Long term average speech spectra in Yolngu Matha and Pitjantjatjara speaking females and males

Hywel Stoakes[1,2], *Andrew Butcher*[2], *Janet Fletcher*[1], *Marija Tabain,*[3]

[1]Linguistics, The School of Languages, The University of Melbourne, Australia
[2]Speech Pathology and Audiology, The School of Medicine, Flinders University, Australia
[3]Linguistics, Faculty of Humanities and Social Science, La Trobe University, Australia

hstoakes@unimelb.edu.au, andy.butcher@flinders.edu.au, janetf@unimelb.edu.au, m.tabain@latrobe.edu.au

Abstract

This paper provides a spectral analysis of two Australian languages Yolngu Matha (YM), Pitjantjatjara (PTJ) and Australian Aboriginal English (AAE) as spoken in two language communities. The aim of this study is to show clear quantitative spectral differences between Australian Aboriginal English and the two Aboriginal languages. Thirteen speakers of Yolngu Matha (Yolŋu Matha) ten male, three female and three female speakers of Pitjantjatjara were recorded reading or retelling a passage in their first language and also in English. The results show that there is a difference between the spectral averages of the two language groups with the AAE having higher amplitudes at higher frequencies when compared to the two Australian languages. The Australian Aboriginal language examples all show higher amplitudes for frequencies between 750 Hz and 2 kHz for all speakers.

Index Terms: Australian Aboriginal, acoustic, spectral average

1. Introduction

Chronic otitis media (COM) develops in the vast majority of Australian Aboriginal infants within a few weeks of birth, leaving up to 80% of Aboriginal children with a significant conductive hearing loss (i.e. greater than 25 dB) [1]. This affects both the low frequency end of the scale (under 500 Hz) and the upper end of the scale (above 4000 Hz). Acoustically the sound systems of Australian Aboriginal languages — including the broader forms of Aboriginal English — are strikingly different from the majority of the worlds languages. They are lacking both in contrasts which depend on low frequency acoustic cues (high vowels, voiced obstruents) and in contrasts which depend on cues at the high frequency end of the spectrum (fricatives, aspirated stops). However they are rich in contrasts which depend on rapid spectral changes in the middle of the frequency range (five or six places of articulation). Thus it appears that Aboriginal languages favour sounds whose characteristics exploit precisely that area of hearing ability which is most likely to remain intact in COM [2]. There has been no research to compare the sounds of Australian languages with the hearing characteristics of the speakers. Before this work can be undertaken an understanding of the spectral profile of Australian languages is needed. The classic study of long term average speech spectra across 12 languages [3] concluded that, "The similarity of the LTASS across samples demonstrates that it is reasonable to propose a universal LTASS which should be satisfactory for many purposes and applicable to most, if not all, languages. Such a LTASS should be suitable for hearing aid prescription procedures and for the Articulation Index, regardless of language" [3]. The aim of this paper is to provide a quantitive analysis of differences in the speech spectra between Australian Aboriginal languages and Australian Aboriginal English. This is done by comparing long term spectral averages of speech samples from each group.

1.1. Australian Languages

Acoustically the sound systems of Australian Aboriginal languages are remarkably similar to one another and at the same time strikingly different from the majority of the world's languages. As far as the consonant systems are concerned, in most Aboriginal languages including the broader forms of Aboriginal English — there is no voicing contrast and no contrast between stops and fricatives. This leaves just a single series of obstruents. On the other hand, all Australian languages have an extensive set of place-of-articulation contrasts [4]. About half of the languages have five different places of articulation and a further one third have six. There are two points to note about these consonant systems: firstly they have unusually few contrasts in the manner-of-articulation dimension and an unusually large number in the place-of-articulation dimension. Secondly, these inventories are rich in sonorants. A typical Australian sound system may consist of 70% sonorants and only 30% obstruents in other words precisely the opposite proportion of sonorants to obstruents to that which is the norm amongst languages elsewhere in the world. As a consequence, these systems are lacking both in contrasts which depend on low frequency acoustic cues such as voicing and also in contrasts which depend on cues at the high frequency end of the spectrum such as friction, aspiration. Additionally, the vowel systems of Australian languages are in general quite small [5] and the majority of them lack any true close vowels [6, 7]. In other words, these systems have no vowel quality distinctions which depend on formant frequencies below about 400 Hz [2].

