Australian Journal of Linguistics

ISSN: 0726-8602 (Print) 1469-2996 (Online) Journal homepage: http://www.tandfonline.com/loi/cajl20

An acoustic comparison between New Zealand and Australian English vowels

Catherine I. Watson , Jonathan Harrington & Zoe Evans

To cite this article: Catherine I. Watson , Jonathan Harrington & Zoe Evans (1998) An acoustic comparison between New Zealand and Australian English vowels , Australian Journal of Linguistics, 18:2, 185-207, DOI: 10.1080/07268609808599567

To link to this article: https://doi.org/10.1080/07268609808599567

Published online: 14 Aug 2008.

An Acoustic Comparison between New Zealand and Australian English Vowels*

CATHERINE I. WATSON, JONATHAN HARRINGTON AND ZOE EVANS

This study presents an acoustic comparison between New Zealand English (NZE) and Australian English (AE) citation-form monophthongs and diphthongs. Formant frequencies were calculated and the data were labelled for vowel target position. Four main kinds of analysis were carried out: F1/F2 formant plots of monophthongs, onglides in vowels exemplified by HEED and WHO'D, formant trajectories of rising diphthongs, and formant trajectories of falling diphthongs. Consistent with other studies, we find the major differences in the NZE and AE vowel spaces are the centralizing and lowering of NZE HID and HOOD, and the raising of the NZE front and high vowels. The major difference in the diphthongs is the HERE/HAIR merger occurring in NZE but not AE. Contrary to other studies we find both HEED and WHO'D in NZE have long onglides and that the extent of these onglides is similar to those in AE. Proposals are included for a modification to the transcription system of NZE.

1. Introduction

It has been documented for over 100 years that New Zealand and Australian English are remarkably similar (e.g. Gordon 1991; Samuel McBurney 1887, reproduced in Turner 1967) but it is also well known that there are distinctive differences in their vowel qualities. Several features that characterize both modern New Zealand and Australian speech were noted in the early part of this century. These include: the shift in the first target of the HAY and HIDE diphthongs from those in the British Received Pronunciation (Bauer 1986); the lowering of the first target in the HOE diphthong; and the raising and fronting of the first target of the HOW diphthong

* We are indebted to Nik Kasabov of the University of Otago for making the speech database available to us. The NZE database was recorded at the University of Otago by the first author, Diana Kassabova, Richard Kilgour, Steven Sinclair, and Mark Laws. We also thank Margaret Maclagan of the University of Canterbury who compiled the word list for the NZE database and for her helpful comments in this study. Finally, we thank the members of SHLRC for their helpful comments and discussion.

0726-8602/98/020185-23

(Bauer 1986, 1994).[1] The similarity remains today and is well documented (see, for example, Bauer 1994; Maclagan 1982; Trudgill & Hannah 1985; Wells 1982). It is often difficult for an outsider to distinguish between New Zealand English (NZE) and Australian English (AE) speakers (Bauer 1994; Trudgill & Hannah 1985; Wells 1982) and at times even NZE speakers cannot readily hear the differences between New Zealand and Australian English, and vice versa (Bauer 1994; Bayard 1990; Wells 1982).

Both NZE and AE are varieties of non-rhotic English (Bauer 1994; Cochrane 1989; Hawkins 1976) and, although there is phonemic equivalence between the vowel systems of NZE, AE and Received Pronunciation (RP), there are also very clear phonetic differences (Hawkins 1976; Trudgill & Hannah 1985; Wells, 1982). Firstly, in comparison with RP, the NZE and AE vowels exemplified by HARD, WHO'D and HUD are more advanced. Secondly HAD and HEAD are raised relative to RP. Thirdly, there are differences in the position of the first target of the rising diphthongs HAY, HOE, HIDE, HOW relative to the monophthong space in NZE and AE compared with RP. Finally, the tense high vowels have a late target in both NZE and AE, but not in RP.

The 'mixing bowl' and 'single origin' theories have been used to explain the similarities between AE and NZE. Proponents of the mixing bowl theory believe that the new accent evolved from the mixture of accents and dialects of the new immigrants (predominantly of British descent) to both New Zealand and Australia (Bauer 1994; Gordon & Deverson 1985; Trudgill 1986). Those who propose the single origin theory believe that NZE and AE are an exported variety of a particular form of British English, with Cockney being an often cited source (Bauer 1994; Cochrane 1989; Gordon & Deverson 1985). Both Bauer (1994) and Gordon (1991) take the single origin theory one step further for NZE, arguing that it is a variety of AE.

Despite the well-acknowledged similarities between NZE and AE, there are also noticeable and distinctive differences. The most frequently noted difference between NZE and AE is the HID vowel: it has centralized in NZE relative to both RP and AE, but it has raised in AE relative to RP (Bauer 1979; Wells 1982). This centralized HID vowel is one of the most distinctive markers of NZE and it may be the result of an extended chain shift involving HAD and HEAD (Labov 1994: 138). However, whether this chain shift is a push-chain or drag-chain has not been established. Bauer (1979) implies it is a push-chain shift that has its source in the fronting of HUD, causing HAD and then HEAD to raise. The raised HEAD resulted in movement of HID, becoming raised in AE but centralized in NZE.

There is some suggestion that NZE HOOD is centralized relative to AE (Bauer 1986; Maclagan 1982). The HAD and HEAD vowels are more raised in NZE than their AE counterparts (Maclagan 1982; Trudgill & Hannah 1985; Wells 1982) and NZE HOARD may also be more raised (Maclagan 1982). Although both NZE and AE have onglides and delayed targets in HEED and WHO'D, the extent to which the target is delayed is believed to be greater in AE, particularly in WHO'D

[1] To avoid confusion that would arise from the different phonemic transcriptions of NZE and AE vowels, we will use the word set from the AE corpus listed in Tables 1 and 2 to refer to *equivalent* vowels in the two accents.

(Bauer 1986; Trudgill & Hannah 1985; Wells 1982). There is some impressionistic evidence that NZE HEAD is becoming diphthongal (Bauer 1986; Maclagan 1982; Turner 1970; Wells 1982), but this has not been noted in AE. However, from examples cited by Bauer (1986) and Turner (1970) (i.e. in the words YES and HERRING) it seems the diphthongization of the vowel could be at least partly caused by the coarticulation effects of the surrounding consonants. There is also some suggestion that NZE HERD has not only fronted but also become rounded relative to AE (Bauer 1979; Maclagan 1982; Wells 1982).

A further source of difference between AE and NZE is in the production of the falling diphthongs in HERE and HAIR. In NZE, these two diphthongs are merging. The direction of the merger is somewhat contentious (Holmes 1997; Maclagan & Gordon 1996). However, Maclagan and Gordon (1996) give a compelling argument that the merger is in the direction of HERE. This merger is not happening in AE.

There is little regional variation in either NZE or AE. There is, however, one exception, the Southland variety of New Zealand English (NZSE) (nomenclature borrowed from Bartlett 1992). Southland is the southern-most province in the South Island of New Zealand. There are a number of phonological and lexical differences between NZSE and NZE. The most notable phonetic feature of the NZSE is the postvocalic /r/. However, Bartlett (1992) has noted that the degree of rhoticism varies widely from speaker to speaker. It is in the 'words like "work" and "first" [that] the /r/ appears to be considerably more strongly maintained than it does elsewhere' (Bartlett 1992).

Although there have been several impressionistic analyses of the NZE vowel system (e.g. Hawkins 1973; Wells 1982; Bauer 1986), there has only been one published acoustic analysis (Maclagan 1982). This is in contrast to AE which has had numerous impressionistic and acoustic studies (e.g. Bernard 1970; Harrington & Cassidy 1994; Harrington *et al.* 1997). There are a few impressionistic studies comparing AE and NZE (e.g. Trudgill & Hannah 1985; Turner 1966 1970; Wells 1982), but the only acoustic comparison is once again the study by Maclagan (1982). In this comparison, Maclagan (1982) compared her data (collected in 1979) with Bernard's AE data, which were collected in the 1960s and were male only.

