this study, at two different recording sessions 20 years apart. The first group of speakers was recorded in a quiet room at Ernabella community in South Australia in 1990, using cassette tape. The second group of speakers, from Areyonga community in the Northern Territory,2 was recorded in a sound-treated room at the La Trobe University recording studio in 2010, directly to computer. All of the speakers were female, with the exception of one male speaker, MW, from Ernabella.

Two types of speech data are used in the present study. The first type consists of readings of isolated words, and the second type consists of read texts. All of the speakers read a list of words which was designed to illustrate the sounds of

---

1 Pitjantjatjara is a dialect of the greater Western Desert language [2,3,4,5,6,7]. It is spoken mainly in the north-west of South Australia, but extends north into the Northern Territory, and west into Western Australia. It has about 2000 speakers, and is a relatively thriving Australian language. It is still being learned by children today, with English, the official language of Australia, being learnt as a second language at school.

2 Langlois [14] presents a vivid overview of teenage Pitjantjatjara speech in Areyonga.
Pitjantjatjara in different positions in the Word (i.e. Word initial and -medial – note that Word-final consonants are not permitted in Pitjantjatjara), and in different vowel contexts (i.e. preceding the three vowels of the language, /a, i, u/). Acoustic data were labelled by paid labellers using the EMU speech software package (http://emu.sourceforge.net/).

In addition, seven of the speakers read a Pitjantjatjara version of the Three Billy Goats story, "Nanikuta", taken from a children's picture book. The three Areyonga speakers also read a Pitjantjatjara translation of the South Wind and the Sun [15] (SWS) passage – "Walpa Ulpagiranya". Any repeated phrases were included for analysis, provided there were no dis-fluencies. Table 1 gives the number of tokens from each speaker according to Words vs. Texts, and also gives the community of the speaker. The number of consonant tokens in different vowel contexts is given in Table 2.

### Table 1: Number of tokens for each speaker, according to Words vs. Texts recordings. Speaker MW is male, all other speakers are female.

<table>
<thead>
<tr>
<th>Community</th>
<th>Speaker</th>
<th>Texts</th>
<th>Words</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Areyonga</td>
<td>CC</td>
<td>311</td>
<td>308</td>
<td>619</td>
</tr>
<tr>
<td></td>
<td>HB</td>
<td>343</td>
<td>332</td>
<td>675</td>
</tr>
<tr>
<td></td>
<td>KW</td>
<td>349</td>
<td>314</td>
<td>663</td>
</tr>
<tr>
<td>Ernabella</td>
<td>KA</td>
<td>98</td>
<td>384</td>
<td>482</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>-</td>
<td>294</td>
<td>294</td>
</tr>
<tr>
<td></td>
<td>MD</td>
<td>68</td>
<td>375</td>
<td>443</td>
</tr>
<tr>
<td></td>
<td>ML</td>
<td>70</td>
<td>367</td>
<td>437</td>
</tr>
<tr>
<td></td>
<td>MW</td>
<td>68</td>
<td>377</td>
<td>445</td>
</tr>
<tr>
<td></td>
<td>T1</td>
<td>-</td>
<td>381</td>
<td>381</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1307</td>
<td>3132</td>
<td>4439</td>
</tr>
</tbody>
</table>

### Table 2: Number of tokens for each stop in each vowel context, according to Word-initial vs. Word-medial position.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Initial</th>
<th>Medial</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>/c/</td>
<td>a</td>
<td>236</td>
<td>508</td>
</tr>
<tr>
<td></td>
<td>i</td>
<td>154</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>126</td>
<td>202</td>
</tr>
<tr>
<td>/k/</td>
<td>a</td>
<td>673</td>
<td>694</td>
</tr>
<tr>
<td></td>
<td>i</td>
<td>94</td>
<td>224</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>687</td>
<td>699</td>
</tr>
</tbody>
</table>

It is notable that there are comparatively fewer tokens of both /c/ and especially /k/ in the context of a following /i/. It can also be seen that there are many more tokens of /k/ than of /c/, and that overall there are the most tokens in the context of /a/. These results reflect overall lexical frequency of these sounds in Pitjantjatjara, given that our data include all consonant instances in the word-list, as well as tokens from read texts.

Spectral analyses of the data were based on a 10 ms Hamming-windowed Fast Fourier Transform (FFT), centred at the stop release. The spectral data were analysed using the four spectral moments – mean, variance/standard deviation, skewness and kurtosis [16] – and spectral tilt (a regression on the amplitude values as a function of frequency, as returned by the FFT). Spectral moments and spectral tilt were calculated for the frequency range 1-6 kHz, based on an initial visual inspection of the spectra and on the comparable analyses presented in [17] for another Central Australian language, Central Arrente.

