Accepted after abstract only review

The centring diphthongs of New Zealand English in the Intermediate Period
An acoustic analysis

Christian Langstrof

Department of Linguistics
University of Canterbury
New Zealand
cla41@student.canterbury.ac.nz / xian@antiflux.org

Abstract

This paper presents the results of an acoustic analysis of the front centring diphthongs in the speech of New Zealanders from the Intermediate Archive. These are speakers born between the late 1890s and the early 1930s. It will be argued that although these speakers tend to have non-merged realisations of /ɪə/ and /ɛə/, there were already precursors of an incipient merger present at that time. That is, it will be shown that in certain contexts, the distributions of the two phonemes approximate each other over the Intermediate period.

1. Background

The merger of the front centring diphthongs /ɪə/ (as in 'NEAR') and /ɛə/ ('SQUARE[1]') is now one of the most thoroughly studied phonetic phenomena in New Zealand English. There is a wide range of auditory, acoustic and perceptual data available (cf. Batterham 2000, Gordon & Maclagan 1990/2001, Maclagan & Gordon 1996, Watson et al. 2000, Warren et al. 2004, Holmes & Bell 1992). By and large, however, these studies have focussed on contemporary New Zealand English, where the merger is nearly completed as a merger of approximation on the higher vowel /ɪə/[2].

In addition, Watson et al. (2000) have compared the front centring diphthongs of late Mobile Unit speakers (born in the late 19th century, recordings of whom are held at the ONZE (Origins of New Zealand English) corpora at the University of Canterbury) with modern New Zealand English and have concluded that the early speakers kept the two vowel phonemes distinct. It can therefore justifiably be hypothesised that the incipient stages of the merger would be found in the speech of people born between the late 1890s and the 1930s.

The following study will present acoustic data from that period and show that the merger was indeed under way in the speech of 3rd and 4th generation New Zealanders.

2. The Corpus

28 speakers from the Intermediate Archive were selected for acoustic analysis. The corpus consists of approximately 140 New Zealanders who were interviewed on topics such as their family background, their personal as well as general local history. The interviews vary in duration, the majority are between 30 and 60 minutes of running speech. As the interviews were conducted in the early 1990s, the recordings are generally of good quality. This analysis is based on data digitised at a sampling rate of 44 khz. The Intermediate Archive is part of the Origins of New Zealand English (ONZE) project based at the University of Canterbury's department of linguistics. All speakers who were analysed in the present study are native speakers of New Zealand English. Although there are speakers from various parts of New Zealand, both the sample as well as the Intermediate Archive as a whole is biased in favour of speakers from the South Island, especially Otago and Southland.

In order to attain a small-step picture of changes over the Intermediate period, the speaker sample was divided up into six groups; three age cohorts and males/females. The early group consists of speakers born between the mid-1890s and 1905, the medium group of speakers born between 1910 and 1920, and the late group of speakers born between 1925 and the early 1930s. The groups will henceforth be referred to as EARLY MALES/FEMALES, MEDIUM MALES/FEMALES and LATE MALES/FEMALES.

3. Method

The data was analysed with the akustyk plug-in for the PRAAT acoustics software (downloadable from http://bartus.org/akustyk/ and www.praat.org, respectively). For each vowel, two types of measurements were taken. The first one was a point measurement at the turning point in the f2 contour of the nucleus (if no such turning point could be identified, as was frequently the case after

[1] Throughout this paper, IPA symbols will be used in the text, lexical set terms in figures, where NEAR = /ɪə/ and SQUARE = /ɛə/ (cf. Wells 1982).

[2] The details of the merger are, however, considerably more complex. Although there seems to be widespread agreement that the predominant pattern is a merger on /ɪə/, Bell and Holmes (1992) as well as Batterham (2000) found merger on /ɛə/ for some speakers. In addition, experimental data shows that even speakers who are not aware of any differences in the two vowels exploit that difference in speech perception (Warren et al. 2004).

velars, the point measurement was taken at the point of maximum amplitude in the nucleus). The second one was a trajectory measurement over the whole range of the vowel with a sampling rate of 10 ms. The data was subsequently coded for internal factors such as place and manner of articulation of the neighbouring consonant as well one of three stress levels. Tokens that were shorter than 30 ms were excluded from the analysis, as were those that were completely unstressed, i.e. reduced to schwa. Each measurement was checked visually, and if necessary, adjusted by fitting the formant track onto the point(s) of maximal energy concentration by adjusting the LPC filter order. A total of 1094 tokens was measured (401 for /ɪə/, 693 for /ɛə/).