The studies comparing NZE and AE often draw on data collected in the 1960s (e.g. Turner 1970; Wells 1982; Maclagan 1982; Trudgill & Hannah, 1985). Thirty years on it is possible that there have been further divergences between AE and NZE that have not been accounted for. To investigate the similarities and differences in contemporary NZE and AE accents, we have compared vowels from the Otago Speech database of NZE, which was collected in 1995 (Sinclair & Watson 1995), with AE speech from the Australian National Database of Spoken Language (ANDOSL) (Millar *et al.* in press). The Australian speech was collected in the early 1990s.

2. Method

2.1. Talkers

2.1.1. New Zealand Speakers

The vowels that were analysed were taken from the Otago speech database (Sinclair

& Watson 1995) which includes speech from 11 men and 10 women, all between the ages of 16 and 33 years. All of the speakers were university educated, with the exception of two of the female speakers who were senior high school students. Most of the speakers were students of the University of Otago. The speakers came from all over New Zealand, including several from Southland.

2.1.2. Australian Speakers

The AE vowels that were analysed were taken from the isolated word materials collected for the Australian National Database of Spoken Language (ANDOSL) (Millar *et al.* 1994, in press; Vonwiller *et al.* 1995). In selecting the talkers for ANDOSL, an attempt was made to cover three age ranges (18–30 years, 31–45 years, 46 + years) and also to select AE talkers from the main accent types that form a continuum from *cultivated* to *broad.* In our study, the criterion we used to select talkers from ANDOSL was based on age only: we used talkers from the 18–30 years range to match them as closely as possible for age with the NZE talkers. There were 21 female speakers and 20 male speakers in the AE group. Their accent types were evenly distributed across the AE accent continuum.

2.2. Materials

2.2.1. New Zealand Database

The Otago speech database talkers read citation-form productions of 129 different words. For this study we selected 69 words, from which we extracted the 19 vowels of NZE in various phonetic contexts. Tables 1 and 2 list the vowels used in this study and the words from which the vowels were extracted. The choice of consonantal context was determined by several criteria. In all cases, the aim was to produce a list of predominantly monosyllabic meaningful words in which the context varied minimally from word to word and in which the context affected minimally the vowel. The consonantal context which conformed as far as possible to these criteria was as follows: for the initial consonant, voiced, bilabial, oral stop; for the final consonant either voiced or voiceless, alveolar, oral stop. We emphasize again that for the purposes of comparison with AE, and to avoid the problems of using two different transcription systems to refer to the same lexical set, we will use capitalized /h/-initial words to refer to the vowels. Therefore, when we discuss NZE HEED, we mean the /i/ vowels that occurred in the words 'eat', 'peat', 'bead', 'pea' of the Otago database (see Table 1).

Casual listening to the recordings showed that 11 talkers rhoticized some words with HARD, HOARD, HERD, HERE, and HAIR vowel nuclei. However, 6 of these 11 talkers only rhoticized the vowel in one word ('are'). The vowel that was rhoticized the most often in varying consonantal contexts by the largest number of talkers was the HERD vowel (as found by Bartlett 1992); only four talkers consist ently rhoticized this vowel. We ran various tests to determine whether there were any significant differences in the vowel spaces in the F1/F2 plane between the remaining vowels of the talkers who rhoticized vowels and those of the other non-rhotic speakers. Since there were none, we decided to retain the speakers in the study after removing all rhotic vowels. The total number of vowel tokens per word that were used in this study are given in Tables 1 and 2.

2.2.2. ANDOSL Database

As described in Millar *et al.* (in press) and Vonwiller *et al.* (1995), the ANDOSL talkers read citation-form productions of 25 different words. For the present paper, we selected the words from an /hVd/ or /hV/ context. In addition, we also selected 'hoist' and 'tour' words to include the HOIST and TOUR diphthong nuclei, respectively, which did not occur in the /hV/ or /hVd/ context. The distribution of AE vowels that were used in this study is given in Tables 1 and 2.

2.3. Recording and Digitizing

2.3.1. New Zealand Database

All subjects were recorded in a quiet room. Each word was read three times. The signal-to-noise ratio for the word collection averaged 44 dB. The speech was sampled at 22.05 kHz and quantized to a 16 bit number (see Sinclair & Watson 1995 for further details). The material for each speaker was recorded over several days owing to its large content, and was presented to the speaker one word at a time from cards to avoid list intonation. The order of the words was random.

2.3.2. ANDOSL

The ANDOSL data were recorded in an anechoic chamber at the National Acoustic Laboratory, Sydney. It was digitized at 20 kHz and quantized to a 16 bit number (for further details see Harrington *et al.* 1997 and Millar *et al.* in press). The material was recorded in a single session, and for the isolated words lists, was presented to the subjects one word at a time to avoid list intonation (see Millar *et al.* in press for further details).

2.4. Labelling

2.4.1. New Zealand Database

The first three formants and their bandwidths for both the NZE and ANDOSL data were automatically tracked using ESPS/Waves (12th order LPC analysis, cosine window, 49 ms frame size, and 5 ms frame shift). All automatically tracked formants were checked and hand-corrections were made when these were considered necessary. Formant tracking errors were especially common in vowels which had F1 and F2 close together (e.g. HOARD) and they were far more common in the female data than the male data. All the labelling was done in EMU, a hierarchical speech data management system (Cassidy & Harrington 1996).

The acoustic onset of the vowel was marked at the onset of voicing as shown in the spectrogram by vertical striations. Additional cues were the onset of periodicity in the waveform, or, if the preceding consonant was a stop, where the aperiodicity caused by the aspiration had ceased.

Table 1. The NZE (columns 1–3) and AE (columns 4–6) monophthongs showing the number of tokens produced by female (columns 2 and 5) and male (columns 3 and 6) talkers, respectively

	New Zealand English			**Australian English**	
	Number of tokens			Number of tokens	
Word	Female	Male	Word	Female	Male
eat	30	33			
peat	30	33	heed	21	20
bead	30	33			
pea	30	33			
ooze	30	33			
boot	30	33	who'd	21	20
rude	30	33			
boo	30	33			
auto	30	32			
port	30	32	hoard	21	20
pawed	30	33			
door	27	30			
are	15	14			
part	30	33	hard	20	20
barb	27	29			
car	27	30			
irk	27	28			
pert	26	24	herd	21	20
bird	30	26			
fur	30	27			
if	29	33			
pit	30	33	hid	21	20
did	30	33			
foot	30	33			
good	27	30	hood	21	20
book	30	33			
push	30	33			
otter	30	33			
pot	30	33	hod	21	20
pod	30	33			
utter	30	33			
but	30	33	hud	21	20
pub	28	33			
egg	30	33			
pet	30	33	head	21	20
beg	30	33			
at	30	33			
sat	30	33	had	21	20
dad	30	33			

Table 2. The NZE (columns 1–3) and AE (columns 4–6) diphtongs showing the number of tokens produced by female (columns 2 and 5) and male (columns 3 and 6) talkers, respectively

	New Zealand English			**Australian English**	
	Number of tokens			Number of tokens	
Word	Female	Male	Word	Female	Male
age	30	33			
date	29	33	hay	21	20
fade	29	33			
day	30	33			
over	30	33			
boat	29	33	hoe	21	20
code	30	33			
go	30	33			
oyster	30	33			
voice	30	32	hoist	21	20
void	30	33			
joy	30	30			
ice	30	33			
pipe	30	33	hide	21	20
hide	30	33			
die	29	32			
out	30	32			
about	30	33	how	20	21
found	30	33			
how	30	33			
ear	23	22	here	21	19
hear	27	25			
air	28	30	hair	21	20
there	26	28			
tour	21	28	tour	18	16

The acoustic vowel target was marked as a single time point between the onset and offset. We marked a single target in monophthongs and two targets in diphthongs. The targets of the monophthongs and first targets of the diphthongs were marked at a point where there was the least movement in the formant tracks. For the high vowels, this point occurred where F2 reached a peak (for the WHO'D vowel this was where both F2 and F3 were close); for open vowels the target was marked where F1 was a maximum; for back vowels the target was marked where F2 reached a trough. If none of the above criteria were satisfied for a given vowel, the target was marked at the point of maximum amplitude in the waveform. The second target of the rising and falling diphthongs was marked according to the same criteria as the first target.

