Comparing acoustic analyses of Australian English vowels from Sydney: Cox (2006) versus AusTalk

Jaydene Elvin & Paola Escudero

The MARCS Institute, University of Western Sydney

j.elvin@uws.edu.au, paola.escudero@uws.edu.au

Abstract

This study presents a comparison of the acoustic properties of Australian English monophthongs produced by 60 monolingual females from Sydney's Northern Beaches reported in Cox's [1] corpus and by the four monolingual females from Sydney recorded within the AusTalk corpus [2]. Cross-corpus discriminant analyses are used to investigate the acoustic similarity between the two corpora to determine whether the values from these corpora would be appropriate for predicting L2 difficulty in future cross-linguistic studies using Western Sydney speakers. Preliminary findings suggest that there is little overall acoustic similarity across these two vowel corpora as classification scores from the discriminant analyses were consistently higher for the Cox corpus than AusTalk. In particular, greatest variation between the two corpora is observed in their productions of front vowels. Limitations for drawing conclusions based on the current data are provided and the need for an additional corpus of Australian English vowels from speakers in Western Sydney for future cross-linguistic studies is proposed.

Index Terms: acoustic analysis, Australian English vowels, discriminant analysis

1. Introduction

The Australian English (AusE) dialect was long considered non-existent or a mere copy of the British dialect until 1946 when Mitchell [3] recognized the need for a set of phonetic symbols that accurately depicted AusE pronunciation [4]. Despite this intention, the vowel transcriptions he proposed still seemed to be reflective of British English rather than a true Australian English standard [4, 5]. Therefore, in response to the need for a more accurate standard, Australian-based phoneticians [5] conducted a new study with the aim of reconsidering the phonetic characteristics of AusE vowels. They also establish a basis for the distinction between broad, general and cultivated accents within the regional accent of Australian English, as AusE was generally discussed in terms of these varieties.

Harrington and colleagues acoustically analysed monophthongs and diphthongs produced by male and female speakers of these three varieties of AusE from the Australian National Database of Spoken Language [5]. It was found that the monophthongs provided fewer acoustic cues than diphthongs in the broad, general and cultivated accents. The authors also proposed modifications to Mitchell's [3] transcription to more accurately reflect the current AusE pronunciation. As a result a new set of symbols were devised and included in this set are the 12 stressed AusE monophthongs which will be analysed in this study, namely /iː, ɪ, e, eː, ɜː, ɐ, ɐː, æ oː, ɔ, ʊ, ʉː/. As the broad, general and cultivated registers of AusE seem to display considerable phonetic overlap [5] for monophthongs, here we will refer to the general register as AusE.

More recently, Cox [1] presented an acoustic examination of AusE vowels produced by 60 male and 60 female adolescent speakers from the Northern Sydney region. This study was the first to provide a large-scale published acoustic analysis of vowels produced by Australian females and to compare formant and duration properties across the two genders [1]. It was therefore designed with the intent of becoming, and has since been used as, a point of reference for those who require a detailed acoustic description of AusE vowel production [1].

Within the past few years a newer corpus of AusE audio-visual speech production, known as the AusTalk corpus [2], has emerged which may be useful for acoustic analyses of AusE vowels. This large-scale corpus is the result of a collaboration funded by the Australian Research Council between 30 Chief Investigators across 13 Institutions [2]. The corpus consists of 1000 Australian English speakers from 17 different locations and comprising of various cultural backgrounds, recognising the need for a corpus that reflects Australia as it is currently, i.e., a multicultural hub. It is also considered a forward step in audio-visual research as data were collected using state-of-the-art audio and video technology and can, therefore, be used by a number of disciplines within the speech science field.

Acoustic analyses such as those reported in [1] are frequently used to predict L2 difficulty. Theoretical models such as The Second Language Linguistic Perception model [L2LP, 6] posit that detailed acoustic comparisons of the native and target language can be accurate predictors of non-native and L2 difficulty. Given that the Cox corpus is the current Australian standard, and that AusTalk has been designed for research across the speech science field, we examine the acoustic similarity of vowels between a subset of the AusTalk data (i.e., AusE vowels by monolingual female speakers from the Sydney region) and the acoustic analysis reported in [1]. In particular, we aim to determine whether the formant and duration values of these two corpora would be appropriate for predicting L2 vowel perception difficulty for native AusE speakers from Western Sydney.

