The perception of coronal stops in Wubuy

Rikke L. Bundgaard-Nielsen\textsuperscript{1}, Brett J. Baker\textsuperscript{2}, Mark Harvey\textsuperscript{3}, Catherine T. Best\textsuperscript{1,4} & Christian Kroos\textsuperscript{1}

\textsuperscript{1}MARCS Institute, University of Western Sydney, Australia
\textsuperscript{2}University of Melbourne, Australia
\textsuperscript{3}University of Newcastle, Australia
\textsuperscript{4}Haskins Laboratories, New Haven, USA

Rikkelou@gmail.com, bjbaker@unimelb.edu.au, mark.harvey@newcastle.edu.au, c.best@uws.edu.au, c.kroos@uws.edu.au

Abstract

Research has indicated that native speech perception may be more difficult than is often assumed. Coronal stop contrast series might fall in this category as they have been reported to be very difficult to distinguish even by native listeners, though this has not previously been systematically examined. It has been claimed that listeners rely on information in preceding vowels in order to correctly perceive these stops, making perception particularly difficult, if not impossible, when the stops are not preceded by a vowel. This paper presents two studies of the discrimination of multiple coronal stops by native speakers of Wubuy and provides strong evidence that native listeners are able to discriminate these consonants, even when they are not preceded by a vowel.

Index Terms: speech perception, coronal stops, Australian languages.

I. Introduction

Several decades of psycholinguistic research have shown that the process of first language acquisition systematically and pervasively influences subsequent native and nonnative speech perception. It might seem trivial to point out that this acquisition process is almost always successful in that learners eventually acquire the phonological system of their native language. However, the problem of first language acquisition may be more complicated than the overall success suggests. Performance on native language speech perception tests is not always perfect \cite{1} and the reasons for this lack of a ceiling effect are not completely clear. Indeed, we know, from previous studies on both native language perception and second language acquisition, that it can be difficult to learn to distinguish phones that are perceived as highly similar, whether they are native or non-native \cite{2, 3}. Also, speakers of first languages with a small(er) inventory of vowels or consonants may experience more difficulty in perceiving new nonnative vowel contrasts than speakers who have larger inventories \cite{4}. Speakers of vowel-rich languages are sensitive to smaller differences in vowels than speakers of other languages, a sensitivity they can apply to a novel language as well \cite{5}. In addition, not all native phonemes are acquired with the same speed and ease by children, whose speech perception and production continues to develop until after they begin school \cite{6}. Thus, not all distinctions are of the same perceptual difficulty, and level of difficulty for a given contrast may be affected by native inventory size and development. Finally, some phonological contrasts are much rarer than others cross-linguistically (e.g., that between dental and alveolar fricatives: \cite{7}, and it has been argued that their infrequency may be due to their low perceptual salience \cite{8}.

The study of (native) perception of Australian Indigenous languages is highly interesting in light of those observations. Many of these languages use a very rare four-way coronal stop consonant series, in which the front of the tongue blocks airflow by forming constrictions with the hard palate in one of several locations in the alveolar region. They are differentiated acoustically only by relatively slight differences \cite{9}, \cite{10}. There has been notable debate as to how perceptible these coronal contrasts are across prosodic contexts (e.g., word-initially as opposed to word-medially). It has been claimed that coronal consonant perception relies on information in a preceding vowel, making the task impossible when they are not preceded by a vowel \cite{2}, \cite{11}. However, recent work on coronal stops in one such language, Wubuy, indicates they are indeed differentiated acoustically, even utterance-initially \cite{9}, thus discrimination of initial contrasts may be possible for native listeners. We investigated this possibility in the studies reported here.

Wubuy is an endangered Indigenous Australian language spoken in Eastern Arnhem Land. It is the first language for adults over the age of ~45 in the community of Numbulwar, NT. Children are no longer acquiring Wubuy as a first language. There are currently perhaps 60 fluent first language speakers of Wubuy.