2.4.2. ANDOSL

The method used in obtaining the formants was the same as for the NZE data. The labelling was done by the third author and was very similar to the process for labelling the NZE data. The acoustic onset of the vowel was marked at the onset of voicing as shown by strong vertical striations in the spectrogram, and by the onset of periodicity in the waveform. The acoustic offset of the vowel in the /hVd/ context was marked at the closure of the [d] corresponding to a cessation of regular periodicity for the vowel and/or a substantial decrease in the amplitude of the waveform. The acoustic vowel target was marked as a single time point between the acoustic onset and offset according to the criteria discussed above. In the diphthongs, two targets were marked using the same sets of criteria as for the monophthongs. Full details of the annotation are given in Harrington *et al.* (1997).

3. Results

In this section we will present a comparative acoustic analysis of NZE and AE. We focus on the monophthongs in the first section; in Section 3.2, we consider whether NZE HEED and WHO'D are as diphthongal as in AE; in the final section we analyse the rising and falling diphthongs.

We will consider the female and male data separately. No speaker normalization was performed on the data because, although normalization can reduce speaker effects, it can also distort the data in unpredictable ways (Disner 1980).

For the most part, the results from the female and male data were very similar. All comments can be assumed to apply to both sets of data unless otherwise stated.

3.1. Monophthongs

Figure 1 shows the ellipse plots in the F1/F2 plane for the monophthongs of female and male NZE speakers, respectively. The centroid of each vowel class is marked by the appropriate item from the lexical set, and each ellipse includes at least 95% of the data for a particular class. Figure 2 shows the centroids of NZE and AE monophthongs for the female and male data, respectively.

The most striking feature of Figure 1 is the centralized NZE HID. This is quite in contrast with the AE HID, which is a high front vowel (see Figure 2). The NZE HID is also lower than the NZE HEAD, suggesting it has both centralized and lowered. Figures 1 and 2 show that the lax front vowels HEAD, HAD are raised in NZE relative to AE, with the NZE HEAD having similar F1 and F2 values to AE HID. The figures also suggest that the NZE WHO'D is possibly more backed, relative to AE, and that NZE HOOD is more central and lower than AE; only a few studies have commented on this feature (e.g. Maclagan 1982; Bauer 1986). The NZE and AE HARD, HUD, and HOD vowels are positioned in very similar places in the F1/F2 plane.

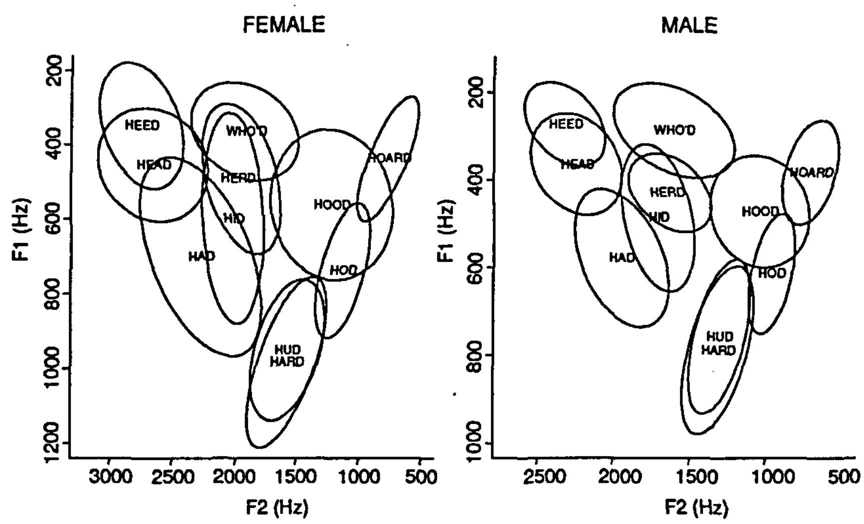

Figure 1. Ellipse plots of the monophthong in the NZE database in F1/F2 space for the female data (*left*) and the male data (*right*). The centroid of each vowel class is marked by the appropriate lexical item, and the perimeter of each ellipse indicates the boundary of the area in which at least 95% of the data for a particular class fall.

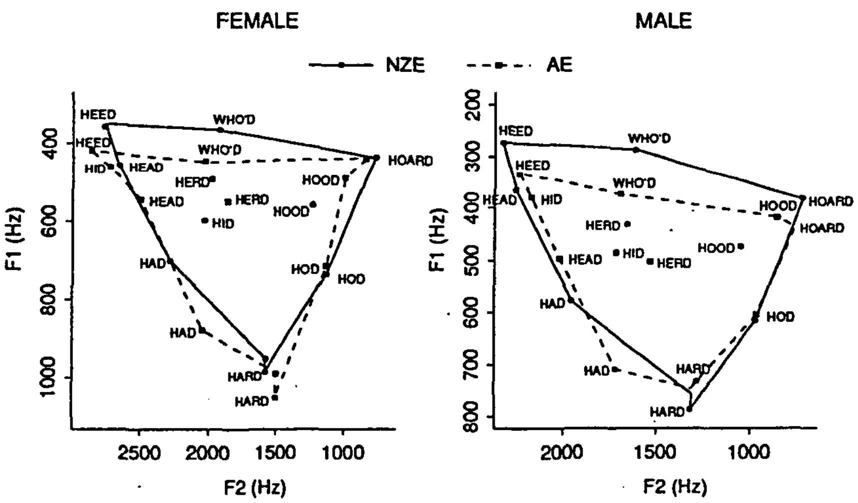

Figure 2. The centroid of the monophthong vowel targets from the NZE and AE for the female data (*left*) and the male data (*right*).

In general, for all the vowels, there is very little difference between the female and male data for either NZE or AE. The exception to this is that the male NZE HOARD is raised relative to the male AE HOARD, whereas there is no difference between the female NZE and AE HOARD.

In summary, the major differences in the NZE and AE vowel spaces are the centralizing and lowering of HID and HOOD and the raising of the front and high vowels. Both Wells (1982) and Trudgill and Hannah (1985) have commented that there is a tendency for HAD and HEAD in NZE to be more raised than in AE. We found that the HAD and HEAD vowels of NZE were definitely more raised than in AE. Furthermore we found all the NZE front vowels (HAD, HEAD, HEED), the high central vowels (WHO'D, HERD) and perhaps HOARD (though this is only for the male data) to be more raised than those in AE. The extent of raising in front and high central vowels of NZE has not been noted in earlier studies.

The question of whether HERD is as raised and retracted as the formant plane plot suggests needs to be investigated further. If HERD is produced with lip rounding (Wells 1982), then this vowel may in fact be slightly lower and more front than its position in the formant plane would suggest. This is because lip-rounding, particularly in the front vowel set, lowers all formant frequencies (Fant 1960; Lindblom & Sundberg 1971). Moreover, lip-rounding causes a progressively greater lowering of the frequency of higher formants, so that lip-rounding (or more specifically the associated vocal tract lengthening) has a greater absolute lowering effect on F2 than on F1. Therefore, if we were able to factor out the contribution of lip-rounding to the F1 and F2 values of HERD, its position in the formant plane plot (due now primarily to the tongue constriction location and size) would be slightly lower and a good deal more fronted, probably intermediate between NZE HAD and NZE HEAD, but closer to HEAD. In other words, it is possible that the tongue positions of NZE HERD and HEAD are quite similar and that the acoustic differentiation between these vowels comes about because HERD has a lip-rounding component which results in an upwards and more central shift in the formant plane plots relative to the front lax vowels.