In addition, the second, third and fourth formants were estimated for the stop burst. This was done using the ESPS-based Pitch and Formant Tool in EMU, using the default settings with the following exceptions: Window was set to Hamming, and Frame Spacing was set to 5 ms.

For the spectral measures analysed here, a Linear Mixed Effects model was produced for each CV pair /ca ci cu ka ki ku/, with Word-initial vs. -medial position as the controlled variable. Significance was set at 0.0083, which is a Bonferroni correction of 0.05, given the six CV pairs. In all LME analyses, Speaker and Word/Text were set as random factors. The lm() function of the nlme package in the R statistical software was used for the Linear Mixed Effects analysis – this function also gives an estimate of the t-value and the p-value for the data being modelled.

It is recognized that Word-initial position conflates Word-initial with Phrase- and Utterance-initial positions – however, we did not have enough data to separate out these factors given the nature of our database. The random factor Word/Text is an attempt to account for this situation, since all Word-initial tokens in the Words database are necessarily Phrase- and Utterance-initial, whereas the Texts database contains a mix. In addition, we would expect a more didactic speech style and less CV coarticulation in the Words recordings. It is also recognized that the nature of the recordings (cassette vs. computer) has an effect on calculated spectral tilt and moment values – we inspected the results separately for the two sets of recordings, but combined them in the statistical analyses, with the random factor Speaker accounting for these recording artifacts (in addition to accounting for the single male speaker in the study).

All data were labeled using the EMU speech software version 2.3, and analyses were conducted using EMU/R version 4.2.0, with R statistical version 2.3, and analyses were conducted using EMU/R version 4.2 or higher, interfaced with the R statistical package version 2.14 or higher [18].

### 3. Results

Figure 1 shows FFT spectra of the stop bursts for /c/ and /k/. Data from Ernabella and Areyonga are shown separately, since it was not thought wise to combine the computer and cassette recordings for visualization purposes. However, all comments below will relate to both the Areyonga and Ernabella spectra. Figure 2 presents mean values for the various measures used here for all speakers (Areyonga and Ernabella) combined, and Table 3 presents a brief summary of the statistical results.

### Table 3: Results from an LME analysis comparing each CV pair in Word-initial vs. -medial position. An asterisk denotes a result significant at 0.0083, and 'tr' denotes a trend, with p < 0.01.

<table>
<thead>
<tr>
<th></th>
<th>ca</th>
<th>ka</th>
<th>ci</th>
<th>ki</th>
<th>cu</th>
<th>ku</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F3</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F4</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tilt</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td>tr</td>
</tr>
<tr>
<td>2nd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>3rd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
An examination of the palatal spectra confirms our previous results [1] that there is a spectral peak at around 3-4 kHz for the palatal. The spectra for the velar have less energy in the higher frequencies compared to the palatal, with the exception of the front vowel context /i/, where both the velar and the palatal have a high-energy peak at around 3-4 kHz. This front vowel context is clearly the most problematic with regard to contrast between the palatal and the velar.

The first point to note with regard to the LME results is that there are more significant results for the velar contexts than there are for the palatal – overall there are eight significant results for the palatals, and 12 significant results (plus one that just failed to meet significance) for the velars. This is in line with the fact that palatals are much more articulatorily stable, compared to velars [19,20].

If we first consider the formant results, it can be seen that where the formants do shift according to prosodic context (Word-initial vs. Word-medial), it does not always serve to enhance the contrast between the velar and the palatal in initial position. For F2, there are significant differences for /ca/, /ci/ and /ku/ - however, F2 for /ca/ in initial position is closer to the values for /ka/, and F2 for /ci/ and /ki/ are both lower in initial position compared to medial position.

For F3 there are significant differences for /ka/ and /ki/ according to prosodic context. /ka/ in initial position has a lower F3, which takes it further from the F3 values of the palatal /ca/ (the F3 mean for /ca/ is higher in initial position, but this result was not significant). For /ki/, F3 is also lower in initial position, but F3 is also lower for /ci/ (again, the palatal result failed to reach significance).

For F4 there was only one significant result, and this was for /ci/ which also showed a lower value in initial position.

Broadly speaking, there appears to be a lower F2, F3 and F4 for both the palatal and the velar in Word-initial position, when the following vowel is /i/ (though this pattern reached significance only four times out of six). This would suggest that in initial position, the spectrum is slightly more dark before /i/.

