# NON-LINEAR ANALYSIS OF A DIPHTHONG MERGER

Paul Warren
Victoria University of Wellington, New Zealand
paul.warren@vuw.ac.nz

## ABSTRACT

This paper presents non-linear analysis, using Generalised Additive Mixed Models (GAMMs), of formant trajectory data relating to the well-attested merger of the centring diphthongs NEAR and SQUARE (/iə/ and /eə/) in New Zealand English (NZE). Previous acoustic analyses have typically focused solely on formant values for the first target of the diphthong, usually measured at the point of peak F2, and have reported apparent-time reductions in Euclidean distance between these diphthongs or increases in the overlap of their distributions in vowel space. The current GAMM analysis includes the entire diphthong trajectory for F1 and F2. Prior results on the degree of separation of the first target are replicated and extended to a larger portion of the vowels. In addition, the analysis reveals both sex differences and apparent-time changes in later parts of the diphthongs, as well as in the timing of the initial F2 peak.

**Keywords**: sound merger, centring diphthongs, New Zealand English, non-linear analysis

## 1. INTRODUCTION

One of the long-noted features of the pronunciation of New Zealand English (NZE) is the merger of the NEAR and SQUARE vowels. For many NZE speakers, word pairs like *cheer* and *chair* or *fear* and *fair* are homophones. The merger, once a topic in the complaint tradition of letters to the editor and opinion columns, is now seldom referred to in such public discourse. While this could be taken as an indication that the merger is now well established, recent research [7] has suggested that such a conclusion may be premature. The primary purpose of the current study is not, however, to examine whether this merger is complete, still progressing, or possibly reversing, but to apply novel methods in the analysis of NEAR and SQUARE vowels from corpus data. In particular, this study broadens the scope of the comparison of the vowels from a focus on the initial target of the diphthongs to a consideration of the entire trajectory, using non-linear analysis through generalised additive mixed effects modelling.

Most early studies of the merger were auditory in nature. Some of these early studies suggested that the direction of the merger might be towards the more open, SQUARE vowel [3, 8, 12]. However, a subsequent more extensive auditory analysis using data collected over a longer time-frame concluded that the merger was proceeding towards NEAR, with the starting point of the SQUARE diphthong becoming less open (higher) over time [4, 10].

The focus in this earlier auditory research on the first element of the diphthongs is not surprising. The first element is closely linked to a possible phonetic motivation for the merger [14], and at the risk of circularity we note that transcriptions that distinguish the two vowels do so primarily in the first element. The second element is typically argued to be identical, usually schwa. At least one study [2] acknowledges the possibility of alternative realisations of the second element (so that the diphthongs are realised as [Vː], [Vᵊ] or [Və]), but these variants are not claimed to differentiate NEAR and SQUARE. Other analyses also allow for monophthongal realisations, although these again do not reliably distinguish the two diphthongs [4, 8].

Acoustic analyses of the diphthongs typically measure F1 and F2 at the point of highest F2 in the initial portion of the vowel, which is assumed to correspond to the auditory first target. To gauge the merger, these formant values are then subjected to Euclidean distance measures or to calculations of distributional overlap (e.g., Hotelling-Lawley trace or Pillai score) [6, 7, 16]. However, by definition, diphthongs are dynamic, and focusing on a single point ignores other potential differences, such as in the trajectory shape and/or the end target of the diphthong. For Australian English, it has been found that classification of diphthongs in fact benefits from the inclusion of multiple spectral slices because diphthongs 'have at least two targets' [5]. Additionally, formant movements in diphthongs are not restricted to straight-line trajectories from one formant value to another, but frequently involve non-linearities.

Recent publications on the dynamic analysis of speech phenomena have introduced statistical tools to the phonetics community for the modelling of dynamic and in particular non-linear trends. These include articulographic analysis of Dutch dialects [19] and of English L1 vs. Dutch L2 speakers [18], as well as tutorial introductions to the use of Generalised Additive Mixed Models (GAMMs) in the analysis of linguistic change [21], of changes in /r/ articulation in Glaswegian English [11], and of pitch contours across

English compounds [1]. These statistical approaches are applied in this paper to the analysis of the NEAR and SQUARE diphthongs of NZE.

## 2. METHOD

### 2.1. Materials

The materials for the analysis reported here are taken from the New Zealand Spoken English Database (NZSED, see [13]). They consist of 18 words with NEAR and 18 words with SQUARE, from the sentence reading section of the database. The two sets are balanced as far as possible for the phonetic context preceding and following the diphthongs.

