Perception and Production of /i:/, /ɑ/ and /e:/ in Australian English

Robert H. Mannell

Centre for Language Sciences, Department of Linguistics, Macquarie University, Sydney, Australia
robert.mannell@mq.edu.au

Abstract

The progressive offglide reduction of /i:/ and /e:/ or onglide reduction of /i:/ in Australian English is the major focus of this study. This paper examines, using synthetic speech tokens, the patterns of vowel perception of female and male speakers of Australian English in (approximately) 1990 and 2007. The relationship between production and perception in 2007 is also examined. This paper provides evidence that monophthongisation of /e:/ precedes that of /i:/.

1. Introduction

The aim of this paper is to examine vowel change in Australian English for three vowels /i:/, /ɑ/ and /e:/ and to especially examine diphthongal (onglide – offglide) patterns and how these patterns affect the perception of these vowels. The degree of /i:/ onglide (delayed target) has been discussed in various papers and, for example, [1] suggest that younger speakers show a lesser degree of /i:/ onglide than older speakers. [2] examines vowel change in Australian English and indicates complex patterns of change in offglide patterns for /ɑ/ and /e:/ which affect the degree or even presence of offglide for these vowels in various contexts. This paper examines patterns of male-female difference in the perception and production of these three vowels over a 17 year period. It also examines the hypothesis that as one vowel becomes more diphthongal, a vowel with similar fronting, height and duration may maintain its perceptual distinction from this vowel by becoming less monophthongal (or vice versa). For example, /i:/ may be becoming more monophthongal for some Australian speakers but it is possible that these same speakers maintain a perceptual distinction between /ɑ/ and /i:/ because they have an /i:/ with a distinct onglide.

2. Experiment 1: Perceptual Vowel Space

2.1. Methodology

2.1.1. Perceptual Tokens

A vowel formant space was modelled with the following characteristics. Firstly, two planes were defined: a short vowel plane and a long vowel plane with vowel lengths set at 150 ms and 300 ms respectively (approximately following the average short and long vowel lengths for /hVd/ citation tokens in [3]). The extreme F1 and F2 values for these planes were derived from the vowel frequency distribution for citation /hVd/ productions by all speakers from [3]. The tokens were synthesised using a purpose-built software parallel formant speech synthesiser based on the synthesiser described by [6]. 112 vowels were defined for each of the long and short vowel planes. These vowels had F1 and F2 values that were all possible multiples of 100 Hz that fit within the defined male vowel space boundary (thick unbroken line approximating the vowel triangle), and also the points on the boundary, represent the F1 and F2 values for each of the 112 tokens for that plane. The same vowel tokens, but with different duration, are defined for the other vowel plane, giving a total of 224 tokens. F3 was specified as a simple function of F2 (appropriate for all vowels except, perhaps, /u/) using the average F3 values for citation /hVd/ tokens from [3]. F4 and F5 were fixed at typical male values (3500 and 4500 Hz). Appropriate coarticulation for the preceding /h/ and the following /d/ was provided using a locus space (“locus matrix”) model, after [4]. The fundamental frequency was centred on 110 Hz. The tokens were randomised and presented to subjects at 5 second intervals. A short training sequence preceded the test materials to accustom the subjects to the voice quality. The same synthesised tokens were used for the 1990 and the 2007 experiments.

2.1.2. Experimental Sessions and Participants

Between November 1989 and November 1991, 20 young monolingual phonetically-untrained speakers of Australian English (7 female, 13 male, age mean 18.5, age std.dev. 2.40) provided orthographic responses to the 224 token dataset described in the previous section. This group is referred to as the 1990 group.

Two groups of 30 young monolingual phonetically-untrained speakers of Australian English participated in a repeat of the 1990 experiment during 2006-2007. One of these groups consisted of 30 female participants (age mean 26.6, age std.dev. 11.03). This group is referred to below as the 2007 female (or 2007-F) group. The second group consisted of 30 male participants (age mean 22.6, age std.dev. 7.03). This group is referred to below as the 2007 male (or 2007-M) group.

2.1.3. Experimental Design and conditions

Subjects were presented the randomised materials in a sound treated room over high quality headphones at 70 dB SPL. Subjects were given a list of possible orthographic responses (a list of h d words and non-word) and were asked to respond using those words. Blank responses were discouraged but not prohibited. All responses were recoded...
into vowel phoneme symbols where possible or recorded as uninterpretable (and excluded from the analysis).

