A PLEASANT MALADY:
THE ELLEN/ALLAN MERGER IN NEW ZEALAND ENGLISH

Brynmor Thomas and Jennifer Hay: *Department of Linguistics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand <Corresponding author: jen.hay@canterbury.ac.nz>*

Abstract

This paper reports on a series of production and perception experiments designed to investigate the merger of /el/ and /æl/ in New Zealand English (NZE). 16 young NZE speakers completed a range of tasks involving nonsense and real words containing pre-lateral /e/ and /æ/. In production, both vowels were produced in an area of acoustic space closer to non-prelateral /æ/ than /e/, and were also noticeably centralised. This was true of all speakers, although the degree of merger varied across participants, and was partially conditioned by social class. While most participants displayed merger in production, all were still largely able to exploit the distinction for the purposes of speech perception. The accuracy rates in the perception tasks were considerably higher for /el/ forms than /æl/ forms, and highest in monosyllabic words.

1. Introduction and Background

New Zealand English has a merger of DRESS[1] and TRAP before /l/. This has been the subject of many indignant letters to the editor, and is typically included in descriptive inventories of features of New Zealand English (e.g. Maclagan 1982, Bauer 1986, Gordon and Deverson 1998, Bauer and Warren 2004. However very little systematic study has been carried out on the merger.

Bayard (1987) included the variable in his study of New Zealand English and found considerable variation in the degree of merger. This variation appeared to be conditioned largely by social class, and did not vary by age. More recently Horsfield (2001) conducted acoustic analysis of 11 speakers. She found that most of the speakers merged DRESS and TRAP before /l/. The merged tokens occupy a space close to non-prelateral TRAP, but slightly more centralised and a little lower. Another recent study was conducted by Buchanan (2001) who conducted production and perception tasks with two participants – one older, and one younger. She found an equal degree of merger in production for both participants. She also found that the speakers had difficulty in accurately identifying words when asked to listen to their own speech.

The experiments to be reported here follow up on the results of a small pilot study reported in Thomas (2003), which yielded perception and production data from four NZE speakers aged between 18 and 25. These subjects were largely merged in production. However all participants performed above chance in a task which required them to distinguish prelateral TRAP and DRESS in the voice of a speaker who did produce the distinction. This suggests that the distinction may still be available to young NZers in speech perception. This paper studies the variable on a larger scale – including more participants, and a wider range of experimental tasks.

The experiments reported in this paper required participants to complete a set of production tasks and a set of perception tasks. Participants completed the production tasks first, followed by the perception tasks. The reasoning for this was governed by the nature of the tasks themselves: the perception tasks required participants to listen to the recording of a speaker, who distinguished between /el/ and /æl/. If participants were to listen to this recording before completing the production tasks, it could potentially have influenced their own productions. In addition, two of the production tasks were designed to elicit productions of the target words (words containing /el/ or /æl/) in isolation (that is, without their minimal pair), hidden amongst many filler words. These tasks were designed to be undertaken without the participant being aware of the target variable. However, had the perception tasks preceded these production tasks, such naive elicitations would be impossible.

The participants in the tasks were all university students or recent university graduates, aged between 18 and 30 years, and all native speakers of New Zealand English. There were no Maori English speakers. In total, there were 16 speakers, 8 males and 8 females.

2. Lexical Items

Ten real word minimal pairs were used in both the perception and production sections. These pairs were: *celery-salary, melody-malady, Ellen-Alan, telly-tally, Kelvin-Calvin, pellet-palate, shell-shall, mellow-mallow, elf-Alf, Ellie-alley*. Two extra near minimal pairs, *elevator-alligator* and *melon-mallet*, were also used in the production tasks.

A set of nonsense words were also included in the experiments. Twenty-two minimal pairs comprised the full set of nonsense words, half of which were monosyllabic words and half were disyllabic words. The reason for having the words divided into monosyllabic and disyllabic categories was that by comparing the degree of merger amongst the monosyllabic words with the degree of merger amongst the disyllabic words one would effectively be comparing the degree of merger in instances where the post-vocalic /l/ falls in a coda position and instances where it does not. This comparison was desirable in order to determine what effect syllable structure may or may not have had on the rate of merger. The nonsense word minimal pairs used in this experiment are listed in Table 1.

Table 1: Nonsense Word Minimal Pairs.

MONOSYLLABIC MINIMAL PAIRS			DISYLLABIC MINIMAL PAIRS		
sal	—	sel	sallit	—	sellit
zal	—	zel	zallit	—	zellit
fal	—	fel	falit	—	fellit
val	—	vel	vallit	—	vellit
chal	—	chel	challit	—	chellit
jal	—	jel	jallit	—	jellit
ral	—	rel	rallit	—	rellit
lal	—	lel	lallit	—	lellit
kal	—	kel	kallit	—	kellit
gal	—	gell	galit	—	gellit
tal	—	tel	tallit	—	tellit
dal	—	del	dalit	—	dellit
nal	—	nel	nallit	—	nellit
mal	—	mel	malit	—	mellit

3. Production Tasks

The production section of the experiment consisted of five separate production tasks: three using real words, and two using nonsense words. For all five tasks participants were presented with a list of printed words and were asked to read the words out loud. The production tasks used the full subset of real words (including the near minimal pairs), and the full subset of nonsense words.

The participants were recorded on a Sony TCM-5000EV portable cassette-corder using a Sony ECM-F8 electret condenser microphone These recordings were digitized at 44.1kHz, and later analysed on computer using Praat v.4.1.2, a computer programme for speech analysis and synthesis (Boersma and Weenink. The application for Praat and its documentation are both available online at: http://www.praat.org).

Care was taken, when extracting the formant frequencies, to draw the frequency measurement from the midpoint of the vowel in order to minimise the effect of the following /l/, and also to minimise the effect of any formant transitions from preceding consonants. All formant readings were checked to ensure the validity of the values.

For a more detailed description of the production tasks methodology, see Thomas 2004.

3.1 Degree of distinction in non-prelateral position

Task one was a control task, intended to yield baseline productions for /e/ and /æ/. Participants were asked to read aloud a word list, which did not include any of the target words, but which included words containing the vowels /e/ or /æ/ preceding /d/. The F1 and F2 values for these target vowels were recorded and used as a basis for comparison in the final analysis. Speakers were asked to read each word only once. The word list contained six words in total.