These materials were produced by 71 speakers (36 female and 35 male) in the age range 17-64. (The database consists of three age groups: younger, mid-age, and older, but analysis of the actual ages shows good distribution across each age range, and so for the current purposes speaker age is treated as a continuous variable.) The speakers came from two urban centres in New Zealand's North Island, Wellington and Hamilton. All were native speakers of NZE. Of 2,556 possible vowel tokens, three were missing (due to mispronunciation).

Segment labelling and alignment at the phone level was carried out using the NZE phoneme models in the *MAUS* on-line tool [9], with hand-correction where required. Formant tracks were produced by *forest* in the *wrassp* package in R [20], again with hand-correction where necessary.

Formants were normalised using the speaker-intrinsic procedure proposed in [17], and based on formant data from eight instances of each of a set of monophthongal vowels produced in sentence contexts by each speaker (see [15]). The diphthongs were time-normalised to produce 25 equally-spaced values across each vowel (but note that the actual vowel durations were included as random terms in the statistical modelling).

### 2.2. Statistical modelling

Statistical modelling was carried out using the *bam* function in the *mgcv* package in R [22]. The dependent variable consisted of the normalised formant values (for F1 and F2), giving a total of 127,650 values (2 formants x 25 time points x 2,553 tokens). Parametric factors included Formant (F1 vs. F2), the inclusion of which allowed the two vowel dimensions (close-open and front-back) to be modelled simultaneously. Vowel (NEAR or SQUARE) and speaker Sex were also included, together with interactions with Formant.

Non-linear smooths included in the model were Trajectory (25 time points for each vowel) and Age, both in interaction with Vowel and Sex. The non-linear interaction of Trajectory and Age was also included, again in interaction with Vowel and Sex.

The random effects structure catered for by-speaker variation in trajectories and in vowel duration across vowels and formants, and for by-item variation in trajectories and age across formants. It also allowed for the influence of preceding and following phonetic contexts, across vowels and formants.

Statistical models were constructed following the procedure presented in [19], using ordered factors for each Formant for NEAR and SQUARE, for female and male speakers, and for the interaction of Vowel and speaker Sex. These ordered factors allowed assessment of the effects on Formant values of Vowel and Sex and their interaction, with reference levels set to SQUARE and to male speakers.

Since in relatively smooth and slow-changing time series values at time t will be correlated with those at time t-1, an autocorrelation parameter was included in the models. This was calculated over an initial simpler model and set at 0.747. Also on the basis of the simpler model, datapoints with residuals of more than 2.5 (less than 1.4% of the data) were excluded from the analysis. Models were evaluated for best fit, and the final model explained 86.4% of the variance, with an estimated $R^2$ of 0.952.

## 3. RESULTS

### 3.1. Random effects

All of the random effects were highly significant, indicating that there was indeed variation by participants and items across the vowels and formants.

### 3.2. Parametric effects

The parametric effects are shown in Table 1. Unsurprisingly, given that the diphthongs are located in the top-right quadrant of the vowel space, normalised F2 was greater than normalised F1.

**Table 1**: Parametric coefficients

| | Est. | SE | *t* | *p* |
|---|---|---|---|---|
| Intercept (F1, male, SQUARE) | 0.92 | 0.032 | 28.80 | <0.001 |
| F2 vs. F1 | 0.53 | 0.062 | 8.61 | <0.001 |
| F1 by Vowel | -0.05 | 0.052 | -0.87 | 0.39 |
| F1 by Sex | -0.09 | 0.022 | -4.06 | <0.001 |
| F1 Vowel effect by Sex | 0.04 | 0.030 | 1.36 | 0.17 |
| F2 by Vowel | 0.01 | 0.078 | 0.16 | 0.88 |
| F2 by Sex | 0.02 | 0.011 | 1.45 | 0.15 |
| F2 Vowel effect by Sex | -0.01 | 0.015 | -0.78 | 0.43 |

Normalised F1 for females was lower than that for males. Note that this comparison is at the reference

level for Vowel, and so indicates that females have an overall raised (i.e. more NEAR-like) SQUARE vowel.