In this study, we will refer to the two corpora as the Cox and AusTalk corpus, respectively. Currently, the Cox corpus is the only available AusE corpus that provides average formant values and vowel duration for a large number of male and female speakers, including variation in acoustic values expressed in standard deviations. However, the data were produced by adolescent speakers, rather than adults, who are commonly investigated in most recent vowel acoustic studies [7-9] including AusTalk. Given that previous studies have shown that formant values for vowels differ for children and adults [10], it is possible that vowel formants also vary between teenagers and adults. As our ongoing research involves an investigation of how university students learn other languages, the values from teenagers may not be appropriate for making

acoustic predictions. However, if the two corpora are comparable then this age difference may be negligible and either corpus is therefore appropriate for cross-language acoustic predictions.

We will show below that considerable variation across corpora is observed. We will discuss that beside age differences in the speakers, the use of different methodologies as well as possible regional differences between speakers may underlie the variation across the corpora. Finally, we will highlight the need for an additional corpus of Australian English vowels from speakers of Western Sydney for future cross-linguistic studies following the L2LP [6] framework.

2. Method

2.1. Speakers

Vowels in the Cox corpus were produced by 60 female and 60 male adolescent monolingual speakers from Sydney's Northern Beaches. They were second generation Australians aged between 14 years 11 months and 16 years 10 months who spoke only English at home.

We selected speakers from the AusTalk corpus who had comparable linguistic backgrounds to those considered in the recent vowel acoustic studies [7-9]. That is, speakers were highly-educated monolingual speakers of AusE, with AusE speaking parents, born and raised in Sydney and aged between 18 and 30.

Of the 1000 AusTalk speakers, 19 partially conformed to these criteria. Twelve were excluded as their parents were from non AusE backgrounds[1]. Here we report the acoustic similarity between the vowels produced by the four monolingual females from Sydney in AusTalk and the 60 females from Sydney's Northern Beaches recorded within Cox's corpus [1].

2.2. Recordings

The data for the Cox corpus were collected in a quiet room at the speakers' school using a portable Marantz CP430 cassette recorder and a Beyer M88 dynamic microphone. Speakers were instructed to read 18 AusE vowels in the hVd context from flashcards presented four times in a random order. Although the Cox corpus includes an analysis for 18 AusE vowels, we selected 12 stressed AusE monophthongs, namely, /iː, ɪ, e, eː, ɜː, ɐ, ɐː, æ oː, ɔ, ʊ, ʉː/ to compare with the data collected from AusTalk.

We selected the most similar context in AusTalk for comparison with the Cox corpus, namely individual words, which were recorded over three sessions and at various universities in the Sydney region. Recordings were made using a standardized self-contained portable recording station known as a "black box" [2].

Speakers read monosyllabic words presented randomly via a computer screen. Each word contained one of 12 Australian English stressed monophthongs mentioned above and were produced in five contexts (hV, hVt, hVd, hVl, hVn). One token with no background noise, hesitation or mispronunciation was selected per context, except hVd for which two tokens were selected. We thus chose 5 tokens per vowel from the available tokens with reasonable quality, for a total of 240 vowel tokens (5 tokens x 12 vowels x 4 speakers).

[1] One of our female speaker's parents was born in New Zealand but reported an Australian accent and was therefore included in the analysis as their results did not affect our overall findings.

Contexts other than the one used in [1], namely hVd, were included in order to apply the same method of analysis as in recent acoustic analysis [7-9], which requires a minimum of four tokens per vowel which would not be possible with the hVd context alone. These previous studies have shown clear acoustic differences for vowels produced in different phonetic contexts. We therefore also examine whether context variation played a role in the comparison across corpora and more generally, in AusE vowel acoustics.

2.3. Data Analysis

Formant analysis for the Cox corpus was conducted using a 12 pole LPC autocorrelation analysis, with F1, F2 and F3 values manually checked and modified if needed. The accuracy of the measurements was validated through intra-judge and inter-judge reliability. For intra-judge reliability, ten percent of the total data set (i.e., 12 subjects) was re-examined and redrawn if necessary. The frequency value of the original measurement was subtracted from the frequency value of the second measurement to obtain a disagreement measure. For inter-judge reliability, another experienced speech scientist independently checked the accuracy of the researcher's measurements for the same set of twelve subjects [1].