Table 2 lists the Euclidean distances (calculated from F1 and F2) between baseline DRESS and TRAP for all speakers. The data in table 2 show that there is variation in the degree to which baseline DRESS and TRAP are kept distinct, but that even so, all of the participants – both male and female – produced a clear distinction

The variation is particularly noticeable in the female speakers, where the Euclidean Distance values range from 115.5 for female 1, to 931.8 for female 3. For the male speakers the degree of DRESS and TRAP distinction was less varied, with Euclidean distances ranging from 240.6 for male 1 to 379.1 for male 6.

Table 2. Euclidean Distance Between Baseline DRESS and TRAP for all Speakers.

	ED		ED
Female 1	115.5	**Male 1**	240.6
Female 2	545.9	**Male 2**	286.5
Female 3	931.8	**Male 3**	332.9
Female 4	263.3	**Male 4**	295.6
Female 5	630.0	**Male 5**	352.3
Female 6	237.6	**Male 6**	379.1
Female 7	424.3	**Male 7**	347.3
Female 8	416.0	**Male 8**	317.4

3.2 Production of Real Words

The participants were then asked to read aloud a word list of 128 real words divided into rows of four. Each row of four words contained a maximum of one target word, which, to avoid the effects of 'list-reading intonation' were not entered as the first or last word of the row. All of the words appeared in a random order. The aim of this task was to have the target words produced in isolation from each other. Not every row contained a target word.

Table 3 shows the averaged F1 and F2 frequencies for /el/ and /æl/ tokens for each speaker.

Table 3. Averaged F1 and F2 Frequencies for all speakers for Real Word Tokens Produced in Isolation (based on twelve tokens per speaker).

	/æl/ TOKENS		/el/ TOKENS			/æl/ TOKENS		/el/ TOKENS	
	F1	F2	F1	F1		F1	F2	F1	F2
Female 1	702.5	1626.2	740.1	1882.2	**Male 1**	481.5	1575.7	494.0	1584.2
Female 2	730.0	1806.4	730.5	1824.0	**Male 2**	557.9	1623.0	528.7	1627.9
Female 3	673.6	1800.6	667.9	1714.7	**Male 3**	570.6	1588.1	557.9	1553.9
Female 4	703.2	1930.5	672.2	1915.0	**Male 4**	857.0	1768.3	842.9	1788.1
Female 5	857.0	1768.3	842.9	1788.1	**Male 5**	587.5	1673.8	600.2	1699.2
Female 6	837.2	2001.1	740.5	1788.7	**Male 6**	831.6	1600.4	670.8	1738.7
Female 7	665.8	1875.8	749.7	2067.4	**Male 7**	579.1	1589.0	565.0	1583.5
Female 8	883.8	1939.0	862.6	2050.5	**Male 8**	625.6	1676.6	641.1	1778.2

The data presented in table 3, show that, in general, neither the male nor the female speakers produce any consistent distinction between pre-lateral DRESS and TRAP in the target real words when read in isolation.

In a later task, participants were asked to produce the same words in minimal pairs. There was no significant difference between the data presented here for the production of words in isolation, and the production in minimal pairs. In neither case was there a consistent difference maintained for DRESS and TRAP before /l/.

For each subject, we calculated the Hotelling-Lawley trace between their TRAP and DRESS vowels, taking the tokens read in isolation and minimal pairs together. This is a measure of the distinction between two distributions. Only 4 of the 16 subjects displayed a significant difference between their pre-lateral DRESS and TRAP vowels (female 3 and 8, and male 6 and 7).

3.3 Production of Nonsense Words

The nonsense words production tasks followed the same design as for the real words.

In the 'words in isolation' condition participants were presented with a list of 175 words, divided into rows of five. The words used in this task were all nonsense words, including the fillers. Not every row contained a target word, and no row contained more than a maximum of two target words.

Tables 4 and 5 show the averaged F1 and F2 frequencies for this task, divided into monosyllabic nonsense words and disyllabic nonsense words respectively.

Looking at the figures in table 4, there does not appear to be an overall consistent distinction between pre-lateral DRESS and TRAP in isolated production of monosyllabic nonsense words for either male or female speakers. However, when DRESS and TRAP are compared for each speaker individually, most speakers (15 out 16) exhibit a higher F1 frequency and a lower F2 frequency for pre-lateral TRAP tokens, and a concurrently lower F1 frequency and higher F2 frequency for pre-lateral DRESS tokens. That is, the TRAP vowels tend to be lower and more central than the DRESS vowels. This differs from the values presented in Table 3 for isolated real word productions, in which the F1 and F2 distributions were less ordered and fewer than half the subjects show differences in the expected direction.

The averaged F1 and F2 frequencies for each speaker's production of disyllabic nonsense words in isolation are presented in table 5. Again, there appears to be a trend for the F1 frequencies to be higher for pre-lateral TRAP

Table 4. Average F1 and F2 Frequencies for Monosyllabic Nonsense Words in Isolation, for All Speakers (based on fourteen tokens per speaker).

	/æl/ TOKENS F1	/æl/ TOKENS F2	/el/ TOKENS F1	/el/ TOKENS F1		/æl/ TOKENS F1	/æl/ TOKENS F2	/el/ TOKENS F1	/el/ TOKENS F2
Female 1	636.3	1991.8	620.9	2019.8	**Male 1**	440.8	1584.7	438.9	1566.6
Female 2	660.9	1808.4	653.6	1845.9	**Male 2**	520.6	1608.9	502.5	1642.7
Female 3	709.2	1751.6	708.4	1785.3	**Male 3**	567.8	1538.8	564.2	1564.2
Female 4	723.8	1859.2	687.6	1949.4	**Male 4**	812.0	1754.9	785.4	1802.6
Female 5	812.0	1754.9	785.4	1802.6	**Male 5**	556.9	1717.7	552.1	1719.9
Female 6	698.4	2026.1	614.4	2168.9	**Male 6**	752.8	1602.8	605.3	1764.9
Female 7	725.0	2086.5	707.1	2117.4	**Male 7**	619.8	1527.9	542.4	1744.3
Female 8	943.8	1624.6	756.6	1935.1	**Male 8**	691.1	1705.7	646.4	1715.3

Table 5. Average F1 and F2 Frequencies for Disyllabic Nonsense Words in Isolation, for All Speakers (based on fourteen tokens per speaker).