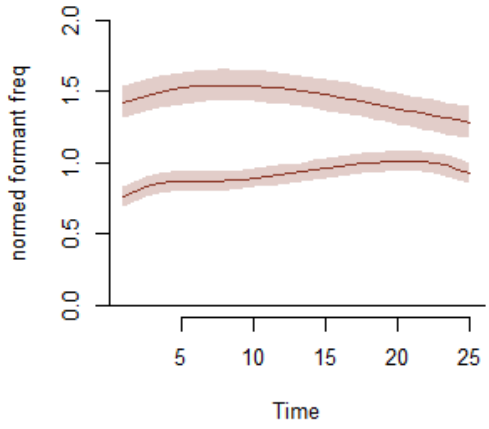
### 3.3. Non-linear smooths

The non-linear effects are given in Table 2. None of the smooth effects for Age were significant and these have been omitted from the table. The smooths over time [s(Time)] show significant non-linear trajectories for both formants (Figure 1), and that both of these differ by Sex.

**Table 2**: Smooth functions and tensor interactions

| | edf | *F* | *p* |
|---|---|---|---|
| s(Time): | | | |
| F1 (for male, SQUARE) | 8.4 | 73.5 | <0.001 |
| F1 by Sex | 7.0 | 4.4 | <0.001 |
| F1 by Vowel | 1.0 | 1.6 | 0.20 |
| F1 Vowel effect by Sex | 3.5 | 2.6 | <0.05 |
| F2 (for male, SQUARE) | 6.6 | 13.0 | <0.001 |
| F2 by Sex | 7.2 | 21.1 | <0.001 |
| F2 by Vowel | 3.4 | 0.8 | 0.46 |
| F2 Vowel effect by Sex | 2.6 | 1.2 | 0.34 |
| ti(Time, Age): | | | |
| F1 (for male, SQUARE) | 5.5 | 4.9 | <0.001 |
| F1 by Sex | 14.2 | 1.4 | 0.15 |
| F1 by Vowel | 1.0 | 0.3 | 0.52 |
| F1 Vowel effect by Sex | 1.0 | 0.5 | 0.56 |
| F2 (for male, SQUARE) | 22.7 | 2.7 | <0.001 |
| F2 by Vowel | 5.1 | 4.6 | <0.001 |
| F2 by Sex | 25.6 | 2.2 | <0.001 |
| F2 Vowel effect by Sex | 1.1 | 0.4 | 0.64 |

**Figure 1**: F1 and F2 trajectories at reference (male, SQUARE).




F1 is consistently lower for females, but the difference is larger at the beginning and end of the diphthong. The Sex effect for F2 is that females have significantly higher F2 than males over the first half of the diphthong, but significantly lower F2 later in the vowel. This indicates a greater front-to-back movement for the females, for the reference vowel SQUARE. The Vowel effect by Sex for F1 reflects a greater difference between NEAR and SQUARE F1 values for males, which reduces towards the end of the trajectories. These effects are consistent with previous findings that females are more likely to merge the vowels (towards NEAR) than males [4, 8].

### 3.4. Non-linear interactions

The bottom half of Table 2 shows non-linear interactions involving trajectories and Age [ti(Time, Age)]. For F1, trajectories start from a lower value (closer starting point) for younger speakers. This is visible at the left edge of the contour plot in Figure 2. The plot is for the reference values of males and SQUARE, and so shows how the starting point of this vowel becomes more NEAR-like over apparent time.

**Figure 2**: F1 trajectories by Age at reference (male, SQUARE).

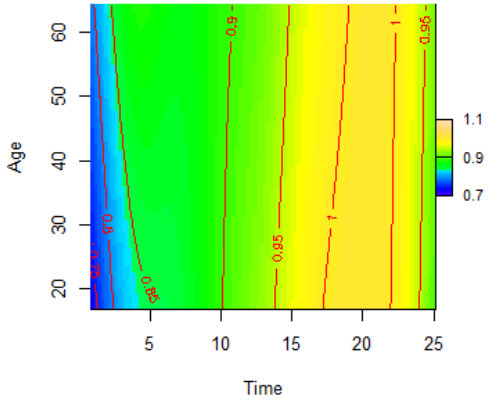
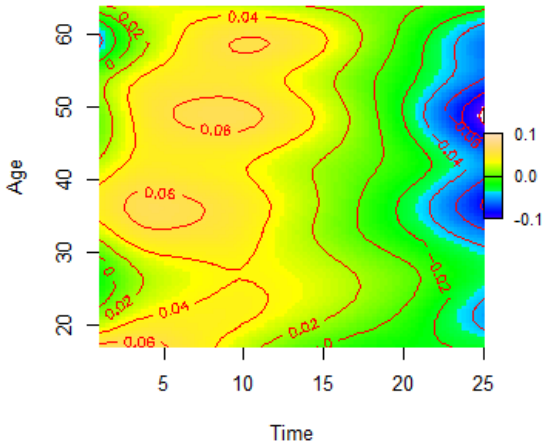



Table 2 also shows that F2 trajectories change over apparent time (Age) in interaction both with Vowel and with Sex. The latter indicates that the difference in the non-linear smooths for female and male F2 trajectories reported earlier is influenced by Age. The peak of the difference gets earlier with decreasing age, as shown in Figure 3.