The data was then normalised on the basis of the algorithm proposed in Lobanov (1971), in order to factor out intrinsic (i.e. physiological) differences in vocal tract size that might confound inter-speaker comparability. All values reported in this paper are normalised.

4. Results

4.1 Properties of the nucleus

4.1.1 Means in two-formant space

This section reviews the developments in the nuclei of /ɪə/ and /ɛə/ in the Intermediate Period. Figure 1 plots the mean position for each of the two phonemes for each speaker group. As for /ɪə/, there is an interesting mismatch between the male and female speakers. Whereas the early male group has a mean second formant value of 2393 hz, i.e. a rather forward realisation of the nucleus of /ɪə/, the medium and late males show f2-values that indicate a position that is considerably further back. In the f1-dimension, all three groups of male speakers show similar values. With respect to the second formant, the opposite holds true for the three groups of female speakers. Both the medium as well as the late group show an increase in f2, which suggests that for the female speakers, the vowel underwent fronting over the Intermediate Period. In addition, the late group of female speakers has significantly lower f1 values than the other two (Wilcox rank sum test, W = 1742, p<.05).

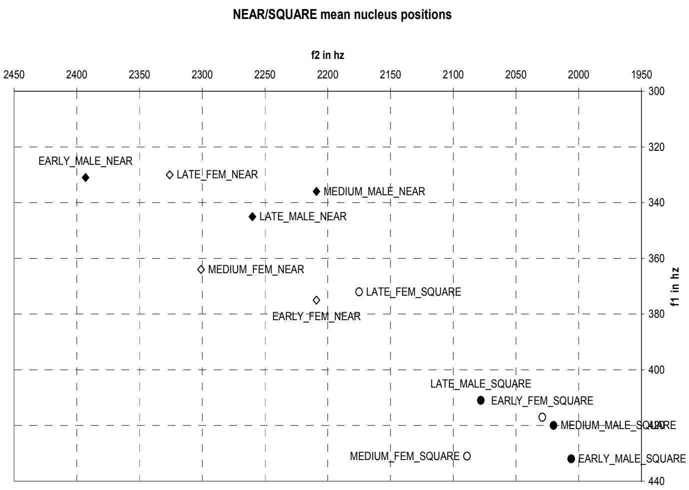

Figure 1 – Means of /ɪə/ and /ɛə/ nuclei in the speech of 6 groups of speakers from the Intermediate Archive.

The behaviour of /ɛə/, on the other hand, is rather more directional in both the male and female groups of speakers. For the males, there are moderate degrees of raising (by 22 hz, W = 6022 p < .01) as well as fronting (by 73 hz, W = 3747 p<.001) over the Intermediate Period. The change over time is rather more dramatic in the speech of the females. The difference in f1 between the early and the late group is in the order of 45 hz (W = 5669, p<.001), the one in f2 is ~146 hz (p<.0001). However, there is also a discontinuity in the female sample: Whereas the early female group has an f1 average of 417 hz, the medium females have a slightly higher mean of 430 hz (W = 8287, p<.05). This shows that the raising and fronting of the /ɛə/ nucleus rapidly gains ground in the speech of the late females, who can therefore be regarded as the main innovators in the sample.

In addition, it can be shown that for the late female group, the distributions of the two vowels are distinct at a considerably lower level of significance than in the other groups. Following Warren et al. (2004), Hotelling-Lawley scores were calculated (Table 1). The Hotelling-Lawley trace is preferable to a distance metric such as the Euclidean distance as it represents the distance between two distributions rather than between two single points. High Hotelling-Lawley scores represent greater distance between distributions. The significance levels associated with the trace represents the degree to which the distributions are significantly distinct. We will return to this point in section 4.3.

speaker group	hotelling-lawley	p
Early females	0.213	< .0001
Medium females	0.454	< .0001
Late females	0.1114	< .05
Early males	1.308	< .0001
Medium males	0.605	< .0001
Late male	0.3136	< .0001

Table 1 – Hotelling-Lawley traces and corresponding significance levels for /ɪə/ - /ɛə/ distributions in six groups of Intermediate speakers.