3.2. Onglides of /i u/

Figure 3 shows averaged formant trajectories for HEED and WHO'D in NZE and AE. In order to match the contexts as closely as possible, the NZE words were taken from 'eat' and 'ooze' which are both closed by an alveolar consonant (to match AE 'heed' and 'who'd'). Each of the formant trajectories was linearly interpolated and time normalized in order to linearly stretch or compress the trajectories such that they were of equal duration. Also given in Figure 3 are the average positions of the vowel targets relative to the acoustic vowel boundaries for both NZE and AE. The criteria for marking the targets were outlined in Section 2.4.

Figure 3 shows that NZE HEED and WHO'D both have delayed targets, i.e. long onglides from a vowel with a schwa-like quality. The figure suggests that NZE WHO'D and the female NZE HEED are more diphthongal than in AE because the NZE data have lower F2 onsets and the targets are more delayed. To investigate whether these differences in NZE and AE were significant, we compared the change in frequency in F2 between the target ($F2_{tar}$) and onset ($F2_{on}$) (see Table 3), and in Table 4 we compared the average target times expressed as a proportion of the vowel duration (i.e. a mean value of 0.5 would indicate that the target occurs, on average, at the midpoint). The tables show there is no significant difference between NZE

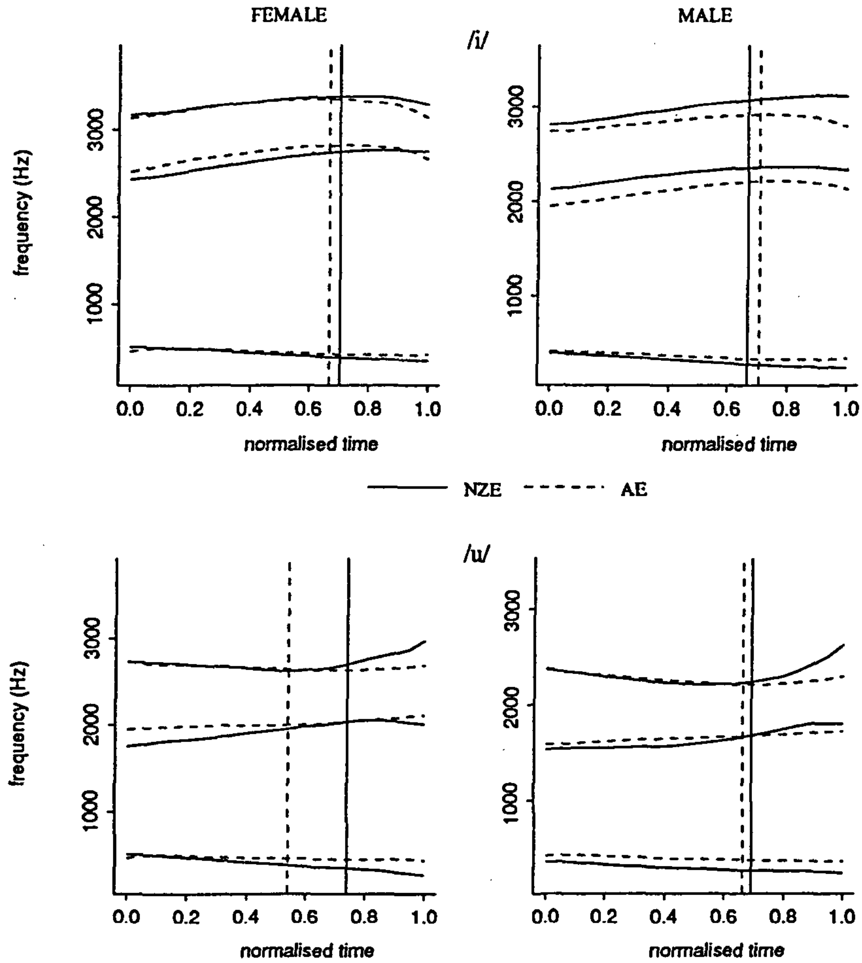

Figure 3. Averaged time normalized formant trajectories from vowel onset to offset for the NZE vowels from the words 'eat' and 'ooze', respectively, contrasted with those for the AE vowels from the words 'heed' and 'who'd', respectively, for the female data (*left*) and the male data (*right*). Superimposed on these trajectories are the average positions of the target relative to the acoustic vowel boundaries.

and AE in the degree of diphthongization of the female and male HEED and the male WHO'D, as judged by these parameters; however, the female NZE WHO'D is more diphthongal than in AE both because the difference $F2_{tar} - F2_{on}$ is significantly greater in NZE and because the target is significantly more delayed.

These findings on onglides contradict the comments made in the literature about the diphthongization of the NZE HEED and WHO'D. Unlike Maclagan (1982), we found that NZE HEED and WHO'D are both diphthongal. Wells (1982) has stated that the degree of diphthongization for the NZE HEED is less than that for AE; we found there was no difference in the degree of diphthongization. Both Wells (1982) and Trudgill and Hannah (1985) have commented that the degree of diphthongization is less for the NZE WHO'D than for AE; however, we found there is no difference in the degree of diphthongization for the male WHO'D and in fact the

Table 3. The mean frequency difference between F2 at the target and at the onset ($F2_{tar} - F2_{on}$) in /i u/ for the NZE and AE data

	NZE $F2_{tar} - F2_{on}$	AE $F2_{tar} - F2_{on}$	Sex	*t*	
i	347	326	female	0.46	
u	272	89	female	4.85	**
i	244	284	male	−1.01	
u	178	104	male	1.70	

**Indicates a significant difference at $p < 0.01$.

NZE female WHO'D is more diphthongal than in AE. Finally, Bauer (1986) states that the degree of diphthongization for the NZE WHO'D is less than the NZE HEED, whereas our analysis suggests that it is as diphthongal as HEED.

The discrepancy between the results from the analysis of our corpora and these impressionistic studies concerning the extent of diphthongization in high vowels warrants further investigation. A possible explanation for these differences is that our Australian corpus includes cultivated talkers who, as shown in Harrington *et al.* (1997) and other studies, tend to have the targets closest to the vowel midpoint and a minimal onglide; however, there is no control for the cultivated-general-broad distinction (to the extent that this is possible) in our NZE corpus. We therefore reanalysed the same data as in Tables 3 and 4, but after removing all the cultivated talkers from the AE corpus. The results showed the same pattern of significant differences as in Tables 3 and 4. This suggests that our finding, that there are no differences in the extent of AE and NZE diphthongization, is unlikely to have been influenced by the lack of F2 onglide or delayed target in the AE cultivated group.

3.3. Rising Diphthongs

Figure 4 shows the NZE averaged trajectories for the three front-rising diphthongs and the two back-rising diphthongs superimposed on the ellipse plots of a selection of NZE monophthongs from Figure 1. The diphthong trajectories were obtained by sectioning the formant trajectories between the first and second target times. Then

Table 4. The average target times expressed as a proportion of the total vowel duration in /i u/ for NZE and AE data

	NZE target time	AE target time	Sex	*t*	
i	0.73	0.69	female	1.35	
u	0.67	0.54	female	2.381	*
i	0.7	0.71	male	−0.09	
u	0.68	0.66	male	0.27	

*Indicates a significant difference at $p < 0.05$.