The analysis of the AusTalk data was conducted for the present study following the technique used in recent vowel acoustic analyses [7-9] to allow for future cross-linguistic comparisons. We first calculated the duration for each of the selected vowel tokens manually by placing boundaries at the start and end point of each vowel using the Praat program. The "optimal formant ceiling" technique [7-9] was used to determine F1, F2 and F3 values at the 50% of the duration of each vowel. That is, for each vowel of each speaker (i.e., 5 tokens per vowel), the "optimal ceiling" was chosen as the one that yields the least amount of variation for the first and second formant within the set number of annotated tokens for the vowel. Formant ceilings ranged between 4500 and 6500 Hz for females.

3. Results

We conducted several discriminant analyses (DA) to statistically determine cross-corpora acoustic similarity. DAs are generally used in perceptual studies to model classification patterns based on the acoustic values of a particular language or dialect and to compare the model's results to real listeners' performance. Many previous cross-language and cross-dialectal acoustic analyses have used this technique to assess whether acoustic values are predictive of listeners' vowel classifications [e.g., 10, 11]. However, other studies have also used this technique to establish the acoustic similarity between two data sets [12].

As the individual values were not available for the Cox corpus, we randomly generated 60 tokens (representing 60 female speakers) for each of the 12 AusE vowels using the reported averaged values of F1, F2, F3 and duration values and their standard deviations. We then trained a DA model on 80% of these randomly generated tokens and used the remaining 20% for cross validation. We then examined correct classifications, i.e. when the speaker's intended vowel was classified as that same vowel by the model, for trained and novel (cross-validation) tokens depending on the acoustic dimensions included in the model, A model that included only F1-F3 values yielded 72% correct classifications across all vowels for trained tokens and 66% for novel tokens drawn

from the same corpus. As vowel duration is a cue to vowel identity in AusE, we trained a second DA model adding duration values, which indeed led to a substantial increase in overall classifications for trained (91.8%) and novel tokens (94.4%). The slightly higher performance for the novel tokens suggests that the randomly selected set had less overlap in duration values across vowels.

For the comparison across corpora, we tested how a DA model trained with all 60 tokens per vowel from the Cox corpus would classify AusTalk vowel tokens produced in all consonantal contexts. We report here the test that yielded the best classification results for AusTalk tokens, namely testing the original DA model with AusTalk tokens from the hVd context, which was the same context used in the Cox corpus. A model with only F1-F3 yielded very similar overall correct classifications as the first model reported above for the trained tokens (71.4%) but only 62.5% for the test tokens from AusTalk, and including duration yielded considerable the same considerable improvement for trained tokens (92.5%) as in the second model above but negligible improvement for AusTalk tokens (64.6%). A small classification improvement (67.7%) for AusTalk was shown when testing with the two individual tokens (rather than their average) from the hVd context.

To establish that this striking difference in the classification of tokens drawn from the two corpora is not a result of using different token numbers, i.e. only 8 per vowel for AusTalk versus 60 per vowel for Cox, , we used the same number of tokens across corpora for training and testing. That is, we randomly selected from our 60 simulated Cox tokens, 8 tokens per vowel representing two tokens per vowel produced by four speakers. A model trained with these smaller set of Cox tokens yielded 96.9% correct classification, while classification of AusTalk tokens remained much lower (68.8%).

Additional DAs were conducted to determine the likelihood that the AusTalk tokens belonged to the same distribution as the vowels in the Cox corpus. We first created four sets of tokens with randomly selected acoustic values for each corpus. Due to the limited number of hVd tokens in the AusTalk corpus, we limited each set to 4 tokens per vowel to ensure the two corpora were comparable. The model trained with the formant and duration values of the 56 Cox tokens (the tokens from the random sets were not included in training) yielded high classification percentages for the four random Cox sets (85.4% - 91.7%), suggesting that classification of random sets of tokens drawn from this corpus is fairly stable. In contrast, the four AusTalk random token sets yielded much lower correct classifications (62.5% - 75%).