4.1.2 Conditioning factors

This section will present the results of a CART (classification and regression tree) analysis of a number of both internal and external factors. CART (Breiman et al. 1984) provides a non-parametric regression model which is especially well-suited to handle continuous data. The data is split up into binary branching trees. No prior assumptions are made about the structure of the data. Figures 2 (a) – (c) show trees that analyse the data in terms of the external factors age and sex. Note that there was no partitioning of the /ɪə/ data set in the f2 dimension. Figures 2 (d) – (g) show corresponding trees for the following internal factors: Preceding phoneme/manner+place+voicing of articulation of the preceding phoneme/stress. All trees are truncated at the fourth branching point. The length of the branches does not indicate the magnitude of that division.

With respect to the external factors, the f1 data set is split up twice in both vowels. For /ɛə/, the first division splits off the late speakers (with a lower f1) from the rest. Within the late group, a further division is found where females have lower f1 values than males. This underlines the suggestion of that group being the most innovative in the upward shift of /ɛə/. In the /ɪə/ data, the first branching sets off male from female speakers and recognises a further split in the latter group, based on age. That is, females have more open realisations of /ɛə/ over the entire sample, but late females have a closer /ɛə/ than their early and medium counterparts.

As for the second formant, external factors seem to be a poor predictor for the variation in both vowels. There is one split in the /ɛə/ data whereby late speakers have more fronted realisations, and no division in /ɪə/.

If only internal factors are considered (fig. 2 (d) - (g), where sh = /ʃ/, dh = /ð/ and th = /θ/), it turns out that the preceding phoneme as well as the stress level play a role in the structure of the data. More specifically, the preceding phoneme is the single most powerful predictive factor in all four trees. However, the specific subgroupings in each tree seem to be rather idiosyncratic in each tree, with the exception of /0/ and /b/, which are grouped together in all four data sets as a conditioning factor of forward and closed realisations in both lexical sets. In addition, stress level shows up as a factor in the f1 dimension of /ɪə/ (where tokens of low stress tend to occupy a position toward the less peripheral parts of the vowel space within the division of close phonemes) as well as in the f2 dimension of both vowels. The division in the f2 dimension of /ɪə/ mirrors that in f1, whereas for /ɛə/ a stress division is found in both 1st order subgroups.

SQUARE f1
root (418.7 hz)

age (395.9 hz) **late** | age (424 hz) **early, medium**

sex (372.3 hz) **female** | sex (410.8 hz) **male**

fig. 2 (a)

SQUARE f2
root (2049 hz)

age (2034 hz) **early, medium** | age (2116 hz) **late**

fig. 2 (b)

NEAR f1
root (350.4 hz)

sex (336.6 hz) **male** | sex (363.5 hz) **female**

age (329.7 hz) **late** | age (369.7 hz) **early, medium**

fig. 2 (c)

SQUARE f1
root (418.7 hz)

p_phoneme (390.6hz) **0,b,m,n,r,sh,t** | p_phoneme (424.6 hz) **d,dh,f,h,k,l,p,w,wh**

p_phoneme (421.1 hz) **dh,f,k,l,p** | p_phoneme (443.4 hz) **d,h,w,wh**

fig. 2 (d)

NEAR f1
root (350.4 hz)

p_phoneme (340.5 hz) **0,b,d,h,j,l,sh,t,w** | p_phoneme (390.2 hz) **f,m,n,p,r,s,z**

stress (333.3 hz) **heavy, medium** | stress (366.3 hz) **light**

fig. 2 (e)

SQUARE f2
root (2049 hz)

p_phoneme (1936 hz) **d,f,p,r,t,w,wh** | p_phoneme (2087 hz) **0,b,dh,h,k,l,m,n,sh**

stress (1886 hz) **medium, light** | stress (1996 hz) **heavy** | stress (2002 hz) **light** | stress (2116 hz) **heavy, medium**

fig. 2 (f)