Phonologically, Wubuy resembles the neighbouring Yolnu languages in having the rare four-way coronal place distinction among the stops /t, t, t̪, c/. According to \cite{12}, the four-way coronal series is contrastive both word-medially and word-initially. Crucially, it is claimed that the apical contrast (i.e., alveolar-retroflex /t̪/-/c/) is difficult to perceive in the absence of a preceding vowel in Wubuy \cite{12}, although Wubuy is one of the few Australian languages said to maintain a contrast in this position \cite{13}.

This paper presents the first systematic examination of native listeners’ discrimination of contrasts from the rare, and purportedly difficult, four-way coronal stop series. We conducted two studies on perception of minimal contrasts among Wubuy coronal stops /t, t, t̪, c/ (apico-alveolar, lamino-dental, apico-retroflex, lamino-alveo-palatal, respectively) by native speakers of Wubuy. Study 1 presented the targets in word-medial position. We expected the native listeners to discriminate among the stops, though we also expected some contrasts to be more difficult than others; in particular, the reportedly difficult contrast /t̪/-/c/ was expected to be discriminated less consistently than the other contrasts. Study 2 tested discrimination of the presumably most difficult stops /t, t̪, c/ in absolute initial position. There was no clear indication from previous research as to whether participants would discriminate these stops in this context. However, given our evidence of their acoustic differentiation in Wubuy, we expected above-chance performance by native listeners, even here. As in Study 1, we expected some contrasts to be more easily discriminated than others. Finally, we compared the listeners’ performance on the medial versus absolute-initial position stimuli, to directly address the relative perceptibility of these stop contrasts in the two contexts.
2. Method

2.1. Stimuli

We recorded three female native speakers (ages 51-61 years), born and raised in the Numbulwar area by parents and family who were native speakers of Wubuy. Two participants also reported speaking the neighbouring Aboriginal language Enindhilyakwa with relatives other than their parents (grandparents, in-laws). All three speakers had acquired English as a second language in a classroom setting. One speaker had acquired English from the age of five, one from the age of eight, and one participant reported acquiring English from the age of 10. All had at least some basic linguistic training, initially for the purposes of Bible translation activities, but all had developed their expertise beyond that base: they were all involved with the local school and language revitalization efforts.

Each of the three participants produced the target consonants /t, ʈ, ɗ/ in the two prosodic contexts given in (1). The target word list is shown in Table 1. Note that the primary stress is on the first syllable of each root word. The lamino-alveo-palatal /c/ was also produced in the VCV context, but not in the /##Ca/ context.

(1) Context 1 Phrase-medial, word-internal /VCV/ targets
Context 2 Utterance initial, word-initial /##Ca/ targets

Table 1. Target words for each context. Note that the orthographic representation (leftmost column for each context) represents the alveolar stop by ‘d’, the retroflex by underscoring ‘d’, and the dental by ‘dh.’ The other columns provide a phonetic transcription of the Wubuy pronunciation of the word in IPA, and its English gloss.

<table>
<thead>
<tr>
<th>Wordlist for the /aCa/ context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Orthography</strong></td>
</tr>
<tr>
<td>madhal</td>
</tr>
<tr>
<td>maada</td>
</tr>
<tr>
<td>ma ga</td>
</tr>
<tr>
<td>maja</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wordlist for the /##Ca/ context</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Orthography</strong></td>
</tr>
<tr>
<td>dhawal</td>
</tr>
<tr>
<td>dawal</td>
</tr>
<tr>
<td>dangle djarra</td>
</tr>
</tbody>
</table>

The participants produced the stops—laminio-dental /ʈ/, apico-alveolar /t/, and apico-retroflex /ɗ/ (and lamino-alveo-palatal /c/ in the /aCa/ context)—in real Wubuy words in the two prosodic contexts, /aCa/ and /##Ca/. The target words were selected to provide a symmetrical vowel context on either side of the consonants in the /aCa/ context. Obviously for the utterance initial /##Ca/ context, there is only the following /a/. An /a/ vowel context offers maximally extensive formant transitions for the coronal stops (see [9], for place of articulation differences in /a/ formant trajectories).

