INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Tracking the New Zealand English NEAR/SQUARE merger using functional principal components analysis

Michele Gubian[1], *Jonathan Harrington*[1], *Mary Stevens*[1], *Florian Schiel*[1], *Paul Warren*[2]

[1]Institute of Phonetics and Speech Processing, LMU Munich, Germany
[2]School of Linguistics and Applied Language Studies, Victoria Univ. of Wellington, New Zealand
m.gubian,jmh,mes,schiel@phonetik.uni-muenchen.de, paul.warren@vuw.ac.nz

Abstract

The focus of the study is the application of functional principal components analysis (FPCA) to a sound change in progress in which the SQUARE and NEAR falling diphthongs are merging in New Zealand English. FPCA approximated the trajectory shapes of the first two formant frequencies (F1/F2) in a large acoustic database of read New Zealand English speech spanning three different age groups and two regions. The derived FPCA parameters showed a greater degree of centralisation and monophthongisation in SQUARE than in NEAR. Compatibly with the evidence of an ongoing sound change in which SQUARE is shifting towards NEAR, these shape differences were more marked for older than for younger/mid-age speakers. There was no effect of region nor of place of articulation of the preceding consonant; there was a trend for the merger to be more advanced in low frequency words. The study underlines the benefits of FPCA for quantifying the many types of sound changes involving subtle shifts in speech dynamics. In particular, multi-dimensional trajectory shape differences can be quantified without the need for vowel targets nor for determining the influence of the parameters – in this case of the first two formant frequencies – independently of each other.

Index Terms: sound change, functional data analysis, New Zealand English

1. Introduction

In the last few decades, there has been an ongoing merger of the New Zealand English (NZE) falling diphthongs /eə, ɪə/ (lexical sets SQUARE and NEAR). In an investigation of 14-15 year old students recorded at five yearly intervals [1, 2, 3], the proportion of participants showing the merger increased from 16% to 80% between 1983 and 1999. The merger is one of approximation [3] in which /eə/ has shifted into the /ɪə/ space [4, 5]. Some [6] have argued that the raising of /eə/ to /ɪə/ is causally linked to DRESS-raising (lexical set DRESS) towards /i/ (FLEECE) that has formed part of a clockwise rotation of the New Zealand English short front vowels in the last 50 years or so [7, 8]. With the exception of [9], showing a greater separation between /eə, ɪə/ in Christchurch than in Wellington, there is little evidence that the merger is regionally conditioned.

There have been several perception experiments [4, 9, 10] in the last 10-15 years probing how the merger interacts with lexical access. These (e.g. [9]) generally show that the merger is less advanced in perception than in production (i.e. that speakers who scarcely produce an /eə, ɪə/ difference nevertheless hear the difference between the diphthongs). These studies also show an asymmetry that is consistent with a shift of /eə/ towards /ɪə/: in general, and as also assessed by semantic priming experiments [4], the perception of NEAR activates both NEAR and SQUARE words whereas the perception of SQUARE words activates only SQUARE words.

Maclagen & Gordon [2] suggest a role for lexical diffusion in which the /eə, ɪə/ merger takes place in some words before others. Todd et al. [11] have developed a computational model in which lexical frequency interacts with sound change in different ways. Central to the model is the idea that high frequency words are recognised in noise or in regions of ambiguity more robustly than are low frequency words; and that listeners are only likely to store exemplars of words if their phonemes are sufficiently discriminable and typical of their phoneme classes. Consequently, when a region of ambiguity is created as a result of a sound change in progress in which A and B are two phonemes and A encroaches on the space of B, then A's high frequency words are predicted to lead the sound change because they are more robustly recognised (and hence are more discriminable) in the region in which A and B progressively overlap due to the A → B sound change in progress. Following this reasoning, high frequency /eə/-words should change faster than low frequency /eə/-words if this is a merger in which /eə/ increasingly encroaches on the /ɪə/ space. On the other hand, there is some evidence [9] that the /eə, ɪə/ merger is not necessarily conditioned by lexical frequency but is instead more advanced when there is a preceding coronal consonant whose high F2-locus causes an F2-raising in /eə/ and hence a synchronic shift towards /ɪə/.

Perhaps surprisingly, the majority of evidence for the NZE merger in the studies sketched above is based on auditory analysis and perception experiments. By contrast, studies in which the main focus is an acoustic analysis of production [8, 12, 13, 14] are much rarer. Moreover, such acoustic analyses are typically based on an analysis of a single time point located somewhere in the first third of the diphthong [5, 9]. A problem with this approach is that this initial target may be difficult to identify, especially since any closer approximation of /eə/ towards /ɪə/ due to tongue fronting could be confounded with F2-locus cues to consonant place of articulation over a similar temporal extent. Such analyses also discount the possibility that the entire shape of the trajectory - including the degree to which the falling diphthongs are monophthongised [2] - could be a factor in separating vowels in NEAR/SQUARE words.