	/æl/ TOKENS F1	/æl/ TOKENS F2	/el/ TOKENS F1	/el/ TOKENS F1		/æl/ TOKENS F1	/æl/ TOKENS F2	/el/ TOKENS F1	/el/ TOKENS F2
Female 1	662.2	1802.7	611.9	1900.6	**Male 1**	458.9	1526.7	432.2	1596.5
Female 2	648.8	1837.6	654.8	1841.7	**Male 2**	507.3	1584.7	496.4	1566.6
Female 3	664.5	1743.1	665.7	1722.6	**Male 3**	564.2	1570.4	556.9	1614.9
Female 4	728.6	1824.1	685.1	1842.3	**Male 4**	807.2	1725.0	820.4	1790.6
Female 5	807.2	1725.0	820.4	1790.6	**Male 5**	594.4	1663.3	595.6	1650.0
Female 6	732.2	1931.2	684.8	1959.8	**Male 6**	725.0	1674.2	631.9	1738.4
Female 7	699.6	2051.5	705.8	2017.1	**Male 7**	623.4	1549.6	556.9	1671.8
Female 8	845.9	1748.0	747.9	1936.6	**Male 8**	738.3	1659.7	683.9	1673.0

and lower for pre-lateral DRESS (10 out of 16), and for the F2 frequencies to be lower for pre-lateral TRAP and higher for pre-lateral DRESS (15 out of 16). However, this trend is not as clear in the disyllabic data as it is in the monosyllabic data, but is still more noticeable than in the real word data presented in table 3.

In a later task, the participants were required to produce the same words in minimal pairs. The overall results were similar, with a slightly greater degree

of distinction maintained in the minimal pair condition than when the words were produced in isolation. Calculating Hotelling Lawley trace scores, and taking together all of the nonsense word productions for each speaker, we find that 8 of the 16 speakers maintained a significant difference between DRESS and TRAP in nonsense words.

3.4 Direction of the merger

Figures 1 to 4 show the acoustic location of all tokens of pre-lateral DRESS and TRAP, divided into real word and nonsense word productions for male and female speakers.

Figure 1. Real Word /el/ and /æl/ Productions for Female Speakers Plotted with their Average Baseline DRESS and TRAP Projections.

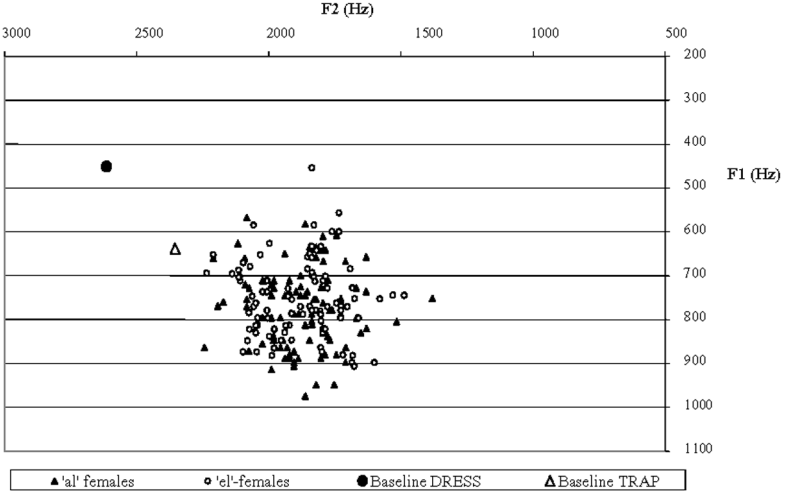


Figure 1 shows the real word /el/ and /æl/ tokens occupying a centralised space closer to the average baseline TRAP projection than to the average baseline DRESS. The same can be seen in the real word productions for the male speakers, given in figure 2.

Figure 2. Real Word /el/ and /æl/ Productions for Male Speakers Plotted with their Average Baseline DRESS and TRAP Projections.

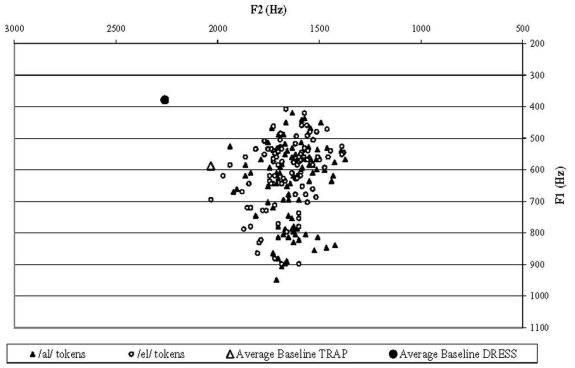


Figure 2 shows that, like the female speakers, the male speakers also produced pre-lateral DRESS and TRAP tokens in a centralised area, closer to baseline TRAP than DRESS. Similar results can be seen for the nonsense word productions, illustrated in figures 3 and 4

Figure 3. Nonsense Word /el/ and /æl/ Productions for Female Speakers Plotted with their Average Baseline DRESS and TRAP Projections.

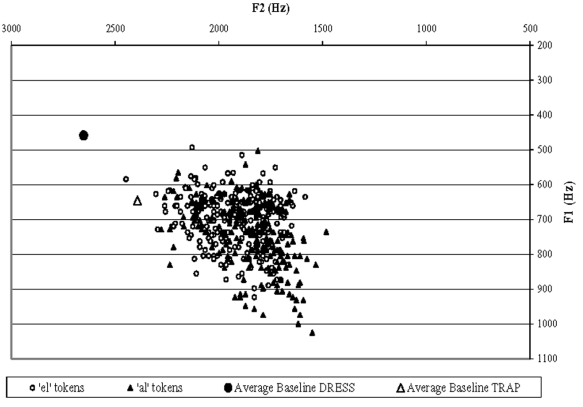


Figure 4. Nonsense Word /el/ and /æl/ Productions for Male Speakers Plotted with their Average Baseline DRESS and TRAP Projections.

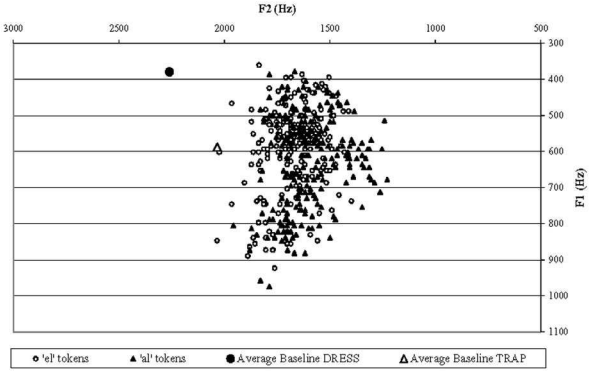


Figure 3 shows that the female speakers' nonsense word productions are falling in an area closer to baseline TRAP than DRESS, and are noticeably centralised. A similar situation can be seen with the male speakers' nonsense word production, presented in figure 4.