The target words were embedded in Wubuy carrier phrases, which were chosen so as to minimize coarticulation with the adjacent edges of the carrier phrase. The carrier phrase for the word-medial elicitations is given in (2) in both Wubuy orthography and IPA. The carrier phrase for the utterance-initial elicitations, in (3), is a re-ordering of the same words. Wubuy syntax is reported to be non-configurational [12], so both carrier phrases are acceptable to native speakers, although there are information structure differences.

(2) ‘ nga-yamana  ada ba’
’[ŋa-jamana _______a[apa]’
1SG-say,PRS now
‘I say ______ now’

(3) ‘______ nga-yamana ada ba’
[ŋa-jamana [atapa],

2.2. Recording procedure and stimulus treatment

The three speakers read the target words in carrier sentences presented in Wubuy orthography on a computer monitor in a fixed order, blocked by the type of consonant. The participants were encouraged to discuss and rehearse the selected words prior to the recording to ensure that all target items were recognized as known words during the recording. During the recording, the participants were instructed to speak in a clear, comfortable voice as though they were speaking to a friend. Five correct utterances (as judged by the speaker herself, as well as by the other speakers who were present in the room during the recording) were recorded for each target, resulting in a total of 35 correct utterances (5 tokens per 4 targets in the /aCa/ context, and 5 tokens per 3 targets in the /##Ca/ context). Recordings of targets containing coughs, stutters or speech or reading errors were discarded.

For the recording, we used a Shure SM10A headset cardioid microphone, an EDIROL UA-25 USB audio interface, and a laptop computer with Cool Edit 2000. All recordings had a 16-bit sampling depth with a sampling rate of 44.1 KHz. The recordings were made in a sound-attenuated room at MARCS Auditory Laboratories in Sydney. The target /aCa/ and /##Ca/ sequences were excised using a Praat script and checked by the first and second author. Each excised token was enveloped with a 20 millisecond ramp-in and a 10 millisecond ramp-out.

2.3. Stimulus presentation

The discrimination task was presented as two separate randomised cross-speaker categorical XAB discrimination tasks (Study 1 and Study 2) programmed in Psychopy2 and Psychopy X, with the stimuli presented over headphones from a MacBook computer. The task was carefully explained to the participants as one in which a ‘teacher’ (the first voice heard) was being imitated by a ‘good student’ and a ‘bad student’ (voices 2 and 3), and it was the job of the participant to indicate (with a key press on the keyboard) which of the two students (voice 2 or 3) was the ‘good student’ who copied correctly what the teacher had said. The participants first completed the discrimination of the target consonants in medial position (the /aCa/ context) and then the discrimination of the target consonants in initial position (the /##Ca/ context).

For both studies, the inter stimulus interval (ISI) was 500 ms. The response window was presented for 2 seconds following the playing of the third target. If the participant did not respond within the 2 second window, the trial was replayed later. The inter-trial interval was 1 second.

In both studies, the participants were presented with 6 unique triads and six repetitions per contrast type, equaling a total of 36 triads per contrast for each listener. There were a total of 12 contrasts (all combinations of the four consonants /t, ʈ, ɗ, c/ as there might be differences in discriminability depending on the order of presentation) in the /aCa/ context (432 trials total per listener) and 6 contrasts (all combinations of the three consonants /t, ʈ, ɗ/, again allowing for differences in discriminability due to order of presentation) in the /##Ca/ context (216 trials total per listener).
2.4. Participants
We tested 10 native speakers of Wubuy (age 40-65 approximately). Of the speakers, some were literate and some semi-literate in Wubuy. In addition to Wubuy, the participants spoke English and Kriol to varying levels of proficiency. Another six Wubuy speakers were tested, but their results were excluded for the following reasons: four participants failed to understand the task, one was reluctant to complete the task, and one participant reported that her first language was not Wubuy but the neighbouring language Enindhilyakwa.