In the following study, we seek to overcome some of these difficulties by analysing the merger in progress based on the dynamic shape differences in the first two formants between /eə, ɪə/ from a large acoustic database of read NZE speech containing speakers from three different age groups and two regions. The shapes of formants for the purposes of vowel classification can be effectively encoded with the discrete cosine transformation [15] as well as with statistical techniques based on generalised additive mixed modeling [16]. However, the first of these techniques is typically applied separately to dy-

http://dx.doi.org/10.21437/Interspeech.2019-2115

namically changing formants, as if they were completely independent of each other. By contrast, a composite analysis of a multidimensional dynamic trajectory that takes into account the changing shapes of both formants together is possible using the technique of the present study, functional principal components analysis (FPCA) [17], [18]. This type of analysis based on a changing multi-dimensional trajectory seems to be especially appropriate in this case, given that the evolving approximation of /eə, ɪə/ could be based on gradual shifts in tongue height and fronting and hence on dynamic changes to both F1 and F2.

Based on the brief literature survey above, we hypothesise that older and younger speakers are likely to differ in /eə/ but not necessarily in /ɪə/, if the sound change is an /eə/ → /ɪə/ approximation [1, 4, 5]. Although region and sex will also be tested in this study, there is only marginal evidence from the literature reviewed above [3] that they will interact with the sound change in progress. We did however test for effects of phrase-finality (essentially whether or not the words containing the falling diphthongs preceded a pause) given that monophthongisation (which, as noted above might have an effect on the /eə/ → /ɪə/ sound change [2]) is less likely to occur in phrase-final position. We also tested changes in /eə/ for two types of lexical effects. The first was whether the /eə/ shift was more advanced in high than low frequency words (on the assumption that this is an /eə/ → /ɪə/ approximation of the same kind by which one phoneme increasingly encroaches on the space of another as in [11]). The second was whether the age differences in /eə/ were least marked following coronals, consistently with the idea that the merger has taken place in this context ahead of others [19].

2. Effects of age group

2.1. Method

2.1.1. Materials, speakers, parameters

The materials were taken from the New Zealand Spoken English Database that is stored and managed by the host institute of the fifth author. The recordings had been made between 1999-2000 and included isolated /hVd/ words as well as 200 read sentences drawn from a mixture of TIMIT [20] and ANDOSL [21] sentences. The isolated words and sentences were read once by each speaker, with speakers drawn from 3 age groups (younger: 18-30 years; mid-age: 31-45 years; older: 46-60 years) and 2 regions (Hamilton and Wellington, both on the north island and separated by just over 500 km). The breakdown of the speakers was 3 (age groups) × 2 (regions) × 12 (6M; 6F) = 72 speakers plus one additional mid-age, female Hamilton speaker. Within the 200 sentences, each word type containing /eə, ɪə/ typically occurred once and was repeated no more than twice. The analysed falling diphthongs were in the syllable with primary lexical stress, except for the words *questionnaire, unfair* and *thoroughfare*, where /eə/ is in a syllable with secondary lexical stress. We removed from consideration function words and exclamations (e.g. *their, there, we're, yeah*), words that had been variably produced by the speakers (*garish, Clara, Sarah*) with an open monophthong or /eə/, as well as /ɪə/ when it did not occur before underlying /r/ (e.g. *ideally*). This left 30 /eə/ (e.g. *airing, aware*) and 23 /ɪə/ (e.g. *appeared, beard*) word types in the 200 sentences resulting in 2328 /eə/ + 1894 /ɪə/ = 4222 analysed falling diphthongs across all speakers.

The speech materials were forced aligned using the Munich automatic segmentation system MAUS [22]. The resulting database was structured and analysed within `emu-webapp` and `emuR` [23]. The first five formant frequencies were derived from the Praat [24] formant tracker set within the range 0-5 kHz for male speakers and 0-5.5 kHz for female speakers and with a Gaussian window of 25 ms and a frame shift of 6.25 ms. Segment boundaries and formant frequencies of the target /eə, ɪə/ diphthongs were manually corrected. Most of the segment boundary corrections occurred at the ends of utterances in which the right boundary extended into silence. Formant errors were most common when F3 had been mis-tracked as F2.