Although there is greater variation in the percentage of correct classification for the above four AusTalk random sets than for the four Cox random sets, we conducted an independent samples t-test to determine the likelihood that these AusTalk tokens came from the Cox distribution. The independent t-test revealed a significant difference between the two corpora $t(6) = 4.742$, $p = .003$, which suggests that it is unlikely that the AusTalk tokens came from the Cox distribution.

Table 1 below shows the percentage of AusTalk vowels classified as Cox vowels from the DA model that yielded the best classification results for both corpora (i.e. the model that included formant and duration values and all tokens in the hVd context).

Table 1: Highest correct percentage classification (in bold) of AusTalk vowel tokens classified as each of the intended Cox vowels (Percentages are rounded to the nearest whole number, and therefore total classifications per vowel (rows) may not add up to 100%).

iː	ɪ	eː	e	ɜː	ɐː	ɐ	æ	oː	ɔ	ʊ	ʉː
100											
	38		38				50			25	13
		25		13			50				
			50								
				75		25					
					75			13	13		
						75	**13**		25		
						88		**100**			
						13			**63**	25	
										100	**100**

In general, back vowels yielded higher correct classifications than front vowels. In particular, there was 100% correct classification for /oː/, /ʊ/ and /ʉː/. The central vowels /ɜː/ and /ɐː/ also yielded high classification results in the discriminant analysis each with 75% correct classifications. With the exception of /iː/ which had 100% correct classification, front vowels yielded lower classification scores. Only 38% of the AusTalk /ɪ/ tokens were correctly classified, with 38% being classified as /e/ and 25% classified as /ʊ/. Only 25% of the AusTalk /eː/ tokens were correctly classified with 50% classified as /æ/ and 13% as /ɜː/ and 13% as /ʉː/. The classification percentage for /e/ was only 50% with the remaining 50% being classified as /æ/. Finally AusTalk /æ/ was incorrectly classified as /ɐ/ 88% of the time and correctly classified only 13% of the time.

4. Discussion

The present study presents an acoustic comparison of 12 stressed AusE monophthongs spoken by female speakers in the Sydney region from the Cox [1] and AusTalk [2] corpora. The findings suggest that overall; the two corpora are not acoustically similar. In the discriminant analyses described above, the classification scores for the Cox corpus are consistently higher than that of the AusTalk data. This finding suggests that it is unlikely that the AusTalk vowel tokens could be from the same distribution as that of the Cox vowel tokens. In particular, as observed in Table 1, the greatest variation between the two corpora is in their productions of the front vowels, /ɪ, e, eː and æ/.

While the present study demonstrates through the use of discriminant analyses that the two corpora are not acoustically similar, what remains uncertain is the principle cause of these differences. As mentioned in the Introduction, previous studies [e.g., 10] have demonstrated that children and adults have differing formant values in their vowel production. This may also be the case for adolescents and adults and could be a possible factor causing these cross-corpora differences. However, aside from age differences and the fact that one AusTalk speaker reported knowledge of a language other than English with no proficiency level provided, other factors such as methodology and regional dialect can be considered as possible explanations for the acoustic dissimilarity between the two corpora. Firstly, each corpus followed a different method of data collection and analysis. For example, formant values for the AusTalk data were extracted at 50% of the duration in line

with previous studies [7-9], whereas the values for the Cox corpus were taken at the target in the vowel which was designated at the point of least formant change [1]. It is possible that the differences we observe in the front vowels could be because the target for these front vowels was not around the midpoint. Or it may be that the acoustic overlap in some of the front vowels in the Cox corpus affected the classifications of the vowels in the discriminant analyses. Further analyses are required using the same methods to confirm that these acoustic differences are not simply a result of different methods of data collection and analysis.

In addition, regional variation could also account for the acoustic differences across the two corpora. Previous acoustic analyses [14-16] have shown regional variation for vowel production by adolescent AusE speakers which even extends to speakers from different parts of Sydney (i.e., Northern Beaches, Northern Suburbs and Western Sydney) [16]. The speakers in [1] are a homogeneous group from Sydney's Northern Beaches. In contrast, the speakers in the AusTalk corpus were from various locations across Sydney, as it was not possible to find four monolingual speakers from the same region within Sydney among the speakers recorded within this corpus. Therefore, future comparisons controlling for age, methodology and regional accent would be necessary to determine the influence of region on native vowel production.