Figure 4 shows that the male speakers produced (nonsense word) pre-lateral DRESS and TRAP tokens across a wider acoustic area than the female speakers, but that most of these tokens fall closer to baseline TRAP than baseline DRESS, and that, again, these tokens are noticeably centralised.

What is clear from figures 1-4 is that almost none of the pre-lateral DRESS or TRAP tokens are as high and front as baseline DRESS. Most of these tokens are lower and further back than baseline TRAP. These results concur with those found in Buchanen (2001), Horsfield (2001) and Thomas (2003).

3.5 The relationship between prelateral and non-prelateral DRESS and TRAP

The production data were analysed to investigate what effect, if any, the degree to which a speaker kept baseline DRESS and TRAP distinct affected the degree to which that same speaker kept pre-lateral DRESS and TRAP distinct. This analysis was conducted on the real words and the nonsense words separately. The relevant data for the real word analysis are presented in table 6. The Euclidean distance measurement simply quantifies the distance

between two points, whereas the Hotelling Lawley (HL) score operates on Bark values derived from F1 and F2 at the first target of the vowel to quantify the degree of phonetic distinction between two vowel distributions (in this case, /el/ and /æl/).

As can be seen in table 6, there is no correlation between the degree to which speakers keep the baseline DRESS and TRAP productions distinct and the degree to which the pre-lateral DRESS and TRAP in real word production are kept distinct (S=634, p-value=0.8).

Table 6. Hotelling-Lawley Trace for Real Words, and Euclidean Distance between Baseline DRESS and TRAP for Each Speaker.

	ED BETWEEN /e/ AND /æ/	HL REAL WORDS		ED BETWEEN /e/ AND /æ/	HL REAL WORDS
Female 1	115.5	0.11506	**Male 1**	240.6	0.006433
Female 2	545.9	0.00221	**Male 2**	286.5	0.008380
Female 3	931.8	0.16790	**Male 3**	332.9	0.150200
Female 4	263.3	0.03240	**Male 4**	295.6	0.027110
Female 5	630.0	0.02711	**Male 5**	352.3	0.005164
Female 6	237.6	0.05669	**Male 6**	379.1	2.999000
Female 7	424.3	0.03148	**Male 7**	347.3	0.271200
Female 8	416.0	0.21500	**Male 8**	317.4	0.047580

A parallel analysis of nonsense words reveals that there is no correlation between the degree to which speakers keep the baseline DRESS and TRAP productions distinct and the degree to which they keep pre-lateral DRESS and TRAP distinct in nonsense words (S=719, p-value=0.83).

This finding suggests that the merging of pre-lateral DRESS and TRAP is not related in any way to the phonetics of non-pre-lateral DRESS and TRAP, as the pre-lateral vowels behave completely separately from the baselines.

3.6 Syllable structure effects

The data from the production tasks was analysed to investigate the effect of syllable structure on the pre-lateral DRESS-TRAP merger. As described earlier, the set of nonsense words used in this experiment was divided into

monosyllabic and disyllabic words. The F1 and F2 frequencies for the mono- and di-syllabic nonsense words were compared and contrasted to investigate what effect, if any, the difference in syllable structure had on the quality of the pre-lateral DRESS or TRAP vowels. The data used in this analysis is presented in table 7. The F1 and F2 frequencies given in the table are averaged over the all the tokens of each given word, both when produced in isolation and when produced in minimal pairs, across all speakers.

Table 7. Averaged F1 and F2 Frequencies for Monosyllabic and Disyllabic Nonsense Words.

	F1		F2			F1		F2	
INITIAL	MONO	DI	MONO	DI	INITIAL	MONO	DI	MONO	DI
/tʃæ/	676.3	695.9	1760.1	1672.4	/tʃe/	614.0	675.8	1857.8	1745.6
/dæ/	655.1	640.4	1756.3	1776.5	/de/	624.1	621.3	1867.2	1852.1
/fæ/	695.3	682.0	1691.1	1656.3	/fe/	646.7	653.7	1809.7	1721.3
/gæ/	663.9	629.2	1845.1	1868.8	/ge/	600.6	601.6	1886.3	1919.5
/dʒæ/	647.6	653.0	1796.8	1768.1	/dʒe/	590.7	592.7	1899.7	1851.5
/kæ/	720.8	701.1	1746.7	1711.7	/ke/	674.8	663.8	1843.2	1711.9
/læ/	707.5	694.9	1644.6	1668.4	/le/	653.2	630.8	1794.3	1779.1
/mæ/	719.8	730.3	1765.4	1689.1	/me/	695.7	688.3	1836.8	1783.4
/næ/	681.0	717.6	1893.6	1775.6	/ne/	656.8	650.8	1925.0	1842.5
/ræ/	706.6	696.5	1623.0	1656.2	/re/	642.9	662.0	1718.6	1709.1
/sæ/	679.6	671.5	1803.4	1676.4	/se/	615.5	649.5	1805.5	1704.7
/tæ/	715.0	725.9	1665.9	1603.2	/te/	694.3	696.8	1742.6	1693.4
/væ/	671.7	636.4	1726.0	1735.9	/ve/	647.7	639.0	1777.9	1638.7
/zæ/	650.8	664.2	1714.4	1716.6	/ze/	592.5	633.9	1812.4	1733.9

The values presented in table 7 show no systematic difference between the F1 and F2 frequencies for monosyllabic and disyllabic nonsense words. For some onsets the F1 average is higher for the monosyllabic word (as is the case for /gæ-/), and for others it is higher for the disyllabic word (as is the case for /se-/). For some, the F2 average is higher in the monosyllabic word (for example, /næ-/), yet in others it is higher in the disyllabic word (for example, /ge-/).

3.7 Phonological environment effects

To investigate what effect, if any, differing phonological environments might have on the quality of pre-lateral DRESS and TRAP, the F1 and F2 frequencies for /el/ and /æl/ were tabulated for each word-initial consonant used in the set of nonsense words. These F1 and F2 values were averaged across all tokens produced by all speakers for each particular initial phoneme. The averaged values are presented in table 8. The data presented in table 8 are sorted from lowest frequency to highest frequency. The corresponding preceding phoneme is entered to the left of the formant value. The appears to be no systematic effect caused by voicing, manner or place of articulation.

Table 8. Averaged F1 and F2 Frequencies for Minimal Pairs of Each Word-Initial Phoneme Ordered by Formant Frequency.