NEAR f2
root (2272 hz)

p_phoneme (2117 hz) **d,m,p,r,s,sh** | p_phoneme (2286 hz) **0,b,f,h,j,l,n,t,w,z**

stress (2219 hz) **light** | stress (2308 hz) **heavy, medium**

fig. 2 (g)

Figure 2 (a) –(g) - CART regression trees for /ɪə/ and /ɛə/ in the speech of 6 groups of speakers from the Intermediate Archive. Fig. 2 (a) – (c) show the partitioning of the overall data in terms of the external factors age/sex, figs 2 (d) – (g) show phonemic factors. The numbers in brackets indicate the formant frequency average at that node. p_phoneme stands for the coding category 'preceding phoneme'.

4.2 Dynamic features

4.2.1 Trajectories

Having discussed the development of the nuclei of the front centring diphthongs in the Intermediate Period of New Zealand English, we will now turn to the results of the trajectory analysis. Figure 3 models the vowel trajectories of both vowels for all six speaker groups. The models were obtained by fitting a non-parametric scatterplot smoother through time-normalised representations of the vowels (cf. Cleveland 1979). That is, each observation is plotted in its relative position within the diphthong token it occurs in.

Accepted after abstract only review

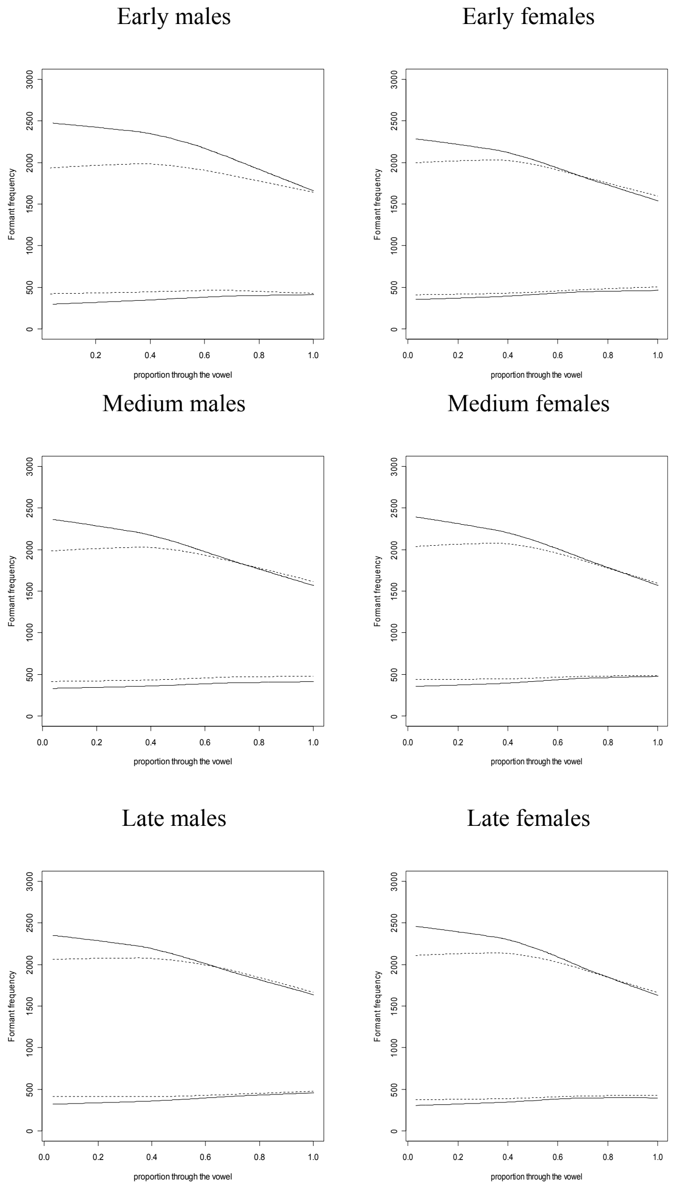

figure 3 – scatterplot smoother models of /ɪə/ (solid lines) and /ɛə/ (dashed lines) trajectories in the speech of 6 groups of speakers from the Intermediate Archive. Males are plotted in the left column, early speakers in the top row.