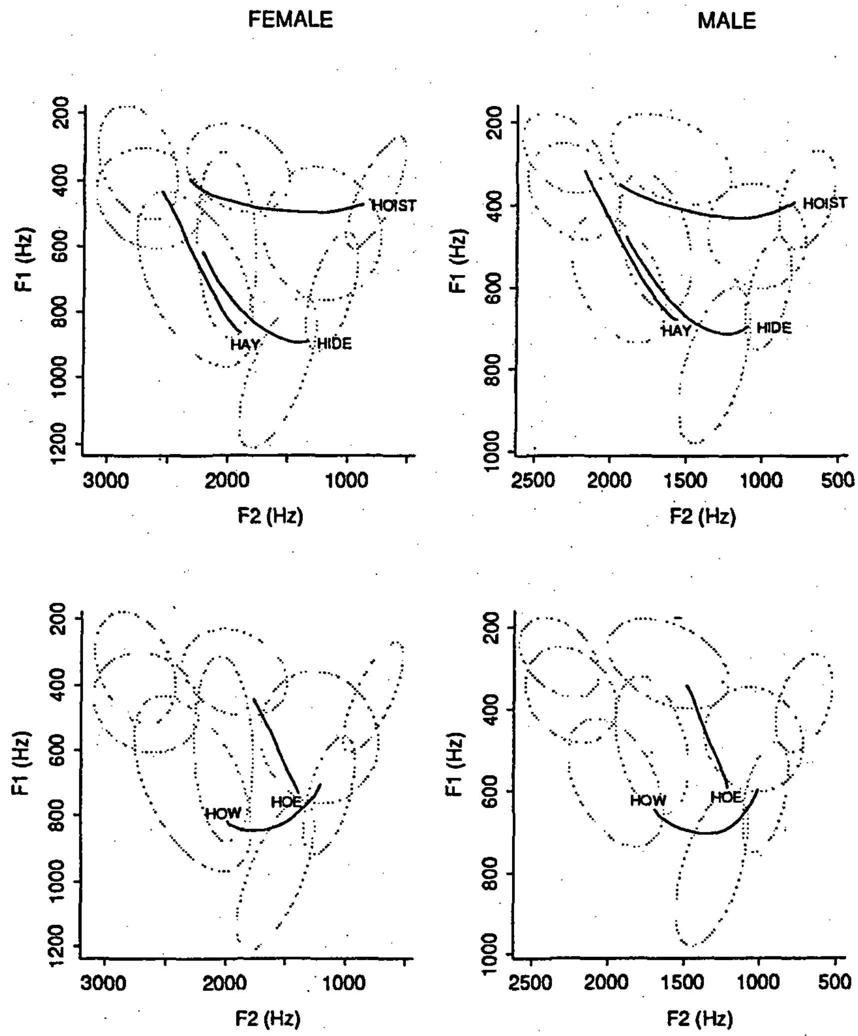

Figure 4. The formant trajectories of the NZE rising diphthongs which have been sectioned between the two targets, linearly interpolated, time normalized, and then averaged for the female data (*left*) and the male data (*right*). The trajectories have been superimposed on a selection of ellipse plots of the NZE monophthongs.

the trajectories were linearly interpolated and time normalized, and finally all the trajectories for each phonetic class were averaged. This method is the same as that used by Harrington *et al.* (1997).

Figure 4 shows some of the well-documented characteristics of the NZE rising diphthongs when contrasted with RP, such as: a low first target of HAY, a fronted first target for HOW, and a low and retracted first target of HIDE. The first targets of HAY, HIDE, and HOIST are close to the NZE HAD, HARD, and HOARD spaces, respectively. The second targets of the front-rising diphthongs seem to be associated with HEAD (slightly lower than the HEAD space for HIDE, in the HEAD space for HAY and at the edge of the HEAD space for HOIST). The glides

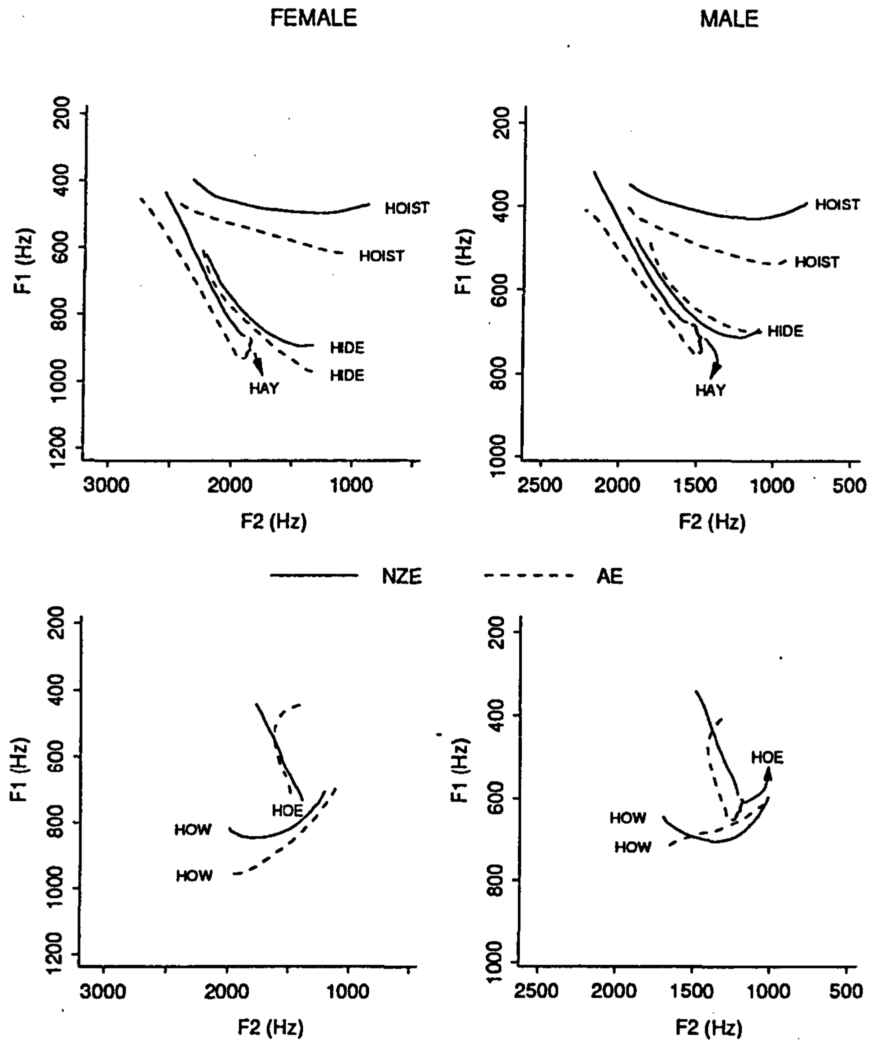

Figure 5. The formant trajectories of the NZE and AE rising diphthongs which have been sectioned between the two targets, linearly interpolated, time normalized, and then averaged for the female data (*left*) and the male data (*right*).

for the two back-rising diphthongs are not as extensive as those of the front-rising diphthongs. The first target for HOW is at the right edge of the HAD space and the second target points towards HOD. For HOE the first target is at the upper edge of the HARD space and the second points towards WHO'D.

Figure 5 contrasts the NZE rising diphthong trajectories with their AE counterparts. The diphthong trajectories for the AE data were calculated in the same way as the NZE ones. It can be seen that the trajectories of the rising diphthongs of NZE and AE are relatively similar: the main difference is that most of the averaged trajectories of the NZE front-rising diphthongs are higher (lower F1) relative to those of AE. The first targets of HAY and HOIST in the NZE data may have raised with NZE HAD and HOARD (see Figure 2). However, since, as Figure 4 shows, the first target of NZE HAY is a good deal lower than that of NZE HAD, the extent of target raising in HAY is much less than in HAD.

The first targets of the back-rising HOW and HOE have a lower F1 in NZE compared with AE which may indicate that they are phonetically more raised. However, the biggest difference between the NZE and AE back-rising diphthongs is in the direction of the second target of HOE. In NZE, the second target points towards the WHO'D space, whereas in AE the direction of the movement is towards HOOD. However, as Harrington *et al.* (1997) have noted, the position of the second target in HOE in AE is contentious, since Bernard (1970) and Cox (1996) both found it pointed towards the AE WHO'D space.

We have demonstrated that the NZE rising diphthongs are very similar to their AE counterparts, as found by Maclagan (1982). This suggests that the second target of the front-rising diphthongs was not affected by the centring of HID, since the second targets point towards the NZE HEAD space (see Figure 4). It does appear, however, that the raised NZE HAD and HOARD may have influenced both the first targets, and indeed the entire diphthong trajectories, of HIDE and HOIST.