AVERAGE F1 (HZ)				AVERAGE F2 (HZ)			
/-æl/		/-el/		/-æl/		/-el/	
g	646.8	**dʒ**	591.7	**t**	1635.6	**r**	1713.9
d	647.7	**g**	601.1	**r**	1639.6	**t**	1724.7
dʒ	650.3	**z**	613.2	**l**	1656.5	**v**	1732.6
z	657.5	**d**	622.7	**f**	1673.7	**s**	1755.9
v	673.5	**s**	632.3	**v**	1703.3	**f**	1765.5
s	675.5	**l**	642.1	**z**	1715.5	**z**	1773.2
tʃ	685.9	**v**	642.7	**tʃ**	1717.0	**l**	1786.8
f	688.6	**tʃ**	644.9	**m**	1727.2	**tʃ**	1801.7
n	699.3	**f**	650.2	**k**	1735.0	**k**	1803.2
l	701.2	**r**	652.5	**s**	1739.9	**m**	1809.7
r	701.5	**n**	653.9	**d**	1766.4	**d**	1859.5
k	714.2	**k**	671.4	**dʒ**	1782.4	**dʒ**	1876.0
t	720.3	**m**	692.0	**n**	1834.6	**n**	1885.1
m	725.0	**t**	706.1	**g**	1856.7	**g**	1902.6

3.8 Social factors

The only social factor we set out to explicitly investigate was gender – half of the participants were male and half were female. However no significant differences between male and female participants were found.

We did not set out to investigate social class in this experiment, and the

participants were selected to be as similar to one another as possible – all are university students, or very recent graduates, aged between 18-30. Nonetheless, there were some differences in the production of the participants, as indicated by a range of Hotelling Lawley scores, which reflects a range of degree of distinction. In order to investigate possible effects of social class, we assigned each participant a score which was derived from the occupation of their parents. We used the Elley-Irving Socio-Economic Scale (Elley and Irving 1985) and investigated whether the combined parents' score related to the degree of distinction maintained in people's speech. The combined parents' Elley-Irving score was well correlated with the degree of distinction in production, both for real words ($r=-.51$, $p<.05$), and nonsense words ($r=-.57$, $p<.03$). If the participants are ranked in terms of their Elley-Irving scores – with the highest socioeconomic status at the top, then the four people who kept DRESS and TRAP significantly distinct in the real word production tasks fall within the top five.

This is consistent with the direction of Bayard's (1987) results, who found that working class participants were more likely to merge /el/ and /ael/ (70%) than middle or upper middle class participants (44% and 23% respectively.)

3.8 Summary

The production results show that there is no relationship between speakers' baseline productions and their pre-lateral productions of DRESS and TRAP, which suggests that the pre-lateral vowels behave completely separately from the baselines; most productions of pre-lateral DRESS and TRAP occupy an area of acoustic space lower and further back than baseline TRAP. There is a social class effect, with individuals from higher socio-economic backgrounds more likely to maintain a significant distinction. However, there were no significant effects of syllable structure or phonological environment.

4. Perception Tasks

The perception section of the experiment consisted of four tasks: two using real words and two using nonsense words. For all four tasks participants listened to a recording of the same female New Zealand English speaker who distinguished between /el/ and /æl/, and were required to identify the word or words they heard. Participants circled their answers on a multi-choice response sheet. The multiple-choice options for each question were presented in random order.

The set of real words used in the perception tasks was the same set as used in the production tasks, though with the exclusion of the near minimal pairs *alligator-elevator,* and *mallet-melon*, giving a total set of ten minimal pairs (20 words). The set of nonsense words used in the perception tasks was a reduced version of the set of nonsense words used in the production tasks. This was because preliminary interviews showed that, without reduction, the perception tasks took too long to complete, causing participants to become bored and lose concentration towards the end of the final task

4.1 Perception of Real Words

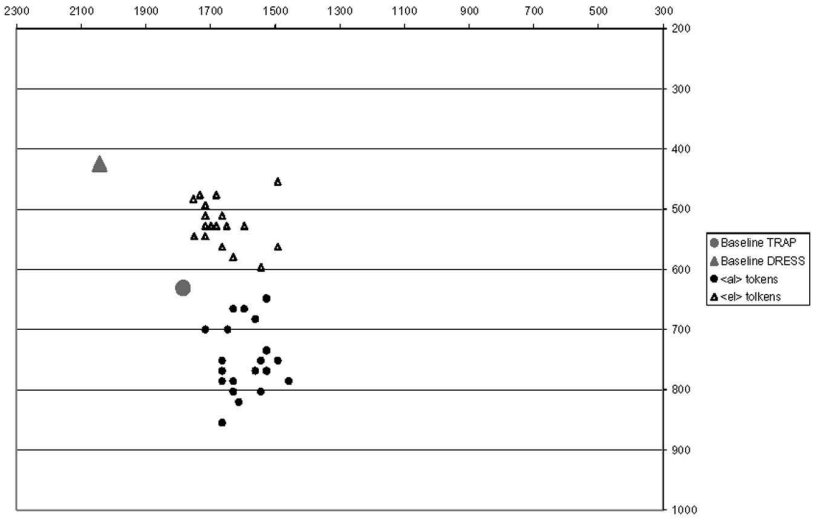
The first perception task was designed to test whether participants could correctly identify the target words in isolation. The individual words appeared in a random order. Participants were played each word once and asked to select the word they heard from the two options given on the response sheet. For example, participants were played the word *telly* and were asked to circle either *telly* or *tally* on the response sheet. Each word appeared only once, with a two second break between words. This task also included the words *head* and *had*, which acted as a control, ensuring that, though participants may not have been able to accurately discriminate between /el/ and /æl/, they were still able to accurately discriminate between /e/ and /æ/ in a neutral environment. Figure 5 shows the distribution of /el/ and /æl/ tokens for the set of real word stimuli. As far as possible, care was taken to select minimal pair tokens which were matched in every way except for the vowel (for example, matched in vowel length, intonation, plosive aspiration, *et cetera*).

The total number of perception errors for perception task one for all speakers is presented in table 9. There were twenty questions in this task.

The data in table 9 shows the range of perception results gained from the identification task involving real words in isolation. Accuracy rates range from only one error (female 4, male 2 and male 6), to nine errors (male 3), with the average number of perception errors for this task being four.

Task one required participants to identify the given stimulus from two options. As such, statistically, there was a 50% chance of getting each answer 'correct'. Therefore, an overall accuracy score of 50% for task one would be predicted by chance. All of the participants completed task one with accuracy scores greater than chance (ranging from 95% accuracy from Male and Male 2 to 55% accuracy from Male 3), though female 5, male 3 and male 4 came close to chance scores. It is also apparent in table 9 that most participants were less accurate at identifying /æl/ words than /el/ words.