It should be noted that there were no significant formant results for /cu/ or /ku/- the back vowel context therefore appears to be relatively stable in terms of formant structure. It is possible that this is because the spectrum in the back vowel context is already weighted towards the lower frequencies.

The pattern of results is quite different when we consider the various broad spectrum measures, namely tilt and spectral moments. For tilt, /ka ki cu ku/ all show significant results. Both /ka/ and /ki/ show lower spectral tilt values in initial position, and in both cases the initial values are further removed from the respective palatal tilt values. The results for /cu/ and /ku/ are a little more ambiguous – although moving in opposite directions (/cu/ has a lower tilt in initial position, and /ku/ has a higher tilt), it cannot be said that the velar/palatal contrast is enhanced by this change in tilt, since the values are very similar for the palatal and the velar.

The results are a little clearer when we consider the various spectral moments. In all but one case (the 1st spectral moment – or centre-of-gravity – for /cu/ vs. /ku/), the results clearly show a greater contrast between /c/ and /k/ in initial position. For the first spectral moment (centre of gravity), both /ka/ and /ki/ have a lower centre of gravity in initial position, and in both cases the initial values are further removed from the respective palatal tilt values. The results for /cu/ and /ku/ are a little more ambiguous – although moving in opposite directions (/cu/ has a lower tilt in initial position, and /ku/ has a higher tilt), it cannot be said that the velar/palatal contrast is enhanced by this change in tilt, since the values are very similar for the palatal and the velar.

The results are a little clearer when we consider the third spectral moment, skewness, we find several significant results. In the case of /a/, both /ca/ and /ka/
show significant results, with /ca/ having a higher skewness value and /ka/ a lower skewness value in initial position compared to medial position – this results in a greater contrast between the palatal and the velar in initial position. A similar pattern is found in the case of /i/, although the result is only significant for the velar, and not for the palatal. Finally, a similar pattern can be seen for /u/, although the result is only significant for the palatal this time, and not for the velar. It should be pointed out that in all cases, the skewness value is positive, which reflects the overall right-skewed nature of the spectral data.

There is only one pair of significant results for the fourth spectral moment, kurtosis, and this concerns the back vowel context /u/. /ku/ has a lower kurtosis value in initial position, and /cu/ has a higher kurtosis value in initial position. The fourth spectral moment is closely related to the second spectral moment, and it is clearly in this aspect of the spectrum that the palatal and the velar differ in the back vowel context.

4. Conclusions

It is clear that formants do not play a role in enhancing the contrast between /c/ and /k/ in Word-initial position – it is possible that the cavity formant affiliations at the moment of release are comparable for palatals and velars, and if an articulation is more forward or more back, it affects the palatales and the velars in a similar way. By contrast the spectral moments, which measure overall spectral shape, are more affected by word position. In particular, the spectral tilt and the third spectral moment (skewness) are especially effective at enhancing the contrast between /c/ and /k/ in the context of a following /i/ or /a/; and the fourth spectral moment (kurtosis) is effective at enhancing the palatal–velar contrast in the back vowel context /u/. These spectral measures broadly reflect the overall balance of energy in the higher vs. lower frequencies of the spectrum. The fact that tilt, skewness and kurtosis play such an important role in contrasting such different sounds in Word-initial vs. –medial position suggests that the spectral differences arise not just from supra-laryngeal adjustments, but either from glottal source effects, or from an interaction between glottal source and the supra-laryngeal cavities.

We should also note that it is not a trivial matter to say that word-initial position is a prominent position in Australian languages, which is what we suggested in the Introduction section. As noted by both [21,22], it is intervocalic position, rather than initial position, which often has a privileged status phonologically in Australian languages – this is usually in relation to the importance of preceding vowel context to the difficult alveolar-retroflex apical contrast. However, our present results suggest that there is indeed evidence that /c/ and /k/ are more distinct in initial position in terms of burst spectrum than in medial position. Word-initial position has been shown to be privileged from a lexical access point of view in many languages [23], and there is no reason that Australian languages should be immune to this trend.

However, the most salient result of the present study is that properties of the stop burst (spectral tilt and the first and third spectral moments) serve to differentiate word-medial from word-initial position. This is evidence that cues to word-boundary lie not only in duration, pitch and energy, but in more intrinsically segmental properties themselves.

Figure 2: Plots of the various spectral measures used in this study, for all speakers combined. Plotted values are means. The different measures are given in columns (first column gives F2 and F3, second column gives F4 an Spectral Tilt, third column gives 1st and 2nd moments, and last column gives 3rd and 4th moments); and the different vowel contexts are given in rows (top row is for /a/ vowel context, middle row is for /i/ vowel context, bottom row is for /u/ vowel context).
5. References