Subjects all passed a simple hearing screening test and no subjects reported hearing problems.

2.2. Results

Figures 1, 2 and 3 display perceptual responses to long vowel tokens by monolingual Australian English speakers tested in 1990 (figure 1: mixed group of males and females) and 2007 (figure 2: females only, and figure 3: males only).

A maximal male vowel space, accounting for the normal spread of vowel targets in the speech of a modeled average sized male speaker (generated by a formant synthesiser), is indicated by the irregular vowel polygon (approximately a "vowel triangle") surrounding the vowel space data. The points at the intersection of the x-axis and y-axis grids (dotted lines at multiples of 100 Hz) and on or inside the "vowel triangle" polygon, represent the F1/F2 values of the tokens presented to the experimental participants. The contours drawn inside the vowel polygons in these figures are "predominance boundaries" [5] which surround data points for which 50% or more of subjects perceived the indicated vowel. The dot within each of these regions is the grand mean of the individual subject means. The skewed nature of the predominance regions for many of these vowels explains why the grand means are sometimes not aligned with the centre of the predominance boundary. Nearly all vowels on the edge of the vowel space have their regions of greatest perceptual certainty adjacent to those edges. The grand means are derived from a scatter of individual subject means and nearly all patterns of individual means conform to a normal distribution in both the F1 and F2 dimensions and so provide a convenient way for statistically comparing measures of central tendency between vowel spaces. Sometimes a vowel's grand mean may appear outside its predominance boundary (e.g. /i/ in figure 2) as the mean is also affected by responses outside the predominance boundary.

A t-test analysis of grand means was carried out. Pairs of same vowel identifications were compared across all three pairings of 1990 (M+F), 2007 (M) and 2007 (F) datasets. There were no significant (p < 0.01) differences between the paired grand means for /i/, /æ/ and /e/.

A chi-square analysis was carried out for each data point, paired across all combinations of the three perceptual spaces. Across each of the three paired vowel space comparisons the number of significantly different points varied between 1 and 6 points out of a total of 112 points (for the long vowel space). Only the following significantly different (p < 0.01) vowel pairs are relevant to the current analysis (i.e. data points in the region where the vowels of interest occur). The comparison between 2007-F and 1990 long vowel space has two significant differences between /æ/ vowel responses (1: F1 = 300 & F2 = 2100, 2: F1 = 300 & F2 = 2300). The comparison between 2007-F and 2007-M long vowel space has only one significantly different data point and it is also a difference between /æ/ vowel responses (1: F1 = 200 & F2 = 2300). These results suggest a difference between /æ/ vowel perception for comparisons between 2007-M and 2007-F perceptual patterns and between 2007-F and 1990 perceptual patterns, which further suggests the possibility of change over time and/or differences between male and female perceptual patterns. These results need to be treated with caution for two reasons:-
3. Experiment 2: Perception vs Production

3.1. Methodology

Over the period 2006 to 2007 perceptual data was collected from 30 male and 30 female young native speakers of Australian English. This procedure is reported above.

Additionally, all 60 participants were recorded reading a list of hVd, hVt or hV real word tokens (heed, hid, head, had, hard, hut, hot, hoard, hoed, who'd, heard, hear, hair).

3.1.2. Perception Data

The audio recordings took place in a sound treated room. F1, F2 and F3 were hand analysed for each vowel and F4 was additionally analysed for "heard". Additionally, the onset F1 and F2 of "heed" and the offset F1 and F2 of "hear" and "hair" were also analysed. Onset and offset values were measured as close to the true onset or offset as was possible without sacrificing the measurability of the formants (e.g. as a consequence of heavy creaky voice at the offset). Finally, the offglides for /i/ and /e/ were impressionistically classified as monophthongal, weakly diphthongal, strongly diphthongal or rhoticised. Rhoticised vowels are excluded from the following analyses of those vowels. Four participants (3 female, 1 male) produced audibly rhoticised /i/ and /e/.

The offglides of /i/ were, for each token, converted to a single distance (in F1/F2 Hertz space) between the F1 and F2 of the onset and the F1 and F2 of the target. Similarly, the offglides of /i/ and /e/ were, for each token, converted to a single distance (in F1/F2 Hertz space) between the F1 and F2 of the target and the F1 and F2 of the offset.