Figure 5. F1 and F2 Frequencies for Target Variables of the Real Words.



A second task was designed to test whether participants could correctly identify the target words when presented with a minimal pair. The set of words used for this task was the same as in perception task one. Participants were played a minimal pair, the individual words of which were separated by a half second gap, and were asked to identify which words they heard and in what order. Each pair of words appeared four separate times throughout the course of the task, each time with the words in a different order. For example, the minimal pair *celery-salary* appeared four times: once as *salary-celery*, once as *celery-salary*, once as *celery-celery*, and once as *salary-salary*. On the multiple-choice response sheet participants were asked to select, from the four possible options, which pair they heard. The stimuli were randomised and counter-balanced. Importantly, in cases where the same word was repeated (for example, *celery-celery*) two different tokens of the word, produced by the same speaker, were used so that an accurate identification could not be put down to recognition of a single repeated token. In cases where the minimal pair was presented, the same token of each word was used for both orders of presentation. For example, the stimuli *celery-salary* and *salary-celery* used the

Table 9. Total Number of Perception Errors for Perception Task One.

	TOTAL ERRORS (OUT OF 20)	MISHEARD /æ/ (OUT OF 10)	MISHEARD /e/ (OUT OF 10)
FEMALE 1	4 (20%)	2 (20%)	2 (20%)
FEMALE 2	3 (15%)	3 (30%)	0
FEMALE 3	4 (20%)	2 (20%)	2 (20%)
FEMALE 4	1 (5%)	1 (10%)	0
FEMALE 5	8 (40%)	6 (60%)	2 (20%)
FEMALE 6	6 (30%)	6 (60%)	0
FEMALE 7	7 (35%)	5 (50%)	2 (20%)
FEMALE 8	6 (30%)	4 (40%)	2 (20%)
MALE 1	6 (30%)	2 (20%)	4 (40%)
MALE 2	1 (5%)	0	1 (10%)
MALE 3	9 (45%)	6 (60%)	3 (30%)
MALE 4	8 (40%)	7 (70%)	1 (10%)
MALE 5	3 (15%)	3 (30%)	0
MALE 6	1 (5%)	1 (10%)	0
MALE 7	2 (10%)	2 (20%)	0
MALE 8	6 (30%)	4 (40%)	2 (20%)
AVERAGE	4 (20%)	3 (30%)	1 (10%)

exact same tokens of each word. This ensured that if there should be any tendency to accurately perceive one order over the other, that that tendency would be a result of the order of the words, and not of differences between separate tokens of the same word.

The total number of inaccurate answers for this task for all speakers is provided in table 10. In table 10, an 'inaccurate answer' is considered to be the inaccurate perception of a pair, not of an individual word, so if a participant was presented with the stimulus *melody-malady*, but perceived it as *malady-melody*, this would have been counted as one 'inaccurate answer'.

The data in table 10 show a wide range of perception accuracy rates for task two, with male 7 only making two errors, and female 7 making 25 errors. There were 40 questions in total for task two, however, since each response was a (forced) four-option multi-choice answer, the accuracy rate that could be generated by chance is 25%, not 50% as in task one. With this in mind, it

Table 10. Total Number of Inaccurate Answers for Perception Task Two for All Speakers.

PARTICIPANT	TOTAL (OUT OF 40)	PARTICIPANT	TOTAL (OUT OF 40)
FEMALE 1	4 (10%)	**MALE 1**	8 (20%)
FEMALE 2	22 (55%)	**MALE 2**	2 (5%)
FEMALE 3	15 (38%)	**MALE 3**	17 (43%)
FEMALE 4	12 (30%)	**MALE 4**	19 (48%)
FEMALE 5	11 (28%)	**MALE 5**	7 (18%)
FEMALE 6	17 (43%)	**MALE 6**	3 (8%)
FEMALE 7	25 (63%)	**MALE 7**	2 (5%)
FEMALE 8	11 (28%)	**MALE 8**	13 (33%)

becomes clear that all participants completed task two with a level of perception accuracy greater than chance (30 inaccurate answers would be expected by 'chance').

4.2 Perception of Nonsense Words

A nonsense words task had the same design as perception task one, but used monosyllabic and disyllabic nonsense words. The aim of the task was to test whether participants could perceive an /el/-/æl/ distinction in isolated target words. The words were played in a random order with a two and a half second break between each. Each word was played only once. Participants were asked to select, from the two minimal options printed on the response sheet, which word they heard. Figures 6 and 7 show the distribution of the target variables in the reduced set of monosyllabic and disyllabic nonsense word stimuli respectively. As with the real word stimuli, care was taken to select minimal pair tokens which were matched in every way except for the vowel.

The total number of perception errors for this task for all speakers is presented in table 11. There were thirty-six questions in this task.

The data in table 11 show that all participants performed task three with a much higher rate of accuracy than that which would be predicted by chance (chance being 50% accuracy, or 18 total errors). There is a range of total error scores, from 4 (10% error rate) from male 6 and male 7, to 14 (35% error rate) from female 8.

Most participants made noticeably more perception errors when presented

Figure 6. F1 and F2 Frequencies for Target Variables of the Monosyllabic Nonsense Words.

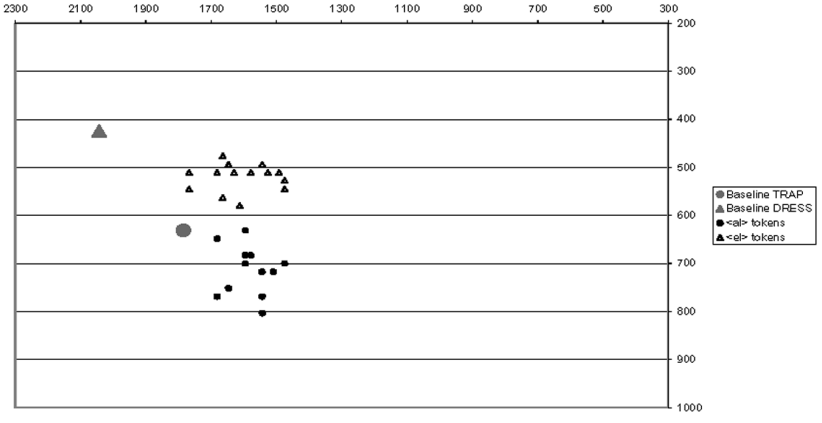


Figure 7. F1 and F2 Frequencies for Target Variables of the Disyllabic Nonsense Words.