All testing took place in a sound attenuated booth at Numbulwar School, Numbulwar, NT. All procedures were explained to the participants in English by the first and second authors and in Wubuy by a native speaker who assisted with interpretation and translation when needed. Each participant was compensated for their time and effort by a $100 payment.

3. Results

3.1. Results of Study 1: /aCa/ discrimination
The results of the /aCa/ /aCa/ discrimination task are presented in Figure 1. Chance performance is 50% correct discrimination, and as is evident from Figure 1, the Wubuy speakers are indeed able to discriminate the four coronal stops in their native language (one-sample t-tests against chance performance confirmed this with p values of <0.001 for each contrast; all significant with Bonferroni correction to p < .004), with no obvious effect of order of presentation within each contrast. For ease of presentation in the following, we have therefore collapsed the 12 contrasts into six: /t/-/c/, /t/-/c/, /t/-/c/, and /t/-/c/ (ignoring stimulus presentation order).

The mean discriminability of the six contrasts was 81% correct. Interestingly, however, the discrimination performance of the participants varied with the contrast in question, with correct discrimination ranging from a high of 88% correct (for /c/ versus /c/), to as low as 70% (for the reportedly fragile /t/ versus /c/).

A One-Way Analysis of Variance (ANOVA) of the discriminability of the six contrasts showed that there was indeed a significant effect of contrast type, F(5,114) = 3.681, p = .004. Tukey HSD post hoc comparisons showed that the significant differences were between the poorest discriminated contrast /t/-/c/, and all three contrasts containing /c/: /t/-/c/, /t/-/c/ (marginally significant), and /t/-/c/.

3.2. Results of Study 2: /##Ca/ discrimination
The results of the word-initial /##Ca/ discrimination task are presented in Figure 2. Chance performance is again 50% correct discrimination. As is evident from Figure 2, the Wubuy speakers are also able to discriminate the coronal consonant contrasts tested in the /##Ca/ context (again, one-sample t-tests against chance performance confirmed this with p values of <0.001 for each contrast; all significant with Bonferroni correction to p < .008). For ease of comparison between the two contexts tested in the present study, in the following, we have collapsed the six contrasts into three: /t/-/c/, /t/-/c/, and /t/-/c/. We note, however, that there appears to be some effect of presentation order for the /t/-/c/ contrast, with the /t/-/c/ presentation successfully discriminated 80% of the time, and the /t/-/c/ presentation successfully discriminated only 64%.

The results of the /##Ca/ /##Ca/ discrimination task, however, the discrimination accuracy appeared to differ for the three contrast types, with scores ranging from 72% correct discrimination for /t/-/c/ and /t/-/c/ to 63% correct discrimination for the reportedly difficult /t/-/c/ contrast. A One-Way Analysis of Variance (ANOVA) of the discriminability of the three contrasts, however, showed that this difference is not statistically significant, F(2,57) = 2.197.

3.3. Does context affect coronal place discrimination?
In order to answer the question of whether word-medial coronal place discrimination is easier than word-initial coronal discrimination, we compared the results from the two discrimination tasks in a series of paired samples t-tests. As Study 2 did not include /c/, we excluded all contrasts with /c/ with each of the contrasts containing /c/ being more discriminable than /t/-/c/.
from the comparison, leaving the three most difficult contrasts /t/-/t̪/, /n/-/n̪/, and /ɳ/-/ɳ̪/ for comparison (Bonferroni correction p < .017).

The t-tests (see Table 3) showed that discrimination performance differed between the two studies only for the /ɬ/-/ʃ/ contrast, where the discrimination accuracy fell from 83% correct discrimination in Study 1 to 72% correct discrimination in Study 2. Thus, there was no significant loss of discrimination for the other two contrasts in the word-initial environment: neither for apical contrast /n/-/n̪/ nor for /ɬ/-/ʃ/.