1.2. Australian Aboriginal English

Many of the basilectal varieties of Australian Aboriginal English (AAE) have significant influence from the Australian language substrate of the speech community in which they are spoken [8]. Phonologically, the heaviest (basilectal) varieties can be almost identical to that of the predominant Australian Aboriginal language of the area. In contrast, the lightest (acrolectal) varieties, may be distinguished from Standard Australian En-

28 – 31 August 2011, Florence, Italy

glish (SAE), only by some small but persistent phonetic difference, such as the use of a clear (non-velarised) /l/ in post-vocalic position [9].

1.3. Aims

The aim of this paper is to provide a quantitative analysis of the differences between Australian Aboriginal Languages and Australian Aboriginal English. This is preliminary to work comparing Aboriginal Languages and Australian Aboriginal English with Standard Australian English.

2. Method

To calculate the long term spectral average, the relative amplitude for each frequency is averaged across the entire speech spectrum of a speech sample. It is possible to compare languages using this long term spectral average profile and this is used in a variety of speech and hearing applications [3]. In the present study the LTASS was calculated for each speaker and then these were combined to find a LTASS for all speakers belonging to a particular language group. Two major Australian language groups are represented and both are Pama-Nyungan languages which are geographically remote from one another.

2.1. Talkers

2.1.1. Yolngu Matha

Thirteen speakers of Yolngu Matha, who are also AAE speakers, participated in this study — ten male and three female talkers of various dialects of YM. The majority identified themselves as Djambarrpuyngu speakers, the dominant dialect of Elcho island. Additionally there was one Gumatj speaker, one Galpu speaker (males) and two Warramiri speakers (females) [1]. The talkers range in age from 25-55 years. There are various levels of English competence and education within the speakers. Some of the younger speakers have had secondary schooling off the island in English speaking schools. Some older speakers have spent some time away from the YM language community mainly in Darwin. Others have never left the Galiwin'ku community. A subset of the YM speakers had hearing tests using pure tone audiometry and all those tested had mild (26-40 dB) to moderate (41-55 dB) hearing loss in one or both ears. This loss being most profound at lower (<750 Hz) and higher (>2 kHz) frequencies.

2.1.2. Pitjantjatjara

Three female speakers of the Areyonga variety of the Pitjantjatjara dialect of the Western Desert group of languages spoken in the south western corner of the Northern Territory of Australia were recorded. Their ages varied from early 20s to early 40s and each had a high level of spoken English competency. There was no audiometric profiling possible for these speakers.

2.2. Recording Equipment

All recordings were made using the same recording equipment and as far as possible under the same recording conditions. Recordings were made using an Edirol R-09HR handheld recorder and a Røde Procaster large diaphragm dynamic microphone. The microphone frequency response is relatively flat with a noticeable proximity effect at close distance. This 'proximity effect' enhances the frequency response in the lower frequencies and can lead to lower frequencies that have a higher amplitude when compared to higher frequencies. The effect is most apparent between $120\,\text{Hz}$ and $220\,\text{Hz}$. There is no explicit compensation for the frequency response of the microphone because all of the recordings used the same equipment and so were subject to the same frequency effects. The microphone was also chosen because of excellent side and rear rejection ensuring that background noise was of minimal concern. The microphone's internal shock mounting there is negligible handling noise. The dynamic microphone has reduced power requirements when compared to a similar large diaphragm condenser and is more reliable in a field setting, this is however, at the expense of a truly flat response. All speech samples were recorded in mono at a 24 bit bit-depth and a $48\,\text{kHz}$ sample rate.

2.2.1. Recording Procedure

For the Yolngu Matha field recordings, the microphone was hand-held, with a soft support for the elbow. The microphone was located approximately $2\,\text{cm}$ from the mouth at a 45° angle below the horizontal plane of the mouth and a 45° angle of incidence. This was done to reduce the influence of high amplitude bursts in plosive sounds which could affect the normalisation procedure (see §3). The recordings for the Pitjantjatjara speakers were conducted in a studio environment with the microphone in a similar position to the YM recordings. In contrast to the PTJ recordings, the YM recordings were conducted outside sitting on the floor taking care to be away from reflective surfaces such as walls and also controlling for external continuous ambient noise sources such as trees, wandering animals or the ocean. As mentioned above in §2.2 the noise rejection qualities of the microphone meant that background noises were of minimal influence.

