AN ACOUSTIC COMPARISON BETWEEN AMERICAN ENGLISH AND AUSTRALIAN ENGLISH VOWELS

Kimiko Tsukada

Division of Speech and Hearing Sciences
University of Alabama at Birmingham (UAB)
tskudaca@uab.edu

ABSTRACT

Five vowels /i, e, æ, a (or o), u/ in isolated /CVC/ words produced by 7 American English (AmE) and 6 Australian English (AusE) talkers were examined with a view to documenting acoustic-phonetic similarities and differences between the two accent types. The effect of Accent was significant on all vowels for at least one of the first two formants. The AusE vowel space was much more compressed relative to the AmE vowel space. In both varieties, however, within-category clustering was quite good, suggesting that these talkers’ productions were fairly consistent within their respective vowel systems. Although the two groups differed considerably in vowel quality, they were quite similar in vowel duration.

1. INTRODUCTION

It would be fair to state that American English (AmE) and Australian English (AusE) give distinct auditory impressions to a great majority of native English-speaking listeners and to some second language (L2) learners who have experience with these varieties of English. One would expect that there are acoustic phonetic cues encoded in the speech signal that listeners use to categorize talkers’ membership to a specific speech community [3]. The purpose of this study is to determine if there are measurable acoustic differences between AmE and AusE in their vowel quality and duration. Although many recent studies have offered detailed phonetic descriptions of vowel systems in AmE [1, 9, 12, 13] and AusE [4, 5, 6, 10, 11, 14, 15], cross-dialectal comparisons between the two are rather rare. The present study thus aims to provide a direct acoustic comparison between AmE vowels and AusE vowels. The question addressed here is whether there are any systematic differences in five monophthongs /i, e, æ, a (or o), u/ in /CVC/ words produced by male AmE and AusE talkers.

Table 1. Target words and the number of vowel tokens produced by AmE and AusE talkers.

<table>
<thead>
<tr>
<th>accent</th>
<th>feet</th>
<th>bed</th>
<th>bat</th>
<th>pot</th>
<th>food</th>
</tr>
</thead>
<tbody>
<tr>
<td>AmE</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>AusE</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
</tbody>
</table>

While acoustic comparisons between different languages or dialects are of interest to phoneticians in their own right, gaining a knowledge of dialectal differences of a language has implications for topics such as L2 learning [8] and pronunciation modeling for automatic speech recognition. It is usually the case that L2 learners find it difficult to understand the target language when it is spoken with an unfamiliar accent even if the talker is a native speaker of that language. However, due to an increased flow of people and information in present-day society, there are more opportunities for native and non-native talkers with different accents to come into contact. Reduced intelligibility due to accent difference is probably inevitable even among native speakers of English, but it appears to have a greater and more adverse impact on non-native listeners. A better understanding of distinctive phonetic characteristics of different varieties of English will facilitate L2 learners’ encounter with unfamiliar accents and help them to gain a realistic experience in their target language. It is quite conceivable that not only human listeners but also speech recognition systems benefit from advanced knowledge of dialectal variations within the same language.

2. METHOD

2.1. Talkers and speech materials

Five monophthongs /i, e, æ, a (for AmE) or o (for AusE), u/ were drawn from two separate data sets representing AmE and AusE. Five words ending in the same consonant (/θ/ or
/d/) were chosen from each talker group (see Table 1). The AmE group included 7 male talkers who were recorded in 4 different states (Alabama - 1, California - 3, Illinois - 2, Texas - 1) in the US. Pictures were presented in random order to elicit the target words from the AmE talkers. They served as control subjects in a larger developmental study and were recorded at two times separated by one year. As it is unlikely that adult native talkers’ speech shows a noticeable change over a period of one year, their data from two times were combined in the present study.

Unlike AmE, AusE is a non-rhotic variety like British English RP (Received Pronunciation). A review of previous studies [4, 5, 6, 10] informs us that AusE is conveniently categorized into three accent types: Broad, General and Cultivated. Cultivated AusE is perceived to have close resemblance to RP and Broad AusE to London Cockney English. General AusE lies between these two extremes and is spoken by the majority of the population. In this study, six male AusE talkers were recorded individually in a sound-treated studio in the Speech, Hearing and Language Research Centre (SHLRC) at Macquarie University in Sydney, Australia. The accent type of these talkers could be regarded as being intermediate between Cultivated and General. The AusE talkers read the target words 7 times as they were displayed on a computer screen one word at a time in random order.

### 2.2. Data processing and analysis

The recorded data were digitized at 20 kHz. Formants were calculated using a signal processing package ESPS/Waves (12th order LPC analysis) and phonetic boundaries for the vowels of interest were marked on the basis of waveform and spectrographic information available in the EMU speech database system [2]. Segmentation was done according to the criteria set in [7]. The R environment was used to retrieve formant frequencies at the vowel’s temporal midpoint and to plot these data points in the F1/F2 formant plane.