The formant frequencies were speaker-normalised using z-score normalisation [25] with respect to the same speaker's /i, ɔ, æ/ vowels (lexical sets: FLEECE, THOUGHT, START). The speaker-normalised formant data were linearly time-normalised to 11 equally spaced time points, and then a 5-point median filter was applied to every 11-point formant contour in order to eliminate short-term random deviations. Finally, each 11-point curve was interpolated using a B-splines basis following the procedure illustrated in [18]; this last step being necessary in order to feed formant tracks as input to FPCA.

2.2. Functional PCA

FPCA [17] provides a data-driven parametrisation of a set of input curves, the latter represented by continuous functions defined on the same time interval. Here each curve is actually a *pair* of formant tracks $F1(t), F2(t)$ defined on normalised time $0 \leq t \leq 1$. The FPCA parametrisation is expressed by the following pair of equations:

$$F1(t) \approx \mu_{F1}(t) + \sum_{k=1}^{K} s_k \cdot PCk_{F1}(t) \quad (1a)$$

$$F2(t) \approx \mu_{F2}(t) + \sum_{k=1}^{K} s_k \cdot PCk_{F2}(t) \quad (1b)$$

where $\mu_{F1}(t)$ and $\mu_{F2}(t)$ are the mean formant tracks, $PCk_{F1}(t)$ and $PCk_{F2}(t)$ are K pairs of Principal Components (PCs), $k = 1, \ldots, K$, which are based on the entire formant track data set, and s_k are weights or *scores*, which modulate PCs differently for each formant track pair. Formally, Eq. (1) follow the same structure of ordinary PCA, namely any input curve $F(t)$ is approximately decomposed into a linear combination of K PCs added to the data set mean $\mu(t)$. What is different from ordinary PCA is that input, mean and PCs are functions of time as opposed to vectors of real numbers, and in this particular case these functions are multi-dimensional, as they take both F1 and F2 values at each point t in time. Crucially, the linear combination expressed by PC scores modulates the PCs for both F1 and F2 dimensions together, i.e. $s_1, s_2, \ldots, s_K$ are the same in Eq. (1a) and (1b). We computed the first $K = 3$ PCs , which combined explain 90.8% of the formant track variance. Figure 1 shows the effect of each PC. Solid curves are the mean formants $\mu_{F1}(t)$ and $\mu_{F2}(t)$, and are the same in all panels. The $\pm$ curves illustrate the modification of the F1 and F2 shapes by each PC in Eq. (1). In each PCk panel, $\pm$ curves are computed by setting in Eq. (1) the s_k score alone to positive and negative values respectively and the other scores to zero. We chose these representative scores to be $\pm\sigma_{s_k}$, i.e. we add to or subtract from each mean curve only one PC curve multiplied by the standard deviation of its corresponding score (0.37, 0.25 and 0.17, respectively). For example, in the PC1 panel on the left, the lower + curve shows the F1 track computed as $\mu_{F1}(t) + \sigma_{s_1} \cdot PC1_{F1}(t)$, and the upper + curve shows its counterpart for F2, i.e. $\mu_{F2}(t) + \sigma_{s_1} \cdot PC1_{F2}(t)$.

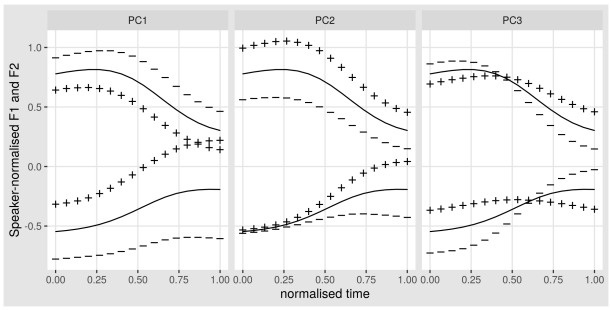

Figure 1: *First three PCs represented as perturbation of the mean. Solid curves are $\mu_{F1}(t)$ (bottom) and $\mu_{F2}(t)$ (top) as in Eq.* (1)*, the same pair of curves in all panels; in panel PCk, $k = 1, 2, 3$, the $\pm$ curves are $\mu_{F1/2}(t) \pm \sigma_{s_k} \cdot PCk_{F1/2}(t)$, where σ_{s_k} is the st. dev. of s_k.*

Figure 1 allows us to relate quantitative changes in PC scores to qualitative, dynamic changes in formant tracks: PC1 captures a rigid vertical translation whereby a positive or negative s_1 moves the formants closer together or further apart from each other respectively. Thus a positive s_1 corresponds to a greater degree of centralisation of the entire diphthong; and when s_1 is negative, the diphthong becomes more peripheral. For PC2, a positive s_2 causes F2-raising especially in the first part of the trajectory and F1-raising beyond the diphthong's temporal midpoint. The corresponding phonetic interpretation is a change from /eə, ɪə/ towards /ɪə, iə/ respectively (i.e. tongue fronting combined with a greater degree of mouth opening in the first part of the diphthong). For PC3, a negative s_3 causes a shift in the formants away from each other in the first part of the trajectory as well as an increased acoustic differentiation between the first and second parts. Therefore, a negative s3 corresponds phonetically to a shift from /eə/ towards /ɪə/ (a positive s_3 has the reverse effect).