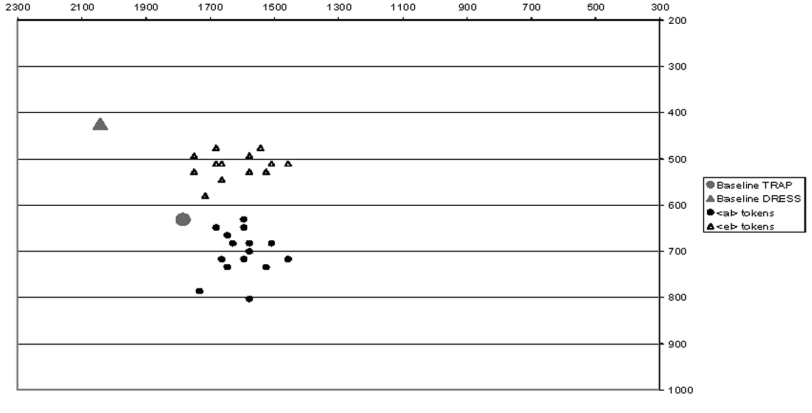


Table 11. Total Number of Perception Errors for Perception Task Three.

	TOTAL ERRORS (OUT OF 36)	MISHEARD /æ/ (OUT OF 18)	MISHEARD /e/ (OUT OF 18)
FEMALE 1	12 (33%)	10 (56%)	2 (11%)
FEMALE 2	10 (28%)	6 (33%)	4 (22%)
FEMALE 3	5 (14%)	3 (17%)	2 (11%)
FEMALE 4	9 (25%)	4 (22%)	5 (28%)
FEMALE 5	13 (36%)	11 (61%)	2 (11%)
FEMALE 6	11 (28%)	10 (56%)	1 (6%)
FEMALE 7	13 (36%)	8 (44%)	5 (28%)
FEMALE 8	14 (35%)	12 (67%)	2 (11%)
MALE 1	11 (28%)	6 (33%)	5 (28%)
MALE 2	6 (15%)	1 (6%)	5 (28%)
MALE 3	9 (25%)	9 (50%)	0
MALE 4	9 (25%)	9 (50%)	0
MALE 5	11 (28%)	8 (44%)	3 (17%)
MALE 6	4 (10%)	3 (17%)	1 (6%)
MALE 7	4 (10%)	2 (11%)	2 (11%)
MALE 8	12 (33%)	9 (50%)	3 (17%)
AVERAGE	10 (28%)	7 (39%)	3 (17%)

with /æl/ stimuli than when presented with /el/ stimuli. This result reflects that evident with the real words (table 9).

The participants also completed a task designed to establish whether participants could accurately perceive an /el/-/æl/ distinction between minimal pairs. This task used as its target words the same set of nonsense words. As with the real words, participants were played a recording of a speaker reading out a pair of words (either minimal or identical) and were required to select from four multi-choice options which words they heard and in what order.

Table 12 shows the total number of inaccurate answers for this task for all speakers. As was the case with table 10, an 'inaccurate answer' is considered to be the inaccurate perception of a pair, not of an individual word. Task four consisted of 56 questions.

As with perception task two, chance for task four would predict an accuracy score of 25% because each response came as a result of a (forced)

Table 12. Total Number of Inaccurate Answers for Perception Task Four for All Speakers.

PARTICIPANT	TOTAL (OUT OF 56)	PARTICIPANT	TOTAL (OUT OF 56)
FEMALE 1	16 (29%)	**MALE 1**	9 (16%)
FEMALE 2	24 (43%)	**MALE 2**	10 (18%)
FEMALE 3	15 (27%)	**MALE 3**	20 (36%)
FEMALE 4	8 (14%)	**MALE 4**	28 (50%)
FEMALE 5	12 (21%)	**MALE 5**	11 (20%)
FEMALE 6	26 (46%)	**MALE 6**	9 (16%)
FEMALE 7	7 (13%)	**MALE 7**	13 (23%)
FEMALE 8	19 (29%)	**MALE 8**	20 (36%)

four-choice selection. With this in mind, it becomes clear that all of the participants performed with greater than chance accuracy (chance predicting 42 total errors).

There is a wide range of results for task four, from only seven total errors (13% error rate) from female 8, to 28 total errors (50% error rate) from male 4.

4.3 Potential Effect of the Stimuli

In order to validate the results from the perception tasks it was important to ensure that there was no relationship between the error rates for the target words, and the degree of separation of the actual stimuli (as measured by Euclidean distance). In other words, a control test was conducted to ensure that participants' perception inaccuracies did not come as a result of the degree of production distinction in the stimuli for the perception tasks. In selecting the actual stimuli for the experiment, every endeavour was made to control for the degree of distinction, but, because the stimuli were taken from natural speech, some slight variation inevitably exists.

The Euclidean distance between each stimulus word pair was tested against the total number of perception inaccuracies (for the words-in-isolation tasks only) for that pair. There was no significant correlation, either for nonsense words or real words. In addition, the full dataset was tested for any correlation between total errors for words presented in isolation and the Euclidean distance of the actual stimuli pairs. There was found to be no

significant correlation between these two factors (Spearman's rho=0.274, p-value=0.16). This lack of correlation is reassuring, as it indicates that the small variation across the stimuli was not responsible for the response patterns observed in the perception tasks.

4.4 /el/ and /æl/ differences

The analysis of the production results showed that pre-lateral DRESS and TRAP were merging in an area of acoustic space lower and more central than baseline TRAP. A related analysis was undertaken on the perception results. The aim of this analysis was to discover which pre-lateral vowel (DRESS or TRAP), if either, was more accurately perceived. Table 13 shows the total number of /el/ perception errors and /æl/ perception errors for real words, for all participants across both the isolation and pairs tasks.

The data in table 13 show that most participants exhibited a higher rate of inaccuracy when presented with pre-lateral TRAP (with the exception of female 1, female 2, female 8, and male 2). Male 4, female 5, and female 6 shows a markedly higher rate of inaccuracy for pre-lateral TRAP tokens. This trend towards a higher error rate with TRAP is significant both by subjects (paired Wilcoxon, V=97.5, p-value<0.005) and by items (V= 51, p-value < 0.02).

The same pattern was found with the nonsense word data. Across subjects, the average error rate with TRAP was 26%. For DRESS words, the error rate was 14%. This is significant both by subjects (paired Wilcoxon, V=111, p-value<0.005) and by items (V= 14, p-value < 0.02).