### 3. RESULTS

#### 3.1. Vowel quality

<table>
<thead>
<tr>
<th>Vowel</th>
<th>AmE</th>
<th>AusE</th>
<th>AmE</th>
<th>AusE</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i</td>
<td>290 (35)</td>
<td>340 (35)</td>
<td>2341 (116)</td>
<td>1934 (166)</td>
</tr>
<tr>
<td>/ɛ</td>
<td>576 (48)</td>
<td>430 (43)</td>
<td>1821 (82)</td>
<td>1836 (117)</td>
</tr>
<tr>
<td>/æ</td>
<td>775 (57)</td>
<td>676 (39)</td>
<td>1733 (120)</td>
<td>1610 (81)</td>
</tr>
<tr>
<td>/ɑ/</td>
<td>779 (52)</td>
<td>613 (53)</td>
<td>1216 (74)</td>
<td>1033 (52)</td>
</tr>
<tr>
<td>/u</td>
<td>327 (38)</td>
<td>300 (23)</td>
<td>1325 (127)</td>
<td>1627 (91)</td>
</tr>
</tbody>
</table>

Table 2. The mean F1 and F2 values (in Hz) in AmE and AusE. Standard deviations are in parentheses.

![Fig. 1. Ellipse plots of the 5 vowel types for the AmE and AusE data. For data having a Gaussian (normal) distribution, the radius of the ellipse is 2.45 times the standard deviation of the mean, covering approximately 95% of the data points. The centroids are the average values of those distributions. In AusE, /ɒ/ indicates /ɒ/.](image)

It is also clear that the /i, ɛ, u/ vowels in AusE are poorly separated compared to the same vowel types in AmE. /æ, ɛ, ɒ/ are much higher in the AusE vowel space than in the AmE vowel space. In fact, the use of the IPA symbol [e] is proposed for an AusE vowel in “head” [10]. The centroid of the AusE /æ/ is intermediate between the AmE /æ/ and the AmE /æ/ in the F1 values. The AusE /i/ is more retracted and /u/ more fronted with respect to the corresponding AmE vowels. In AusE, these two vowel categories show some overlap. The AmE /u/ is most closely aligned with /u/ along the F2 axis whereas the more fronted AusE /u/ is aligned with /æ/ and the AusE /ɔ/ in the F1 values. The AusE /i/ may not have reached its target yet, as the formants were tracked at the vowel’s midpoint. It is known that the AusE /i/ has a long onglide starting at lower position in the vowel space and reaching its most extreme values later in the vowel [4, 10, 14].

While the AusE vowel space is much more compressed relative to the AmE vowel space, it is questionable whether
this difference can be attributed to the different tasks that the two groups of talkers performed. Both the AmE and AusE groups produced isolated /CVC/ words in similar phonetic contexts. Despite the different regions occupied by the five vowel categories in their respective vowel space, the two groups do not appear to differ substantially in the extent of within-category variation.

Although the AmE talkers are not geographically homogeneous, their vowel tokens show a reasonable clustering within each category and there is hardly any overlap between the categories. The formant values are generally in modest to good agreement with those reported in previous studies [1, 9, 13]. In another recent study [12], the /e/ and /æ/ vowels showed a very high degree of overlap when only F1 and F2 values at steady state were considered. This is in sharp contrast with the present results. In particular, the /æ/ vowel in this study has a much higher mean F1 value (by 187 Hz) and a lower mean F2 value (by 219 Hz) compared to the corresponding values in [12]. [9] also noted a big difference in the formant values for /æ/ between [12] and his study on California English vowels. The AmE /u/ in this study was more fronted and its mean F2 value was closer to the value reported in [1, 9, 13] than in [12]. It is noteworthy that the AusE /u/ is even more fronted than any group of the AmE talkers reviewed here. This is consistent with the finding in previous studies [4, 5, 10, 11], i.e., the fronted or centralized /u/ is a clear marker of AusE, keeping it distinct from other varieties of English.

Two-way ANOVA with Accent as the between-subject factor and Vowel Type as the within-subject factor was conducted separately for F1 and F2 values. The dependent variables were each talker’s mean F1 and F2 values.

For F1, both main factors reached significance \(F(1, 11) = 54.1, p < 0.001\) for Accent and \(F(4, 44) = 410.622, p < 0.001\) for Vowel Type. The Accent x Vowel Type interaction was also significant \(F(4, 44) = 20.895, p < 0.001\). The simple effect of Accent was significant for /ɛ, æ, a (or o)/ as is clearly seen in Figure 1. These vowels had higher F1 values in AmE than in AusE. The simple effect of Vowel Type was significant for both AmE and AusE. Post-hoc Scheffé showed that, in AmE, vowels were divided into 3 groups in terms of height. High vowels /i/ and /u/ did not differ significantly. /e/ was the only mid vowel which differed significantly from high vowels /i, u/ and low vowels /æ, a/, which did not differ from each other. For AusE, too, /i/ and /u/ did not differ significantly in F1. These two had lower F1 values than /e/ which in turn had lower F1 than /æ, a/. /æ/ had the highest F1 of the five vowels in AusE.