3.1.2. Perception Data

In experiment 1, perception data analysis relied heavily on impressionistic visual analysis of vowel space images as well as simple measure of central tendency and space-by-space data point comparison. For the present experiment a different approach to quantifying the perception data was carried out. In a visual examination of the long and short vowel perceptual spaces it was found that the predominance of the target and the F1 and F2 of the offset.

The onglides of /i/ were, for each token, converted to a single distance (in F1/F2 Hertz space) between the F1 and F2 of the onglide and /g1852/g1317/ and /e/g1949/ and /e/ fall within these boundaries. The previous analysis, particularly grand-means based on subject means, were based equally on each subject regardless of the relative strength (high number perceived) or weakness (low number perceived) of their perception of each vowel. A simple way of avoiding this problem is to simply count the number of tokens within this sub-space that are perceived as each of these vowels. The reason for including the two lax vowels is that they are also adjacent vowels across the long-short vowel planes and a few /i/ and /e/ vowels were perceived in this sub-space of the long vowel plane. This analysis resulted in tallies for each subject by each vowel across the 2007-F and 2007-M datasets.

3.2. Results

3.2.1. Impressionistic Analyses

Table 1. Impressionistic analyses of /i/ and /e/

<table>
<thead>
<tr>
<th></th>
<th>2007-F</th>
<th>2007-M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/i/</td>
<td>/e/</td>
</tr>
<tr>
<td>Monophthong</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Weak Diphthong</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Strong Diphthong</td>
<td>19</td>
<td>11</td>
</tr>
<tr>
<td>Rhoticised</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

ANOVA of the above data showed only one relative weak male/female difference for /e/ (F(1,54) = 5.15, p = .027).

A Pearson Correlation analysis showed highly significant correlations between the impressionistic analysis and the acoustic F1/F2 offglide distance of /i/ (r(55) = .853, p = .000) and of /e/ (r(55) = .826, p = .000). Analyses of the impressionistic categorizations (for both males and females) showed that all subjects that are monophthongal for /i/ are also monophthongal for /e/ (but not the converse) and all subjects that are weakly diphthongal for /i/ are also either monophthongal or weakly diphthongal for /e/ (but not the reverse). This pattern suggests that monophthongisation of /e/ precedes that of /i/.

3.2.2. Production Acoustics

Table 2. Acoustic analyses of /i/ onglide and /i/ and /e/ offglides ("tr" = target, "o" = onset or offset, df12 = frequency difference on F1/F2 plane)

<table>
<thead>
<tr>
<th></th>
<th>2007-F</th>
<th>2007-M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/i/</td>
<td>/e/</td>
</tr>
<tr>
<td>tr F1 mean</td>
<td>310</td>
<td>320</td>
</tr>
<tr>
<td>tr F1 s.d.</td>
<td>72</td>
<td>44</td>
</tr>
<tr>
<td>tr F2 mean</td>
<td>2810</td>
<td>2730</td>
</tr>
<tr>
<td>tr F2 s.d.</td>
<td>166</td>
<td>160</td>
</tr>
<tr>
<td>o F1 mean</td>
<td>335</td>
<td>702</td>
</tr>
<tr>
<td>o F1 s.d.</td>
<td>96</td>
<td>270</td>
</tr>
<tr>
<td>o F2 mean</td>
<td>2653</td>
<td>1898</td>
</tr>
<tr>
<td>o F2 s.d.</td>
<td>228</td>
<td>345</td>
</tr>
<tr>
<td>dt12 mean</td>
<td>196</td>
<td>936</td>
</tr>
<tr>
<td>dt12 s.d.</td>
<td>123</td>
<td>448</td>
</tr>
</tbody>
</table>

ANOVA indicates a significant male/female difference in the acoustic extent of the offglides of both /i/ (F(1,58) = 14.33, p = .000) and /e/ (F(1,58) = 7.68, p = .007) with, in both cases, the female data showing the higher mean extent of the offglide. There is a similar, but weaker, pattern for the /i/ onglide (F(1,58) = 6.79, p = .012). That is, females show a stronger pattern of diphthongisation. These differences may be due (at least in part) to the larger size of the female vowel space. This analysis needs to be repeated on either auditorily scaled or normalised data.
3.2.3. Perception

Table 3. Tally of token perception of /i/., /a/ and /e/.