It can be observed that although the trajectories in figure 3 support the above conclusions that the merger of the front centring diphthong was well under way during the Intermediate Period, the 'partitioning' of speaker groups seems to be somewhat different. That is, whereas we have concluded above that the merger gains momentum in the late female group, the trajectories suggest a sub-grouping along the lines of early males vs. all others. More specifically, it is in that group where the trajectories look as one would expect from a dialect that keeps the two vowel phonemes apart. Both formants are distinct along the entire trajectory, whereas there is overlap in one or both formants in all the other groups. The medium males keep the first formants of both vowels distinct over the entire trajectory, but show overlap in f2 toward the offglide. The late male group shows the same kind of f2 overlap (if somewhat earlier in the trajectory) as well as approximating f1 contours toward the offglide.

As for the females, all three groups have some degree of overlap in the f2 contour toward the offglide as well as closely bundled f1 contours over the entire trajectory. In keeping with current notions in sociolinguistics that males generally show more conservative vowel realisations in sound change from below than females of the same age, we could therefore justifiably date the precursors of the /ɪə/ - /ɛə/ merger (if we accept overlap in the offglide as a precursor to the modern pattern of complete overlap) further back than the nucleus analysis suggests.

However, a qualification is in order with respect to the trajectory patterns: The trajectories plotted in figure 3 look somewhat different from those given in Watson et al. (1998) for Australian and New Zealand English. Specifically, it seems that the crossover point in the f2 contour is rather more well-defined in the Australian English data, which does not have merger of /ɪə/ and /ɛə/. That is, it is a property of an unmerged system that the formant contours of /ɛə/ are fully included within those of /ɪə/ toward the offglide. On the other hand, it should be noted that the methodologies are rather different between this study and Watson et al's. Whereas the data presented here is from running speech and incorporates any phonemic environment as well as different stress levels, Watson et al.'s work is based on word list production with well-specified phonemic environments (only tokens followed by an alveolar stop were considered in the NZ data, and /hVd/ and /hV#/ in the Australian data). However, although this might account for the degree of difference in the offglide patterns between the two studies, it is hard to see why this should bring about differences in type. A hypothesis could be that there is a disproportionally higher number of following environments in /ɪə/ which trigger monophthongal realisations of the vowel or, more generally, higher f2 loci toward the offglide. No analysis of this has been carried out yet, but will be topic of future research. However, the data presented in the next section suggests that it is the offglide of /ɪə/ that more consistently specified.

4.2.3 Properties of the offglide

Having identified the offglide as the potential locus of early overlap between the front centring diphthongs, we will now turn to a more detailed discussion regarding offglide position and target specification. Since offglides tend to be somewhat weakly specified in comparison to nuclei, it was not always possible to identify comparable positions for point measurements in the offglide. The following analysis is therefore based on the average formant value between 70% and 90% through each diphthong token. Hence, 'offglide' here refers to an average value of a number of (between 1 and 6, depending on the overall length of the diphthong) observations. Table 2 states the mean formant frequency values as well as the standard deviation and the first and third quartile of the offglide for all six speaker groups. Although there do not seem to exist clear directional patterns over the entirety of the sample, two points are worth noting: In all three groups of male speakers, the standard deviation of the formant frequency values is consistently higher in /ɛə/ than in /ɪə/, which implies that the former shows more variation in the offglide than the latter for those speakers. Secondly, the highest degree of freedom in the offglide realisation of both vowels seems to hold in the speech of the late speakers, although not in a very straightforward way: Whereas the late females have the highest STDevs in the f2 dimension, the late males allow for a higher degree of freedom in f1.