For F2, Vowel Type \(F(4, 44) = 306.537, p < 0.001\) and Accent x Vowel Type \(F(4, 44) = 36.524, p < 0.001\) were significant, but Accent was not \(F(1, 11) = 3.248, p = 0.09\). The simple effect of Accent was significant for /i, æ, o/ (or /u/), /u/. /e/ and /æ/ had higher values in AmE than in AusE and the reverse was the case for /u/. The simple effect of Vowel Type was significant for both AmE and AusE. F2 of AmE vowels had 3 levels in back-front dimension. /i/ was the highest and most front vowel. /e/ and /æ/ did not differ significantly and nor did the two back vowels /u/ and /æ/. For AusE, /i/ and /æ/ did not differ in F2. These two vowels had significantly higher F2 values than /æ/ and /u/ which did not differ from each other. The /u/ vowel, as was noted earlier, is not a back vowel in AusE. /o/ had the lowest F2 of the five vowels.

### 3.2. Vowel duration

<table>
<thead>
<tr>
<th>vowel</th>
<th>AmE</th>
<th>AusE</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>138 (14)</td>
<td>155 (22)</td>
</tr>
<tr>
<td>e</td>
<td>199 (26)</td>
<td>181 (27)</td>
</tr>
<tr>
<td>æ</td>
<td>197 (33)</td>
<td>157 (27)</td>
</tr>
<tr>
<td>o/æ</td>
<td>152 (35)</td>
<td>126 (23)</td>
</tr>
<tr>
<td>u</td>
<td>232 (44)</td>
<td>265 (43)</td>
</tr>
</tbody>
</table>

Table 3. The mean vowel duration (in ms) in AmE and AusE. Standard deviations are in parentheses.

Table 3 shows the mean vowel durations for AmE and AusE. Contrary to a large accent difference in the spectral domain, vowel duration in the two talker groups is fairly similar. The order of vowel duration from shortest to longest was /i/ < /æ/ < /æ/ < /æ/ < /æ/ in AmE and /æ/ < /æ/ < /æ/ < /æ/ < /æ/ in AusE. /æ/ was the longest vowel in both. A largest mean durational difference was observed for the /æ/ vowel at 40 ms. The relatively long duration of /æ/ and /æ/ is presumably an influence of the following voiced stop /d/.

Two-way ANOVA with Accent as the between-subject factor and Vowel Type as the within-subject factor was conducted. Each talker’s mean durational value was used as the dependent variable. Vowel Type x Accent \(F(4, 44) = 7.509, p < 0.001\) were significant, but Accent was not \(F(1, 11) = 0.333, p = 0.5754\). The simple effect of Accent was significant only for /æ/, which was longer in AmE than in AusE. The simple effect of Vowel Type was significant for both AmE and AusE. Post-hoc Scheffé test showed that, in AmE, two vowels /æ/ and /æ/, both of which preceded /d/, did not differ significantly. The other 3 vowels /i, æ, o/ formed one group and did not differ significantly. /æ, æ, æ, æ, æ, æ/ formed another group, but /æ/ was significantly longer than /æ/. /æ/ was significantly longer than the other four vowels.

### 4. DISCUSSION

We observed many phonetic differences between AmE and AusE vowel systems. Do these differences lead to reduced
intelligibility and interfere with efficient communication? It seems unlikely that, between native talkers of AmE and AusE, mutual intelligibility is entirely lost. It would be more plausible to think that there is some kind of perceptual normalization process that the AmE and AusE listeners employ. To understand how this mechanism operates, it is necessary to examine how AusE listeners assimilate AmE vowels to their own vowel categories and vice versa.

Given these clear differences in the phonetic realization of vowel categories examined here, another question of interest would be if one group of talkers perceive the other group to be foreign-accented. Alternatively, do they simply hear vowels produced by the other group to come from native English talkers of some other dialect? If such results were obtained, a possible interpretation might be that making judgements on foreign accent involve other acoustic cues in addition to or instead of the parameters examined here. It would be quite useful for L2 learners to understand how non-native accent is differentiated from native accent.

In this study, acoustic comparison was made only with respect to static samples of the formant pattern, i.e., formant frequency values taken at the vowel’s midpoint. However, a possibility remains that the AmE and AusE vowels exhibit a marked difference in the extent of dynamic formant movement and this may give rise to different auditory impressions between the two. There is evidence that AusE diphthongs, but not monophthongs, significantly benefit from the inclusion of dynamic spectral information, resulting in increased category separability [11, 15]. It may be the case that, at least for some vowel types, AusE vowels show less degree of formant movement than AmE vowels do.

5. SUMMARY

We observed that AmE and AusE vowels differ significantly in vowel quality while the difference in vowel duration was of limited magnitude. The two groups also differed in the degree of dispersion/compression for the 5 vowels in the F1/F2 plane. It is likely that some of these acoustic phonetic differences are related to the auditory impressions associated with the two varieties of English.

An increase in people’s mobility in today’s world makes it difficult to choose the absolute model which language learners should try to acquire. It would be more beneficial if a detailed knowledge of a variety of models were made available. This study provided two types of “normative” data against which the speech of L2 learners can be compared.

6. REFERENCES