While we cannot draw definite conclusions due to the aforementioned reasons, a difference between corpora has important implications for our ongoing research on L2 vowel perception by Western Sydney undergraduates. Recall the L2LP model [6] which explains that detailed acoustic comparisons of the native and target language are needed for accurate predictions of non-native and L2 difficulty. However, for accurate predictions, L2LP posits that acoustic analyses must be based on the same target population for which the predictions are intended.

The fact that the two corpora from this study are different from each other and we cannot confirm whether these differences are caused by differing methodologies, age or region, it is highly likely that their acoustic values will be different from those in the vowels of another group of AusE speakers. If we are to follow L2LP model and aim at making accurate predictions about Western Sydney speaker's non-native vowel learning, a corpus from this population is required, as it is likely that their vowels will not have values that are similar to those reported in either the Cox or AusTalk corpora.

To this end, we are currently collecting data for a new corpus of Australian English vowels from Western Sydney monolingual speakers, using the same methods as in previous studies [7-9] to ensure cross-linguistic comparability. This new corpus will not only allow for future cross-linguistic, dialectal and regional comparisons, but it will also serve as a reference point for cross language perception, word recognition and production studies on Australian English speakers from Western Sydney.

5. References

[1] Cox, F., "The Acoustic Characteristics of /hVd/ Vowels in the Speech of some Australian Teenagers", Australian Journal of Linguistics, 26(2), 147–179, doi:10.1080/07268600600885494, 2006.

[2] Burnham, D., Estival, D., Fazio, S., Viethen, J., Cox, F., Dale, R., & Hajek, J., "Building an Audio-Visual Corpus of Australian English: Large Corpus Collection with an Economical Portable and Replicable Black Box", in INTERSPEECH, pp. 841-844, 2011.

[3] Mitchell, A. G.,"The pronunciation of English in Australia", Sydney, Angus and Robertson, 1946.

[4] Cox, F., & Palethorpe, S., "Australian English". Journal of the International Phonetic Association, 37(03), 341–350, doi:10.1017/S0025100307003192, 2007.

[5] Harrington, J., Cox, F., & Evans, Z., "Australian Journal of An acoustic phonetic study of broad, general, and cultivated Australian English vowels", Australian Journal of Linguistics, 17(2), 155–184, 1997.

[6] Escudero, P. "Linguistic Perception and Second Language Acquisition", PhD Dissertation, Utrecht University, 2005

[7] Van Leussen, J.-W., Williams, D., & Escudero, P., "Acoustic properties of Dutch steady-state vowels: Contextual effects and a comparison with previous studies", 1149–1197, 2011.

[8] Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. a H., "A cross-dialect acoustic description of vowels: Brazilian and European Portuguese", Journal of the Acoustical Society of America, 126(3), 1379–93. doi:10.1121/1.3180321, 2009.

[9] Chládková, K., Escudero, P., & Boersma, P., "Context-specific acoustic differences between Peruvian and Iberian Spanish vowels", Journal of the Acoustical Society of America, 130(1), 416–28, doi:10.1121/1.3592242, 2011.

[10] Peterson, G., & Barney, H., "Control methods used in a study of the vowels", Journal of the Acoustical Society of America, 24, 175-184, 1952.

[11] Strange, W., Bohn, O.-S., Trent, S. a., & Nishi, K. "Acoustic and perceptual similarity of North German and American English vowels", Journal of the Acoustical Society of America, 115(4), 1791. doi:10.1121/1.1687832, 2004.

[12] Escudero, P., & Vasiliev, P., "Cross-language acoustic similarity predicts perceptual assimilation of Canadian English and Canadian French vowels", Journal of the Acoustical Society of America, 130(5), EL277–EL283, 2011.

[13] Morrison, G., & Escudero, P., "A cross-dialect comparison of Peninsula-and Peruvian-Spanish vowels", 2007.

[14] Cox, F., & Palethorpe, S., "The border effect: Vowel differences across the NSW/Victorian border", in Proc. 2003 Conference, Australian Linguistics Society, 2004.

[15] Butcher, A. R., "Formant frequencies of/hVd/vowels in the speech of South Australian females", in Proceedings of the 11th Australasian International Conference on Speech Science & Technology, 449-453, 2006.

[16] Cox, F., & Palethorpe, S., "Regional variation in the vowels of female adolescents from Sydney", in ICSLP, 1998