3.4. *Falling Diphthongs*

Figure 6 shows the NZE averaged trajectories for the three falling diphthongs superimposed on the ellipse plots of a selection of NZE monophthongs from Figure 4. The diphthong trajectories have been calculated in the same manner as those for the rising diphthongs in Figure 4. Both HERE and HAIR have similar trajectories: the main difference is the raised first target for HERE vowels. The second targets of both HERE and HAIR point towards the lower region of the HID space. The first target of TOUR starts in the middle of the HOOD. The second target points towards the lower HID space.

The small difference in the HERE and HAIR trajectories in Figure 6 is evidence of the well documented HERE/HAIR merger (Bauer 1994; Holmes 1997; Maclagan & Gordon 1996; Trudgill & Hannah 1985; Wells 1982). Further evidence for the merger can be seen in Figure 7 which shows averaged time normalized formant trajectories for AE and NZE HERE and HAIR (these were calculated in the same way as for the rising diphthongs). The NZE HERE and HAIR formant trajectories are closer together than in AE. Furthermore, differences between the AE HERE and HAIR trajectories can be noted across the entire trajectory, particularly for F2. However, the most noticeable differences in the NZE HAIR and HERE trajectories are between the onset and first target; the differences in the trajectories at the offset are negligible.

Figures 6 and 7 suggest that the extent of the HERE/HAIR merger is greater in the female NZE data than in the male NZE data. This difference in the degree of the merger between the NZE female and male data was also noted by Maclagan and Gordon (1996). They found that for young female and male non-professional subjects the merger was almost complete. The young female professionals still retained some distinction about 20% of the time. However, the young male pro-

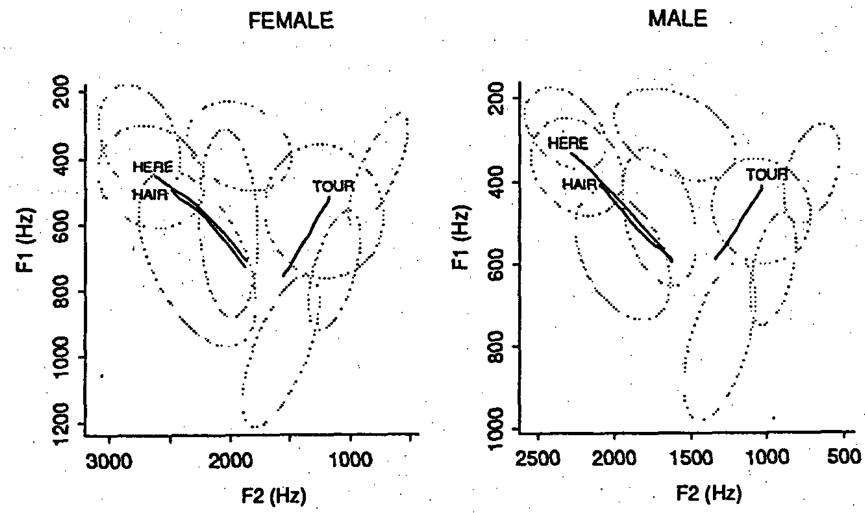

Figure 6. The formant trajectories of the NZE falling diphthongs which have been sectioned between the two targets, linearly interpolated, time normalized, and then averaged for the female data (*left*) and the male data (*right*). The trajectories have been superimposed on a selection of ellipse plots of the NZE monophthongs.

fessional subjects were least likely in this group to have merged HERE/HAIR; the distinction was maintained in approximately 40% of the talkers.

In summary, the most noticeable difference between AE and NZE diphthongs was in the falling HERE and HAIR diphthongs, with a merger occurring in NZE, but not in AE. The first target of the NZE HERE is in a very similar position to the first target of the AE HERE (see Figure 7 and Harrington *et al.* 1997); however, the first target of the NZE HAIR is a good deal more raised than in AE (see Figure 7 and Harrington *et al.* 1997). This supports the view of Maclagan and Gordon (1996) that in NZE the direction of the merger is towards HERE and that the first target of HAIR has been affected by the raising of the HEAD vowel.

4. Discussion

The results of this acoustic phonetic comparison between AE and NZE vowels are broadly consistent with earlier impressionistic studies. We have shown that NZE HID has a central quality relative to the NZE vowel space and that it is certainly a good deal more central than its AE equivalent. The results also confirm that HERE/HAIR have almost merged in NZE and that most NZE rising diphthongs have shifted in a way that could be predicted from the movement of the NZE monophthongal vowel space relative to that of AE. Contrary to some suggestions in the literature, our results show that HEED and WHO'D vowels have delayed targets in NZE and that the extent of the onglide (and therefore the temporal delay in the target relative to the vowel midpoint) is not substantially different in AE and NZE. Another finding to emerge from our analysis is that the NZE HERE/HAIR merger

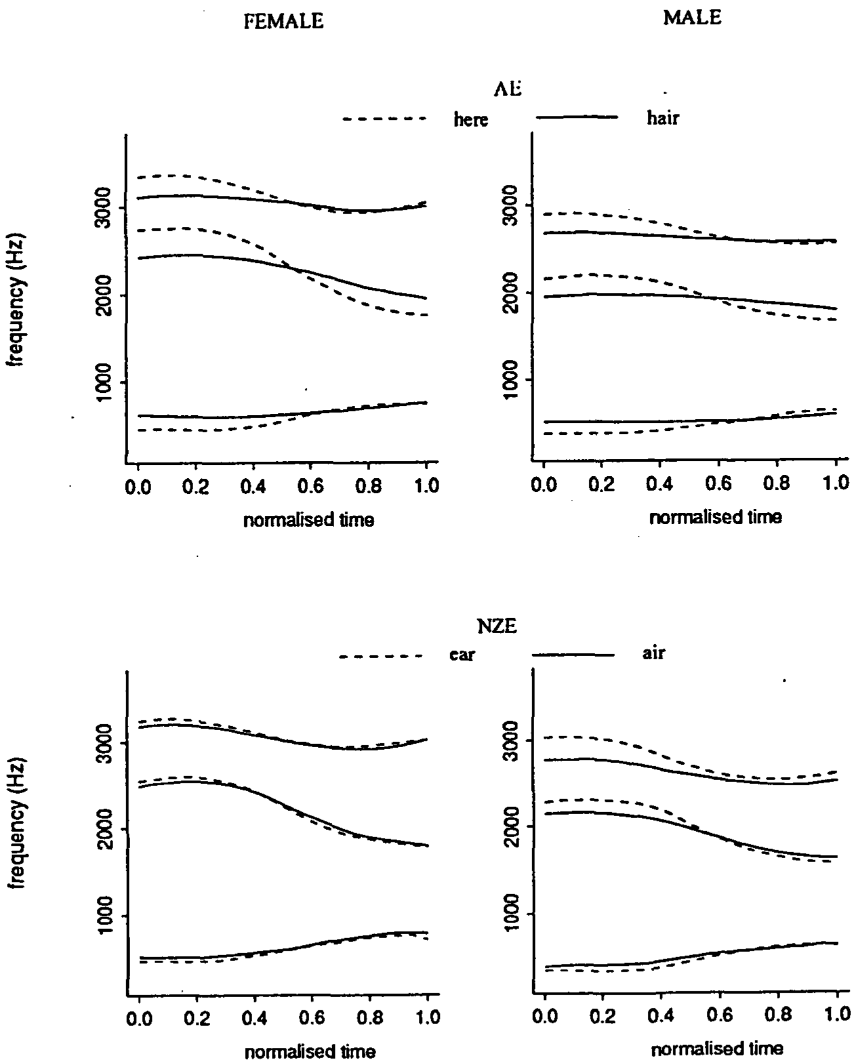

Figure 7. Averaged time normalized formant trajectories for the NZE vowels from the words 'ear' and 'air', respectively, contrasted with the averaged time normalized formant trajectories for the AE vowels from the words 'hear' and 'hair', respectively, for the female data (*left*) and the male data (*right*).

is in progress. Although their formant trajectories are similar (and a good deal more so in NZE than in AE), they show that NZE HAIR has a lower first target than HERE. Moreover, the fact that female talkers have a more complete merger than the male talkers (see also Maclagan & Gordon 1996) is in accordance with the general principle that female talkers tend to be innovative in sound change (but conservative once the change has become established: see Labov 1990; Holmes 1997). Our analysis is also consistent with the evidence in Maclagan and Gordon (1996) which shows that the direction of the merger is a shift from HAIR towards HERE and not the other way round.