<table>
<thead>
<tr>
<th></th>
<th>2007-F</th>
<th>2007-M</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>5.07</td>
<td>6.50</td>
</tr>
<tr>
<td>/a/</td>
<td>6.30</td>
<td>8.73</td>
</tr>
<tr>
<td>/e/</td>
<td>6.63</td>
<td>6.50</td>
</tr>
</tbody>
</table>

ANOVA indicates a significant (F(1,58) = 12.53, p = 0.001) male/female perceptual difference for the /a/ vowel (stronger /a/ perception pattern for females), but no significant difference for the other two vowels.

There were negative correlations between /i/ and /a/ (r(59) = -.526, p = .000) and between /e/ and /a/ (r(59) = -.694, p = .000). These are mutually exclusive choices (e.g. if you choose /i/ for a token then you don't choose /a/).

3.2.4. Perception and Production

Pearson Correlation analysis of pooled male and female data shows that the only significant (p < 0.01) correlation between production data (glide distance in Hz) and perception data is a positive correlation between strength of /e/ perception and production data (glide distance in Hz) and perception data is a positive correlation between strength of /e/ perception and degree of /e/ ofglide production (r(59) = .366, p = .004).

A separate examination of the Pearson Correlations for male and female data confirms this correlation only for female subjects. It is important to note that all perception tokens were hVd, whilst the production tokens for /e/ and /a/ were hV and the production tokens for all other vowels were hVd.

3.3. Discussion

In this experiment evidence was found for a strongly significant relationship between impressionistic classification of, and acoustic evidence for, degree of vowel ofglide (diphthongisation) for /i/ and /e/.

Additionally, impressionistic analysis of patterns of diphthong-monophthong status of /a/ and /e/ suggested that monophthongisation of /e/ precedes that of /a/.

Females were found to produce significantly stronger /a/, /e/ ofglides and /i/ onglides than males.

A significant pattern of mutually exclusive vowel choices was found for /a/ and /i/, /e/ and /e/. That is, when the perceptual pattern of one vowel changes it affects the perceptual pattern of similar vowels. In the perception task a choice must be made between competing alternative vowels. For example, a long vowel with F1 = 300 and F2 = 2300 will most likely be identified as either /i/ or /a/.

As the perceptual tokens are monophthong, we might assume that a subject will choose vowels according to his/her expectations of which tokens are monophthongal, we might assume that a subject will choose vowels according to his/her expectations of which tokens are monophthongal /e/). This is equivalent to a negative hypothesis that listeners base their perceptual choices upon their own production. It is possible that the perceptual choices made here are based upon personal patterns of hVd production but that the collected hVd /e/ production data isn't a good guide to hVd /e/ production. It has previously been observed [1] that some subjects have monophthongal productions of /a/ and /e/ in hVd context but diphthongal or bisyllabic productions in hV context.

4. Experiment 3: Differences in Male and Female Perception for 1990 and 2007

4.1. Methodology

By combining perception data from two separate 1990 vowel perception experiments it was possible to create two separate groups of 20 male (1990-M) and 20 female (1990-F) subjects.

The same analysis as described in section 3.1.2 was carried out. These results were compared with the 2007-M and 2007-F datasets (N= 30) described in sections 2.1.2 and 2.3.2.

4.2. Results

Table 4. Tally of token perception of /i/., /a/ and /e/.

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>6.00</td>
<td>6.45</td>
</tr>
<tr>
<td>/a/</td>
<td>4.80</td>
<td>2.95</td>
</tr>
<tr>
<td>/e/</td>
<td>7.10</td>
<td>6.20</td>
</tr>
</tbody>
</table>

An ANOVA analysis indicated significant differences for /a/ (F(3,96) = 6.313, p = 0.001). A post hoc Tukey analysis showed significant (p < 0.01) differences, for /a/ only, between (a) 2007-F and 1990-M, and (b) 2007-F and 2007-M but not between 1990-F and the male data. This indicates a significant change in male-female perceptual differences between 1990 and 2007.

5. Conclusion

This paper provides evidence that monophthongisation of /e/ precedes that of /a/. Females show a stronger pattern of ofglide production for /a/ and /e/ (in hVd context) and onglide production for /i/ than males. Females, but not males, show a significant negative correlation between /e/ perception and production patterns. Females also show significantly stronger degrees of /a/ monophthong perception in hVd contexts than males, and there is also evidence for a significant change in the pattern of male-female difference between 1990 and 2007. Future work will need to collect both hVd and hV production data for these vowels as well as perception data and ideally the perception experiments should also include both hVd and hV tokens.

6. References