**Figure 3**: Sex differences (female-male) in F2 trajectories by Age.

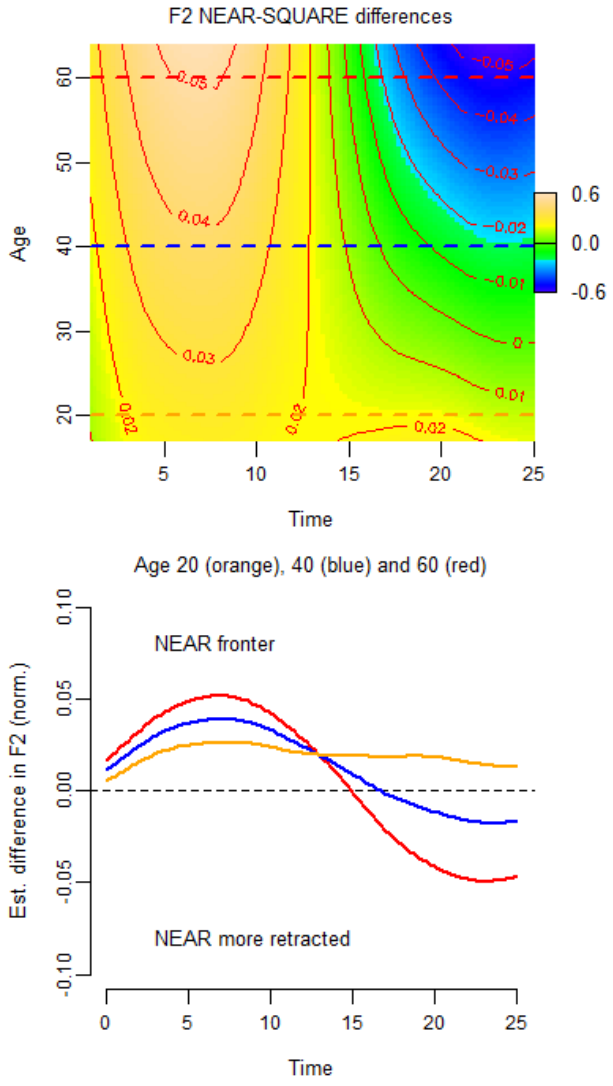



Note again that this is adjusted for the reference vowel SQUARE, suggesting that women not only have

more NEAR-like trajectories for SQUARE (with greater initial fronting) but also that they achieve this NEAR-like target earlier in the vowel the younger they are, i.e. the greater front-to-back movement noted earlier is more evident in the speech of the young females. Additional visualisations of male and female by-Age changes in F2 trajectories confirm that the effect in Figure 3 is carried by the female speakers, with the males showing little change over apparent time.

The Vowel (NEAR-SQUARE) difference in F2 trajectories over Age is shown in Figure 4. The contour plot at the top shows the Vowel differences, and the lower plot shows smooths extracted from the modelled data at three ages (20, 40, 60, corresponding to the dotted lines overlaid on the contour plot). These plots show clearly that the differences between F2 trajectories for NEAR and SQUARE become reduced over apparent time. Importantly, they also show that these changes affect not only the first part of the diphthongs, typically associated with the NEAR-SQUARE contrast, but also the final portion, reflecting the increasing diphthongal nature of SQUARE.

**Figure 4**: F2 trajectory: Vowel differences by Age.




## 4. DISCUSSION

The results of the non-linear GAMM analysis both confirm and extend the findings of previous acoustic studies of the merger of the NEAR and SQUARE diphthongs in NZE. These previous studies have focused on the initial target for the diphthongs, typically represented by single-point F1 and F2 values at the point of maximum F2 early in the vowel. They have generally (at least for the time frame and geographic regions represented in the current data) shown increased merger of the diphthongs over apparent time, based on those values. Most usually, this is reflected in SQUARE becoming more like NEAR, i.e. with lower F1 and higher F2. Also, a typical claim is that the change has been led by women.