2.3. Speech Material

For the YM speakers the speech material consists of two recordings each containing $60\,\text{s}$ of connected speech. This was repeated in YM and English for each person. The YM speakers recorded a spoken text in both Yolngu Matha and English on a subject familiar to the speaker. It was not possible to record exactly the same passage for each speaker but they each talked on a similar theme and used similar lexical items and consequently similar phonemic segments. There has been agreement in previous studies '... that the choice of material is not critical provided that it is not grossly unrepresentative phonemically with speech passages that contain the repetition of a few phrases' [3]. For the PTJ speakers each speaker recorded a version of *The South Wind and the Sun*. A 60 second portion of this was extracted and used for analysis.[2] This text was spoken in both English and Pitjantjatjara.

3. Analysis

The recordings were broadly labelled and extraneous transient non-speech sounds were removed using the Emu Speech

[1] These are all dialects/related languages of the Yolngu Matha macro-language. There are however significant levels of multilingualism in most Aboriginal communities making the term *first language* a misnomer. It is more accurate to talk of *mother* language, which would be the initial language that a child would be exposed to and the first language to be acquired to competence.

[2] A version of Æsop's fable *The North Wind and the Sun* used in Illustrations of the IPA in the *Journal of the International Phonetic Association*, adapted for the southern hemisphere.

Database [10]. The speech sounds were analysed using the commands within the *seewave* library [11] a package available for the *R* statistical environment [12]. All samples were peak normalised to 75 dB and any dc-offset was removed. When all the speech from each language was combined it was possible to extract a mean spectrum for each language. This mean spectrum could then be used to calculate a long term average speech spectrum (LTASS).

4. Results

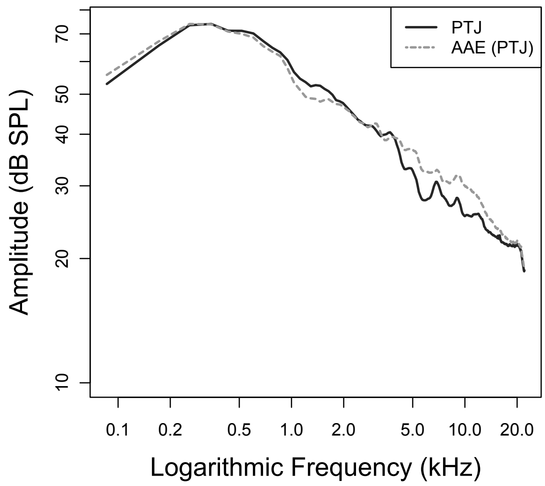

Figure 1: *Average spectra of female Pitjantjatjara talkers speaking both Pitjantjatjara and Aboriginal Australian English for a* 60 s *speech sample.*

The plots of the long term average spectra are shown in Figures 1, 2 and 3. Both frequency and amplitude are plotted using a logarithmic scale to more accurately represent hearing perception. When these data are plotted, clear patterns emerge. In the combined average spectra for speakers separated by sex and language group, the spectrum for AAE rises in amplitude relative to the associated Australian language above approximately 2 kHz. In contrast the signal between 750 Hz and 2 kHz has higher amplitudes in the Australian languages. Although this is a very small difference, the divergence holds true for all speakers regardless of sex or language group membership. It should be noted that many important phonetic distinctions are made between these frequencies, specifically in the domain of place of articulation and they are particularly salient for the perception of speech sounds in general. The female speakers show the greatest differences in the frequencies above 2 kHz with Australian Aboriginal English showing higher amplitudes (see Figures 1 and 2). As discussed below this could be due to the inclusion of fricatives in the phonology of AAE which are absent in the Australian languages. The male speakers (see Figure 3) show a greater similarity between the English sample and the Australian language sample than the females both individually and as a group and the difference in higher frequencies is less pronounced. Despite the similarity however, nine out of ten speakers have higher amplitudes in YM for the frequencies between 750 Hz and 2 kHz when compared to AAE. To give a