Accepted after abstract only review

	E Ma N		*E Ma SQ*	
	F1	F2	F1	F2
sd	65.6	201	99.7	250.8
mean	417.9	1921	491	1803
1stQ	382.8	1810	430.6	1640
3ndQ	445.5	2065	526	1920

	M Ma N		*M Ma SQ*	
	F1	F2	F1	F2
sd	74.6	223.1	90.19	229
mean	430.1	1745	494.2	1790
1stQ	381.8	1595	433.5	1617
3ndQ	456.8	1880	543.8	1926

	L Ma N		*L Ma SQ*	
	F1	F2	F1	F2
sd	118.6	195.6	144.1	216
mean	469.9	1808	493.9	1859
1stQ	397.3	1701	404.8	1698
3ndQ	519.5	1927	533.7	1859

	E F N		*E F SQ*	
	F1	F2	F1	F2
sd	108	223	84.5	231.5
mean	475.7	1736	493.7	1751
1stQ	392.4	1561	428.9	1597
3ndQ	549.6	1860	558.3	1886

	M F N		*M F SQ*	
	F1	F2	F1	F2
sd	89	183.4	96	206.1
mean	481.8	1764	509.4	1781
1stQ	422.7	1664	443	1633
3ndQ	543.7	1880	568.7	1924

	L F N		*L F SQ*	
	F1	F2	F1	
sd	105.5	259.5	109.7	298.5
mean	420.6	1804	459.3	1836
1stQ	369.7	1702	379.9	1638
3ndQ	456.7	1960	514.9	2006

Table 2 – Mean formant frequency averages, standard deviations and 1st and 3rd quartile in the offglide of /ɪə/ and /ɛə/ in the speech of six groups of Intermediate speakers. E/M/L = early/medium/late speakers; Ma/F = male/female speakers; N = /ɪə/; SQ = /ɛə/.

One might wonder how much importance should be assigned to offglide properties in the first place, given that the most salient property of the merged system in current NZE is overlap in the nucleus. In addition, the traditional phonemic notation of the offglide as /ə/ in most non-rhotic English dialects indicates that there has not been much of a salient auditory difference in the past between the offglides of /ɪə/ and /ɛə/. Yet, the development in the speech of (especially the male) speakers in this study suggests a rather linear development where the two trajectories are kept apart in the earliest male group, and subsequent approximation toward the offglide region of the diphthongs.

4.3 Stress

Finally, we will consider the role of stress in the developments in /ɪə/ and /ɛə/ over the Intermediate Period. Maclagan and Gordon (1996) stated that stress might have a role to play in the degree of displacement of /ɛə/ tokens, whereby the higher the degree of stress is, the closer the vowel. Table 3 gives Hotelling-Lawley scores and significance levels for all speaker groups and stress levels. In section 4.1.1 it was noted that although the late females keep their overall distributions of the two vowels distinct, they do so at a considerably lower level of significance than is the case in the other groups. Additionally, it can be shown that this effect is much stronger if the data is resolved for different levels of stress. In comparison to the data given in table 1, it can be seen that the front centring diphthong distributions in the speech of the late females are not kept apart at levels of significance below .05 if the data is resolved for stress. In addition, it should be noted that whereas there is at least a trend towards keeping them distinct for tokens of stress level 2 and 3 (i.e. medium and low), sounds in heavily stressed position show a much lower Hotelling-Lawley score and a p-value of .3, which indicates a fairly advanced stage of overlap between these sub-distributions.

A similar pattern holds in the late male group, where the Hotelling-Lawley score is considerably lower in stressed contexts, although the vowels are kept distinct in this group for all stress levels, but at an appreciably lower level of significance in tokens of stress level 1. With respect to the early and medium groups of both sexes, it can be observed that solid distinctions between /ɪə/ and /ɛə/ are maintained for any level of stress.

Two general points are worth noting: First, for any two comparable sex cohorts, it is the female speakers who tend to show lower levels of differentiation of the two vowel phonemes. This effect is particularly striking in the late sample, but also present in the sub-distributions of vowels of stress level 3 in the early and medium groups. Furthermore, there is a striking mismatch between the absolute distance measures (i.e. the H-L traces) and the significance levels in the early female group, which suggests a rather tight clustering of the two distributions in these groups. This effect, however, as well as the global one mentioned above with respect to significance levels in the late samples is probably in part ascribable to smaller token numbers in those sub-samples. This caveat is also supported by the author's auditory impression, which suggested that although the realisation of /ɛə/ is rather high in articulatory space in especially the late female group, most tokens were heard 'correctly' (~ 1/6 of all /ɛə/ tokens were auditorily coded as /ɪə/ in that group, compared to ~ 1/15 for the entire sample including the late female group, and ~ 1/17 excluding that group).