Table 3. Paired samples t-tests of the discrimination accuracy of the three contrasts /t/-/t̪/, /n/-/n̪/, and /ɳ/-/ɳ̪/.

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Study 1 (in %) vs. Study 2 (in %)</th>
<th>t</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɬ/-/ʃ/</td>
<td>83% vs. 72%</td>
<td>6.071</td>
<td>19</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>/n/-/n̪/</td>
<td>76% vs. 72%</td>
<td>1.195</td>
<td>19</td>
<td>ns</td>
</tr>
<tr>
<td>/ɳ/-/ɳ̪/</td>
<td>70% vs. 63%</td>
<td>1.613</td>
<td>19</td>
<td>ns</td>
</tr>
</tbody>
</table>

4. Discussion

This paper presents two studies of native consonant perception in Wubuy. To our knowledge, these studies are the first systematic examination of native discrimination of a four-way coronal stop contrast in any language to date. The results provide new insights into questions of the accuracy (and potential limitations) of native speech perception by showing that the native Wubuy speakers are able to discriminate the coronal stop contrasts. This is the case both when these consonants are presented word-medially (as in Study 1), and when they are presented word-initially (as in Study 2).

There has been little doubt in the literature on Australian languages that listeners would be able to discriminate the stops in the /s/c/ context, and our result confirms this to be the case. However, not all contrasts are discriminated equally well. The observed perceptual asymmetry corresponds well to reports that these stops differ acoustically (and articulatorily) to varying degrees [9]. For instance, it is clear that the lamino-alveo-palatal stop /c/ is more easily distinguished from the other coronals than any of the others are from each other (corroborating the previous identification study of [2]). This result therefore provides evidence that multiple coronal contrasts may be inherently difficult to perceive, without having to infer this perceptual difficulty from the degree of acoustic (and/or articulatory) overlap between the phonemes.

The results from Study 2 provide new insights into the phonetic information native listeners use in discrimination. The results show that native Wubuy speakers are able to discriminate three Wubuy coronal contrasts /t/-/t̪/, /n/-/n̪/, and /ɳ/-/ɳ̪/ in word-initial position, and they are able to do so, in two of these cases, equally well as in the word-medial context. This finding is obviously at odds with previous claims that correct perception of these stops depends on acoustic information carried on a preceding vowel, and also contradicts observations that these stops are virtually indistinguishable when not preceded by a vowel ([2], [11]). Remarkably, the apical contrast /ɬ/-/ʃ/ remains just as discriminable to the listeners in the initial position when there is no preceding vowel as it is when there is a preceding vowel. This lends perceptual support to the conclusions of [9] that place of articulation information can be carried by the following vowel (perhaps in addition to acoustic information in the consonant release) in Wubuy when coronal stop consonants are produced word-initially. It may be significant that Wubuy is the first language with an initial apical contrast to be tested in this way; previous studies were on languages where the apical contrast is neutralised word-initially.

Turning to the overall discriminability of the stops, we would like to suggest two possible explanations for the non-ceiling discrimination scores, which might be regarded as low in both studies (80% in Study 1 and 70% in Study 2). The first is that discriminating coronal contrasts appears to be difficult even for native listeners. Although [2] used an identification rather than a discrimination task, our conclusions are consistent with her finding that the coronals (except the lamino-alveo-palatal) are more difficult to perceive than are more ‘peripheral’ places of articulation, such as labials [also 14]. Our findings are also consistent with the variable performance on native vowel perception presented in [3]. These results demonstrate that even native speakers can have some difficulty with consonant contrasts in their own language, if the consonants are taken out of context and contrasted with native consonants that are acoustically very similar.

5. Acknowledgements

The authors would like to thank the Wubuy participants. We also thank the University of Melbourne for funding this project. The research was approved by the University of Melbourne’s Human Research Ethics Committee.

6. References