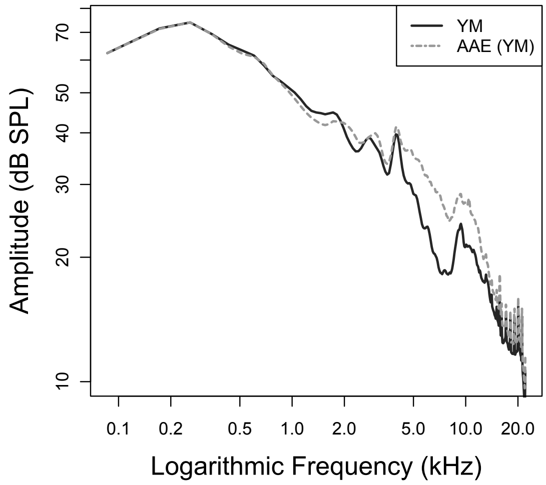

Figure 2: *Average spectra of female Yolngu Matha talkers speaking both Yolngu Matha and Aboriginal Australian English for a* 60 s *speech sample.*

clearer picture of the variance of these data, the combined mean ($\bar{x}$) and standard deviation (*s.d.*) for a selection of frequencies between 500 Hz and 4 kHz is presented in Table 1. The variance for the female speakers of YM is shown in Table 2. As we can see in each of the tables there is very little difference between the means of the Australian language and AAE but the differences that are evident are consistent across speakers.

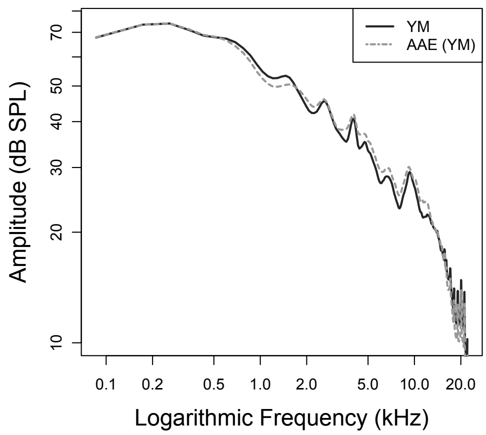

Figure 3: *Average spectra of male Yolngu Matha talkers showing both Yolngu Matha and Aboriginal Australian English for a* 60 s *speech sample.*

5. Discussion and Conclusions

In this sample, the differences in frequency amplitude are relatively small yet they are very consistent across speakers. This

	YM		AAE (YM)	
freq. (Hz)	$\bar{x}$ (dB)	*s.d.* (dB)	$\bar{x}$ (dB)	*s.d.* (dB)
250	74	0.37	74	0.36
500	67	3.9	68	3.3
750	64	4.6	63	5
1000	55	4.7	53	5.4
1250	51	4.8	49	5.3
1500	52	4.7	49	6.8
2000	43	3.8	44	5.2
4000	39	8.7	39	7.3

Table 1: *Mean ($\bar{x}$) and standard deviation (*s.d.*) for LTASS for* ***male*** *Yolngu Matha talkers for both Yolngu Matha and Aboriginal Australian English.*

	YM		AAE (YM)	
freq. (Hz)	$\bar{x}$ (dB)	*s.d.* (dB)	$\bar{x}$ (dB)	*s.d.* (dB)
250	74	0	74	0
500	62	7.6	63	6.8
750	56	7.4	57	7.6
1000	50	4.2	49	5.4
1250	45	5.2	44	5.3
1500	44	6.2	41	6.7
2000	39	8	42	6.4
4000	39	4.2	40	4.4

Table 2: *Mean ($\bar{x}$) and standard deviation (*s.d.*) for LTASS for* ***female*** *Yolngu Matha talkers for both Yolngu Matha and Aboriginal Australian English.*