Accepted after abstract only review

Speaker group	Stress	Hotelling-Lawley trace	p	n
Early females	1	0.1721	< .0001	100
	2	0.3911	< .0001	81
	3	0.3145	< .01	51
Mediumfemales	1	0.6059	< .0001	74
	2	0.823	< .0001	79
	3	0.3277	< .01	50
Late females	1	**0.09426**	**.2964**	46
	2	**0.2410**	**.08349**	36
	3	**0.23539**	**.08796**	27
Early males	1	1.427	< .0001	68
	2	1.306	< .0001	59
	3	2.029	< .0001	55
Medium males	1	0.5442	< .0001	81
	2	0.655	< .0001	100
	3	0.8193	< .0001	51
Late males	1	0.1326	< .05	57
	2	1.0895	< .0001	48
	3	0.9248	< .001	31

Table 3 – Hotelling – Lawley scores and significance levels for /ɪə/ - /ɛə/ distributions in the speech of six groups of Intermediate speakers, resolved for three levels of stress. Numbers in bold indicate distributions that are not significantly distinct at $p < .05$. Stress levels are stated in descending order prominence, i.e. 1 = heavy stress, 2 = medium stress and 3 = low stress.

5. Conclusion

The main conclusion to be drawn from the data analysed above is that with respect to the vowel phonemes /ɪə/ and /ɛə/ in New Zealand English, the modern merged system can be shown to have precursors in the Intermediate Period. Over that period, the nuclei of /ɛə/ undergoes a substantial shift in the direction of /ɪə/. It was furthermore pointed out that overlapping f2 contours toward the offglide already occur in the early female sample. Finally, it was shown that the nuclei of stressed vowels approximate each other at higher rates than those that have neutral or low stress.

6. Acknowledgements

This research was funded by scholarships from the *Deutscher Akademischer Austauschdienst* (DAAD) as well as the University of Canterbury. The data was collected by Rosemary Goodyear, Sandra Quick and members of the ONZE team. The work done by members of the ONZE project in preparing the data, making transcripts and obtaining background information is also acknowledged.

In addition, I would like to thank Dr. Jen Hay and Dr. Margaret Maclagan for comments on earlier drafts of this paper.

References

Batterham, M. (2000). The apparent merger of the front centring diphthongs – EAR and AIR – in New Zealand English. In Bell & Kuiper (eds.) *New Zealand English*. Wellington: Victoria University Press.

Breiman, L., Friedman J. H., Olshen, R.H. & Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth: Belmont.

Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74: 829-836.

Gordon, E. & Maclagan, M. (1990). A longitudinal study of the 'ear/air' contrast in New Zealand speech. In Bell and Holmes (eds.) *New Zealand Ways of Speaking English*. Clevedon: Multilingual Matters.

Gordon, E. and Maclagan, M. (2001). Capturing a Sound Change: A Real Time Study over 15 Years of the NEAR/SQUARE Diphthong Merger in New Zealand English. *Australian Journal of Linguistics* 21.2: 215-38.

Holmes, J. & Bell, A. (1992). On shear markets and sharing sheep: The merger of EAR and AIR diphthongs in New Zealand English. *Language Variation and Change* 4: 251 – 273.

Lobanov, B.M. (1971). Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* 49: 606 – 608.

Maclagan, M. & Gordon, E. (1996). Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change* 8: 125-149.

Watson, C. I., Harrington , J. & Evans, Z. (1998). An acoustic comparison between New Zealand and Australian English vowels. Australian Journal of Linguistics 18.2: 185 – 207.

Watson, C. I., Maclagan, M. A. & Harrington, J. (2000). Acoustic evidence for vowel change in New Zealand English. *Language Variation and Change* 12: 51-68.

Warren, P., Hay, J. and Thomas, B. (2004). The loci of sound change effects in recognition and perception. Paper presented at the ninth conference on Laboratory Phonology in Champaign-Urbana in June 2004.

Wells, J.C. (1982). *Accents of English*. Cambridge: Cambridge University Press.