The acoustic phonetic comparison between these two accents shows clearly that AE and NZE HUD have quite a similar quality (and the same applies to AE and NZE HARD, which can be considered to be a lengthened version of HUD in both accents) and this may have some implications for the push-chain theory (Bauer 1979) in which it is proposed that the fronting of these open vowels (relative to RP) may have caused the raising of the lax vowels resulting ultimately in the lowering and centralization of NZE HID. The unresolved issue here is why NZE HAD and HEAD should be considerably higher than their AE equivalents if the shift in the lax vowels can be attributed to the movement of HUD, which is largely undifferentiated in these two accents. That is, there would be much clearer support for such a push-chain theory if there were evidence to show that NZE HUD and HARD are more fronted than in AE. Our data are therefore at least as compatible with the alternative theory that the shift in the NZE lax vowels is the result of a *drag*-chain which was initiated by the lowering and centralization of NZE HID, causing a subsequent raising of the NZE HEAD and HAD vowels.

A further issue which needs to be resolved, and which is not unrelated to the drag vs push-chain theories outlined above, is whether present-day NZE HAD and HEAD can still be categorized as lax vowels. Although our data and earlier impressionistic studies show quite clearly that these vowels are raised relative to AE, there is no evidence that the raising has been accompanied by any shortening. Consider in the light of this the well-known finding from acoustic phonetics (e.g. Lehiste & Peterson 1961) that in English and many languages there is an inverse relationship between vowel height and length which can be explained by the greater distance through which the articulators have to move in low vowels (e.g. the magnitude of jaw displacement is a good deal more extensive in a low vowel than in a high vowel). If NZE HAD and HEAD have raised but not shortened, then they would have a greater length than would be expected from their height and this may well cause them to be perceived as tense (since length also cues vowel tensity in English). Therefore, the vowel quality changes in the (so-called) NZE front lax vowel set could have occurred not only in height and backness, but also in the tense/lax dimensions. For example, it is possible that lax NZE HID centralized and lowered (and has stayed lax because it has not lengthened) and that NZE HEAD and HAD became more tense as they raised. Moreover, these vowel quality changes involving the tense/lax dimension would now be in accordance with Labov's (1994) general principles that (a) lax vowels tend to fall non-peripherally (this applies to NZE HID) and (b) tense vowels rise on a peripheral track (under this modified theory, this would apply to NZE HAD and NZE HEAD). It would therefore be of interest to assess both whether NZE HAD and HEAD have maintained approximately the same duration over the last 30–40 years and also to compare vowel durations with AE vowels in an equivalent context (this is not possible in the present study because of the differing consonantal contexts of our NZE and AE corpora).

We have argued so far that the NZE vowel shift may involve changes not only to vowel height and backness, but also to vowel tensity; a further possibility is that the maintenance of their perceptual contrast involves some changes in the unrounded/ rounded dimension as well. Although the formant plots show differences between

NZE HERD and HEAD, we have argued that this need not imply that their tongue configurations are markedly different. This is because if, as has been suggested, NZE HERD is produced with a lip-rounding component, its formants would all be lowered thereby maintaining its contrast with NZE HEAD. In fact, if we assume firstly that NZE HEAD has lengthened perceptually (as discussed above) and secondly that, as it raised it overlapped progressively more with HERD, which was also becoming more fronted, then the development of lip-rounding in HERD may have been necessary precisely to maintain its contrast with HEAD. We are currently using kinematic procedures to compare both NZE HERD and HEAD on tongue positions and to assess whether they differ in the magnitude of lip-rounding.

We offer finally some comments on a proposed transcription system for NZE which takes into account our acoustic phonetic analysis and proposals that have been made by Bauer (1986) and Wells (1982). In proposing our NZE transcription system, we wish to highlight both the relationship to AE but also to reflect the differences. Another principle is to take into account the evidence suggested by the experimental data wherever possible.

Our proposed transcription system for NZE is given in Table 5; also listed is the transcription system for AE proposed by Harrington *et al.* (1997) and the transcription systems for NZE proposed by Bauer (1986) and Wells (1982).

Our transcriptions for the long monophthongs in HEED, WHO'D, HOARD, HARD are the same as the ones we have proposed for AE in Harrington *et al.* (1997). We are therefore consistent with Wells (1982) in our inclusion of length marks but, as for AE, we have [ɐː] for HARD to reflect a quality approximately intermediate between cardinal vowels 4 and 5. The transcription of HERD is problematic because we do not yet have enough experimental evidence about its tongue configuration and lip-rounding status. Assuming that it is rounded and that it has a tongue height not dissimilar to NZE HEAD, then [øː] or possibly even [ʏː] would be appropriate. Alternatively, for an unrounded transcription we would propose [ɘː] which is higher than [ɜː] (thus also reflecting the evidence showing that NZE HERD is higher than in AE).

Our transcription of HID, HOOD, HUD, HAD agrees with Bauer (1986). The transcription of HID with [ə] is appropriate to reflect both its central quality and lowering relative to AE; [ɛ] for HAD is justified for the same reasons. The motivation for [ɐ] is firstly that NZE HUD is similar in quality to its AE equivalent, and secondly because HUD differs from NZE (or AE) HARD only in length. In our *Australian* English transcription system, HOD is transcribed with [ɔ] to reflect that this vowel is somewhat raised relative to RP; since our experimental data suggest very few differences between the NZE and AE qualities for this vowel, we have proposed the same symbol for NZE HOD. For NZE HEAD, we would tentatively propose [ɪ] to show that the vowel is raised relative to AE (for which we have [e]). This would imply that AE HID and NZE HEAD have a similar quality: while this is suggested by our acoustic data, we would certainly recognize the need to collect further evidence before being definitive about the relationship between these vowels.

Our proposed transcription system for the diphthongs diverges more extensively from Bauer (1986) and Wells (1982). Firstly, for HAY, the acoustic phonetic evidence from these corpora shows that the first target may be a more open production of NZE HAD, possibly intermediate in quality between NZE HAD and NZE HUD. Since there is some ambiguity as to whether HAY is more closely affiliated with HAD or HUD, we calculated, separately for the NZE male and NZE female data, whether it is probabilistically closer to HAD or HUD. This was done by training a (Bayesian) classification model on a two-dimensional F1 × F2 space using only these two monophthong classes, and then classifying the first target of HAY as either HAD or HUD. The results showed that 58% of the first target HAY tokens were classified as HAD (the remainder as HUD) in the female data; for the male talkers, 61% were classified as HAD. These results simply confirm that the first target of HAY is intermediate between HAD and HUD. Since we are proposing [ɛ] for NZE HAD and [ɐ] for HUD, we would propose [æ] for the first target of HAY to reflect an intermediate quality. Moreover, since the second component of HAY extends into the HEAD space, Figure 4, and since we are proposing a transcription of [ɪ] for HEAD, our transcription of NZE HAY is [æɪ].

The first target of NZE HOW is somewhat closer than that of NZE HAY to the centroid of NZE HAD. Moreover, since a probabilistic classification similar in design to the one carried out for HAY (discussed in the preceding paragraph) shows that 78% (female) and 87% (male), respectively, of the first targets of HOW were closer to NZE HAD than to NZE HUD, we arrive at [ɛɔ] for NZE HOW (but [æɔ] for AE). The transcription of HIDE is more problematic. For AE, Harrington *et al.* (1997) proposed [ɑe] to reflect firstly the backer quality of the first component relative to AE HARD and to show the association of the second component to AE HEAD. Since the evidence suggests that the AE and NZE diphthong trajectories traverse a similar space in the formant plane (see Figures 4 and 5), we could propose the same transcription of [ɑe] for NZE HIDE. Notice that, since we have [ɪ] for NZE HEAD, such a transcription implies that the second component of NZE HIDE is intermediate in height between NZE HAD and NZE HEAD (for which we have [ɛ]). This is partly substantiated by the evidence in the top panels of Figure 4 which shows that the diphthong trajectory of NZE HIDE terminates above the centroid of HAD; alternatively, if subsequent experimentation shows that the second component of NZE HIDE is closely affiliated with NZE HAD, [ɑɛ] could also be appropriate. Our suggestions for the transcription of the HOE and HOIST rising diphthongs are the same as for AE; i.e. [əʉ] and [oɪ]; for HOIST, this again highlights the relationship between its second component and NZE HEAD (but AE HID).