Young New Zealanders are still exposed to distinct DRESS and TRAP before /l/, and these results suggest that they are able to exploit this distinction to some degree in speech perception. However their error rates in this task reflect their overall speech experience. Young New Zealanders produce DRESS forms in an acoustic space which is closest to /æ/. Thus, /æ/ forms to which they have been exposed are frequently ambiguous, whereas /e/ forms are not. The result that the error rate in perception is higher for TRAP forms suggests that this exposure affects speech perception, such that DRESS forms are less likely to be processed as ambiguous.

4.5 Syllable Structure

When we compare the overall error rate for the monosyllabic and disyllabic nonsense words, there is a small but significant difference in error rate. The average error rate for monosyllabic words is 25%, whereas the average error

Table 13. Total Number of /el/ and /æl/ Perception Errors for Real Words across All Participants.

	TOTAL ERRORS /æl/ (OUT OF 50)	/el/ (OUT OF 50)
FEMALE 1	3 (6%)	5 (10%)
FEMALE 2	17 (34%)	17 (34%)
FEMALE 3	16 (32%)	13 (26%)
FEMALE 4	10 (20%)	7 (14%)
FEMALE 5	17 (34%)	9 (18%)
FEMALE 6	21 (42%)	8 (16%)
FEMALE 7	24 (48%)	17 (34%)
FEMALE 8	8 (16%)	9 (18%)
MALE 1	12 (24%)	9 (18%)
MALE 2	8 (16%)	2 (4%)
MALE 3	23 (46%)	13 (26%)
MALE 4	25 (50%)	6 (12%)
MALE 5	7 (14%)	3 (6%)
MALE 6	4 (8%)	0
MALE 7	3 (6%)	8 (16%)
MALE 8	12 (24%)	8 (16%)
AVERAGE	13 (26%)	8 (16%)

rate for disyllabic words is 31%. The difference in error rates between one and two syllable words is significant ($W=81$, $p<.001$).

This result is interesting, in light of the fact that there was no significant acoustic difference in the degree of separation between the minimal pairs for monosyllabic versus disyllabic stimuli. There was also no syllable-based difference in the degree of distinction produced by these participants in their own speech. Yet they are significantly more accurate at hearing the distinction in one syllable than two syllable words.

One potential difference between the one syllable and two syllable nonsense words is the vowel length. The one syllable words have the crucial vowel appearing in the final syllable of the word, and so it may be subject to final lengthening (Wightman et al. 1992). Indeed, the vowels in the mono-

syllabic words were significantly longer than those in the disyllabic words (paired Wilcoxon, V=171, p<.001). The vowels in the monosyllabic words average 0.162ms, and those in the disyllabic words average 0.112ms. This slight difference in length may be the reason for the greater accuracy in perceiving the distinction in monosyllables. Participants are exposed to the vowel for a fraction of a second longer, and so have more evidence on which to base their analysis.

5. Conclusion

16 New Zealand English speakers completed a range of production and perception tasks involving nonsense and real words containing pre-lateral DRESS and TRAP. The experiments used both nonsense and real word stimuli.

In production, both pre-lateral DRESS and pre-lateral TRAP were produced in an area of acoustic space closer to baseline TRAP than baseline DRESS, and also noticeably centralised. This was true of both merged speakers and unmerged speakers. There is some evidence that there is a social class effect – with participants from higher socio-economic backgrounds more likely to maintain a distinction.

In perception there was a highly significant trend for /æl/ word stimuli to result in more errors than /el/ word stimuli. This reflects speakers' exposure to many speakers who produce both forms in an acoustic space closer to TRAP.

There was also a higher rate of perception accuracy when presented with a monosyllabic nonsense word than when presented with a disyllabic nonsense word. It was speculated that this was due to the difference in the vowel length for mono- and di-syllabic words, with the monosyllabic stimuli having a significantly longer vowel.

These results demonstrate that, while TRAP and DRESS are merged before /l/ in the speech of many young NZers, all are still largely able to exploit the distinction for the purposes of speech perception. It will be intriguing to track the degree to which this remains true in future generations. In what ways will the significantly reduced exposure to speakers who produce a reliable distinction between /el/ and /æl/ affect the processing of these sounds in speech perception?

Notes

1. We use Wells' (1982) lexical set terminology.

References

Bauer, Laurie. 1986. 'Notes on New Zealand English Phonetics and Phonology.' *English World-Wide* 7(2): 225–258.

Bauer, Laurie and Warren, Paul. 2004. New Zealand English phonology. In Bernd Kortmann, Edgar W. Schneider, Kate Burridge, Rajend Mesthrie and Clive Upton (eds) *A Handbook of Varieties of English: A Multimedia Reference Tool* (vol. 1). Berlin: Mouton de Gruyter. 580–602.

Bayard, Donn. 1987. 'Class and Change in New Zealand English: A Summary Report.' *Te Reo* 13: 3–36

Buchanan, Hannah. 2001. 'Neutralisation of DRESS and TRAP before /l/ in New Zealand English.' *Wellington Working Papers in Linguistics* 13: 15–29.

Elley, Warwick B. and James C. Irving. 1985. 'The Elley-Irving Socio-Economic Index: 1981 census revision.' *New Zealand Journal of Educational Studies* 20: 115–128.

Gordon, Elizabeth and Tony Deverson. 1998. *New Zealand English and English in New Zealand*. Malaysia: New House Publishers Ltd.

Horsfield, Rachel. 2001. 'The Changing Vowels of New Zealand English.' Honours Thesis. University of Otago.

Maclagan, Margaret. 1982. 'An Acoustic Study of New Zealand English Vowels.' *The New Zealand Speech Therapists' Journal* 37: 20–26.

Thomas, Brynmor. 2003. 'A Study of the /el/-/æl/ Merger in New Zealand English.' *New Zealand English Journal* 17: 28–44.

Thomas, Brynmor. 2004. 'In support of an exemplar-based approach to speech perception and production: a case study on the merging of pre-lateral DRESS and TRAP in New Zealand English.' Unpublished Masters thesis. Christchurch, NZ: University of Canterbury.

Wells, John. 1982. *Accents of English*. (3 Vols). Cambridge: Cambridge University Press.

Wightman, Colin W., Shattuck-Hufnagel, Stefanie, Ostendorf, Mari, and Price, Patti J. 1992. 'Segmental durations in the vicinity of prosodic phrase boundaries.' *Journal of the Acoustical Society of America* 92: 1707–17.