may be due to the fact that the YM and PTJ speakers when they speak English, reduce the contrasts between obstruents (frication and voicing) and in addition use a reduced vowel space when compared to SAE (but see [8]). The higher frequencies such as those above 5 kHz are normally not perceptually salient in the speech sounds of most languages except for fricative sounds. As there are no fricative phonemes in the vast majority of Australian languages this may be a reason for the differences in average spectra shown in Figures 1,2 and 3. There is much less of a difference between the spectra in the male YM speakers. This difference in the YM male group is not thought to be due to the larger sample size as this same difference is observable in individual speakers. The sex differences may be primarily due to the hyperarticulation of fricatives by the majority of female speakers in the AAE sample and that the majority of the females use a relatively acrolectal variety of AAE when compared with the males. In the analysis, one problem was that we were unable to control the recording conditions in a field environment. In three of the recordings of Yolngu Matha (two females and one male) there was continuous insect noise that was centred around 4 kHz and a harmonic just above 8 kHz. This noise was well below the speech level in the raw recordings but is visible on plots of the normalised long term spectral averages. There is a clear quantitative difference that emerges between the long term average spectrum of Australian Aboriginal English when compared to both Yolngu Matha and Pitjantjatjara. The frequencies above 2 kHz have a higher amplitude for AAE and the frequencies between 750 Hz and 2 kHz have higher amplitude in the Aboriginal languages examined. We can only speculate whether the differences between AAE and Australian Aboriginal languages have implications for the speech perception of the speakers, but this will be investigated in future work with both perceptual and audiometric studies underway. It has long been recognised that, because the onset of COM-induced hearing loss occurs in the first two years of life, normal speech and language development may be disrupted [13]. There has also been an assumption that this disruption occurs irrespective of the language being spoken. However, it may be the case that acoustic aspects of Aboriginal languages make them inherently more robust than English as a medium of communication for those with the type of hearing loss associated with COM.

6. Acknowledgements

The authors would like to thank Michele Swanborough, Dorothy Bukulatjpi, Peter Datjing Burarrwanga, Shepherdson College at Galiwin'ku and the speakers who kindly gave their time to participate in this study. Thanks to four anonymous reviewers for their comments, all errors remain our own. This research was made possible with funding from the Australian Research Council.

7. References

[1] Coates, H., Morris, P., Leach, A., and Couzos, S. (2002). Otitis media in Aboriginal children: tackling a major health problem. *The Medical Journal of Australia*, 177(4):177.

[2] Butcher, A. R. (2006). Consonant-salient phonologies and the 'place-of-articulation imperative'. In Harrington, J. and Tabain, M., editors, *Speech Production: Models, Phonetic Processes and Techniques*, pages 187–210. Psychology Press, New York.

[3] Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., et al. (1994). An international comparison of long-term average speech spectra. *Journal of the Acoustical Society of America*, 96(4):2108–2120.

[4] Butcher, A. R. (1995). The phonetics of neutralisation: the case of Australian coronals. In Windsor Lewis, J., editor, *Studies in General and English Phonetics. Essays in honour of Professor J.D. O'Connor*, pages 10–38. Routledge, London.

[5] Dixon, R. M. W. (2002). *Australian Languages: Their Nature and Development*. Cambridge University Press, New York.

[6] Butcher, A. R. (1994). On the phonetics of small vowel systems: evidence from Australian languages. In Togneri, R., editor, *5th Australian International Conference on Speech Science and Technology*, volume 1, pages 28–33, Canberra. ASSTA,.

[7] Fletcher, J. and Butcher, A. R. (2002). Vowel dispersion in two Northern Australian languages: Dalabon and Bininj Gun-Wok. In Bow, C., editor, *9th Australian International Conference on Speech Science and Technology*, pages 343–348, Melbourne. ASSTA.

[8] Butcher, A. R. (2008). Linguistic aspects of Australian Aboriginal English. *Clinical linguistics & phonetics*, 22(8):625–642.

[9] Butcher, A. R. and Anderson, V. (2008). The vowels of Australian Aboriginal English. *Proceedings of the 9th Annual Conference of the International Speech Communication Association*, 22-26 September 2008, pp. 347-350. Brisbane, Australia.

[10] Harrington, J., *et.al.* (2009). *emu: Interface to the Emu Speech Database System*. R package version 4.2.

[11] Sueur, J., Aubin, T., and Simonis, C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18:213–226.

[12] R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[13] Aithal, S., Yonovitz, A., and Aithal, V. (2008). Perceptual Consequences of Conductive Hearing Loss: Speech Perception in Indigenous Students Learning English as a 'School' Language. *Australian and New Zealand Journal of Audiology*, 30(1):1–18.