Finally, we have proposed [iə] and [ɪə] for those talkers who make a pronunciation distinction between NZE HERE and HAIR. This transcription is justified for three reasons and further validates the choice of [ɪ] for NZE HEAD. Firstly, [i] correctly reflects the association of the first component of HERE to NZE HEED. Secondly, the choice of [i] vs [ɪ] implies that the qualities of the first components of these two falling diphthongs are very similar (again appropriate for NZE) and certainly more so than in AE for which we have [ɪ] vs [e] (note that although Harrington *et al.* 1997 did not explicitly propose transcriptions for HERE and HAIR, these were given the phonemic symbols /iə eə/, respectively, in the text).

Table 5. Proposed transcription system of New Zealand English (*left*), the transcription system of the Australian data used in this study as proposed by Harrington *et al.* (1997) and the transcription systems of NZE proposed by Bauer (1986) and Wells (1982), respectively

		Vowel system		
Lexical set	Proposed NZE	AE Harrington *et al* (1997)	NZE Bauer (1986)	NZE Wells (1982)
HEED	iː	iː	əi	iː
WHO'D	ʉː	ʉː	ʉ	uː
HERD	øː *or* ɘː	ɜː	øː	ɜː
HOARD	oː	oː	oɐ	ɔː
HARD	ɐː	ɐː	a	aː
HID	ə	ɪ	ə	ə
HOOD	ʊ	ʊ	ʊ	ʊ
HEAD	ɪ	e	e	e
HOD	ɔ	ɔ	ɒ	ɒ
HUD	ɐ	ɐ	ɐ	ʌ
HAD	ɛ	æ	ɛ	æ
HAY	æɪ	æe	ɐe	ʌɪ
HOE	əʉ	əʉ *or* əʊ	ɒʊ	ʌʊ
HOIST	oɪ	oɪ	oe	ɔɪ
HIDE	ɑe	ɑe	ɒe	ɑɪ
HOW	ɛɔ	æɔ	ɐʉ	æʊ
HERE	iə	ɪə	iɐ	iə
HAIR	ɪə	eə	eɐ	eə

Thirdly, the proposed transcription appropriately conveys the impression that NZE HAIR is more similar to AE HERE than to AE HAIR. The choice of [ə] is also not inappropriate because, as Figure 6 shows, these diphthong trajectories terminate well within the NZE HID space.

References

Bartlett C 1992 'Regional variation in New Zealand English: the case of Southland' *New Zealand English Newsletter* 6: 5–15.

Bauer L 1979 'The second great vowel shift' *Journal of the International Phonetic Association* 9(2): 57–66.

Bauer L 1986 'Notes on New Zealand English phonetics and phonology' *English World-wide* 7(2): 225–258.

Bauer L 1994 'English in New Zealand' in R Burchfield (ed.) *The Cambridge History of the English Language* (vol. 5) Cambridge University Press Cambridge: 382–429.

Bayard D 1990 '"God help us if we all sound like this": Attitudes to New Zealand and other English accents' in A Bell & J Holmes (eds) *New Zealand Ways of Speaking English* Multilingual Matters Clevedon: 67–96.

Bernard JRL 1970 'Towards the acoustic specification of Australian English' *Zeitschrift für Phonetik* 2/3: 113–128.

Cassidy S & J Harrington 1996 'Emu: An enhanced hierarchical speech data management system' in P McCormick & A Russell (eds) *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*: 361–366.

Cochrane G 1989 'Origins and development of the Australian accent' in P Collins & D Blair (eds) *Australian English: the Language of a New Society* University of Queensland Press St Lucia: 176–185.

Cox FM 1996 *An acoustic study of vowel variation in Australian English* Unpublished doctoral dissertation Macquarie University.

Disner SF 1980 'Evaluation of vowel normalisation procedures' *Journal of the Acoustical Society of America* 67: 253–261.

Fant G 1960 *The Acoustic Theory of Speech Production* Mouton The Hague.

Gordon E 1991 'The development of spoken English in New Zealand' in G. MacGregor (ed.) *Dirty Silence: Aspects of Language and Literature in New Zealand* Oxford University Press Auckland: 19–29.

Gordon E & T Deverson 1985 *New Zealand English: an Introduction to New Zealand Speech and Usage* Heinemann Auckland.

Harrington J & S Cassidy 1994 'Dynamic and target theories of vowel classification: evidence from monophthongs and diphthongs in Australian English' *Language and Speech* 37: 357–373.

Harrington J, F Cox & Z Evans 1997 'An acoustic analysis of cultivated, general, and broad Australian English speech' *Australian Journal of Linguistics* 17: 155–184.

Hawkins P 1973 'The sound patterns in New Zealand English' in *Proceedings and Papers of the 15th Congress of the Australasian Universities Language and Literature Association* (vol. 13) Sydney: 1–8.

Hawkins P 1976 'The role of New Zealand English in a binary feature analysis of English short vowels' *Journal of the International Phonetic Association* 6(2): 50–66.

Holmes J 1997 'Setting new standards: sound changes and gender in New Zealand English' *English World-wide* 18(1): 107–142.

Labov W 1990 'The intersection of sex and social class in the course of linguistic change' *Language Variation and Change* 2: 205–254.

Labov W 1994 *Principles of Linguistic Change* (vol. 1: *Internal Factors*) Blackwell Oxford.

Lehiste I & G Peterson 1961 'Transitions, glides and diphthongs' *Journal of the Acoustical Society of America* 33: 268–277.

Lindblom B & J Sundberg 1971 'Acoustical consequences of lip, tongue, jaw, and larynx movement' *Journal of the Acoustical Society of America* 50: 1166–1179.

Maclagan M 1982 'An acoustic study of New Zealand vowels' *NZ Speech Therapist's Journal* 37(1): 20–26.

Maclagan M & E Gordon 1996 'Out of the AIR and into the EAR: another view of the New Zealand diphthong merger' *Language Variation and Change* 8: 125–147.

Millar J, J Harrington & J Vonwiller (in press) 'Spoken language data resources for Australian Speech Technology' *Journal of Electrical and Electronics Engineering Australia.*

Millar J, J Vonwiller, J Harrington & P Dermody 1994 'The Australian National Database of Spoken Language' in *Proceedings of the International Conference on Acoustics Speech and Signal Processing* (vol. 2) Adelaide: 67–100.

Sinclair S & C Watson 1995 'The development of the Otago speech database' in N Kasabov & G Cogill (eds) *Proceedings of the 2nd New Zealand International Conference on Artificial Neural Networks and Expert Systems* IEEE Computer Society Press Los Alamitos: 298–301.

Trudgill P 1986 *Dialects in Contact* Basil Blackwell Oxford.

Trudgill P & J Hannah 1985 *International English, a Guide to Varieties of Standard English* (2nd edn) Edward Arnold London.

Turner GW 1970 'New Zealand English today' in W Ramson (ed.) *English Transported, Essays on Australasian English* Australian National University Press Canberra: 84–101.

Turner GW 1966 *The English Language in Australia and New Zealand* Longman London.

Turner GW 1967 'Samuel McBurney's newspaper article on colonial pronunciation' *Australian Universities Modern Language Association* 27: 81–85.
Vonwiller J, I Rogers, C Cleirigh & W Lewis 1995 'Speaker and material selection for the Australian National Database of Spoken Language' *Journal of Quantitative Linguistics* 3: 177–211.
Wells JC 1982 *Accents of English: Beyond the British Isles* Cambridge University Press Cambridge.