In the current analysis, the overall F1 trajectory for SQUARE (i.e. not just the part associated with the initial target) is more NEAR-like for females, resulting in a smaller difference in NEAR-SQUARE trajectories for females than for males. However, while SQUARE is more open throughout its trajectory for males than for females, this difference reduces over apparent time. As far as the early part of the trajectory is concerned, younger speakers show closer SQUARE articulations.

The trajectories for F2 show that females have higher values than males over the first part of the SQUARE reference vowel, but importantly also reveal lower values over the last part of the vowel. This suggests greater front-to-back movement during SQUARE for females. Over apparent time, the differences between NEAR and SQUARE F2 trajectories become reduced, and interestingly this also affects the second part of the vowel as well as the first, i.e. the apparent-time changes in NEAR-SQUARE differences are not limited to the initial target typically investigated in previous acoustic studies. A further new finding was that female speakers achieve an earlier F2 peak for SQUARE the younger they are.

In sum, the formant trajectories of the NZE NEAR and SQUARE diphthongs are non-linear and therefore suited to modelling using a non-linear approach. This approach reveals patterns of similarity and difference between the diphthongs across the trajectories, confirming previous findings relating to the initial target (higher and fronter for NEAR, with SQUARE becoming more similar to NEAR over apparent time), and additionally showing marked differences in the final portion of the vowels, which also change over apparent time, as well as differences in the alignment in the SQUARE vowel of the F2 peak.

## 5. REFERENCES


[1] Baayen, H., et al. 2018. Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In: Speelman, D., Heylen, K., Geeraerts, D. (eds), *Mixed Effects Regression Models in Linguistics*. Berlin: Springer, 49-69.

[2] Batterham, M. 2000. The apparent merger of the front centring diphthongs - EAR and AIR - in New Zealand English. In: Bell, A., Kuiper, K. (eds), *New Zealand English*. Wellington: Victoria University Press, 111-145.

[3] Durkin, M. 1972. *A study of the pronunciation, oral grammar and vocabulary of West Coast schoolchildren*, MA, University of Canterbury.

[4] Gordon, E., 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), 215–238.

[5] Harrington, J., Cassidy, S. 1994. Dynamic and target theories of vowel classification: Evidence from monophthongs and diphthongs in Australian English. *Language and Speech* 37(4), 357-373.

[6] Hay, J., Drager, K., Warren, P. 2009. Careful who you talk to: An effect of experimenter identity on the production of the NEAR/SQUARE merger in New Zealand English. *Australian Journal of Linguistics* 29(2), 269-285.

[7] Hazenberg, E. 2017. *Liminality as a lens on social meaning: A cross-variable analysis of gender in New Zealand English*, PhD, Victoria University of Wellington.

[8] 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(3), 251-273.

[9] Kisler, T., Reichel, U.D., Schiel, F. 2017. Multilingual processing of speech via web services. *Computer Speech & Language* 45, 326-347.

[10] Maclagan, M.A., Gordon, E. 1996. Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change* 8(1), 125-147.

[11] Sóskuthy, M. 2017. Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction. *arXiv preprint arXiv:1703.05339*.

[12] Starks, D., Allan, S., Kitto, C. 1998. *Why vernacular speech? Speech samples from the taped Auckland rapid and anonymous survey*. Sixth New Zealand Language and Society Conference, Wellington, NZ.

[13] Warren, P. 2002. NZSED: building and using a speech database for New Zealand English. *New Zealand English Journal* 16, 53–58.

[14] Warren, P. 2006. Word recognition and sound merger. In: Luchjenbroers, J. (ed), *Cognitive Linguistic investigations across languages, fields, and philosophical boundaries*. Amsterdam: John Benjamins, 169–186.

[15] Warren, P. 2017. Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association*, 1-26.

[16] Warren, P., Hay, J., Thomas, B. 2007. The loci of sound change effects in recognition and perception. In: Cole, J., Hualde, J.I. (eds), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 87–112.

[17] Watt, D., Fabricius, A. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1-F2 plane. *Leeds Working Papers in Linguistics and Phonetics* 9, 159–73.

[18] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86-116.

[19] Wieling, M., et al. 2016. Investigating dialectal differences using articulography. *Journal of Phonetics* 59, 122-143.

[20] Winkelmann, R., Bombien, L., Scheffers, M. 2018. *Interface to the 'ASSP' Library*, vsn 0.1.8. https://github.com/IPS-LMU/wrassp.

[21] Winter, B., Wieling, M. 2016. How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution* 1(1), 7-18.

[22] Wood, S. 2018. *mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*, vsn 1.8-24. https://cran.r-project.org/web/packages/mgcv/.