2.2.1. Linear Mixed-Effects models

Linear mixed-effects (LME) models [26] were run separately with any of the PC scores s_1, s_2, or s_3 as a dependent variable and with various combinations of the five fixed and two random factors shown in Table 1. The idea is that factors like Diphthong, Age, Sex, etc. should predict formant shapes, which in turn are parameterised by PC scores and interpreted as illustrated in Section 2.2. For example, we expect that a binary factor coding for Diphthong /ɪə/ vs. /eə/ would predict a lower (/ɪə/) or higher (/eə/) value of s_3 whenever the diphthongs are realised canonically, as this follows the interpretation of PC3 (see Figure 1). Age was reduced to two levels from the original three because there were few differences between the mid-age and younger groups. Models were fitted that initially included all fixed factors, their two- and three-way interactions, as well as intercepts and slopes as specified in Table 1. These were subsequently pruned when terms were not significant. In the pruned models, marginal Pseudo-R^2 scores, which indicate the proportion of variance explained by the fixed effects only [27, 28], were 7.4%, 2.9% and 15.1% for the dependent variables s_1, s_2 and s_3 respectively: this shows that the selected set of fixed effects was more informative in predicting s_3 than in predicting the other scores.

Table 1: *Factors in the linear mixed-effect models. Tick marks indicate the presence of a random slope with respect to the fixed factor (e.g. row 1 denotes* `(Diphthong|Speaker)`*).*

		Random slopes	
Fixed factor	**Levels**	**Word**	**Speaker**
Diphthong	/ɪə/, /eə/		✓
Age	older, younger	✓	
Region	Hamilton, Wellington	✓	
Sex	male, female	✓	
Phrase Final	yes, no	✓	✓

2.3. Results

The focus is on whether there was an Age × Diphthong interaction which is relevant for testing both the presence and direction of the merger. For s_1, there was a significant effect of Diphthong ($\chi^2 = 12.5$, $p < 0.001$). This comes about (predictably) because /eə/ was acoustically more centralised (i.e. with a higher F1, lower F2 and hence formants that are closer together) than /ɪə/ . Neither Age nor its interaction with Diphthong were significant. For s_2, neither Diphthong, Age, nor their interaction were significant. For s_3, there was a significant influence of Diphthong ($\chi^2 = 34.5$, $p < 0.001$) that comes about for two reasons (Figure 2): firstly, because /eə/ was more monophthongal i.e. with flatter formants; and secondly, because F1 and F2 were closer for /eə/ than for /ɪə/ in the first part of the diphthong. There was also a significant Age × Diphthong interaction ($\chi^2 = 18.8$, $p < 0.001$) but no further three-way interactions involving these fixed factors. The significant Age × Diphthong interaction comes about because these acoustic differences between /eə/ and /ɪə/ were more marked for older than for younger speakers (Figure 2). Post-hoc tests showed that the age groups differed significantly from each other in /eə/ ($s_{3,\text{older}} - s_{3,\text{younger}} = 0.08$, s.e. $= 0.018$, d.f. $= 92.4$, $p = 0.001$) but not in /ɪə/. Overall, these results support the hypothesis that the sound change has affected /eə/ but not /ɪə/, and that the direction is indeed /eə → ɪə/.

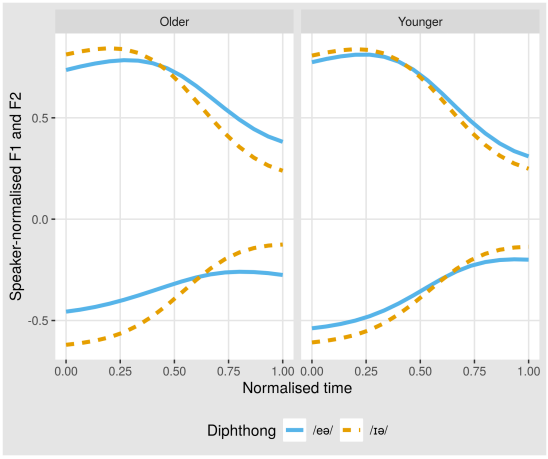

Figure 2: *Formant tracks (F2 top, F1 bottom in each panel) as modulated by Eq.* (1) *, where s_3 values are expected means predicted by the corresponding LME model for all combinations of Diphthong × Age.*

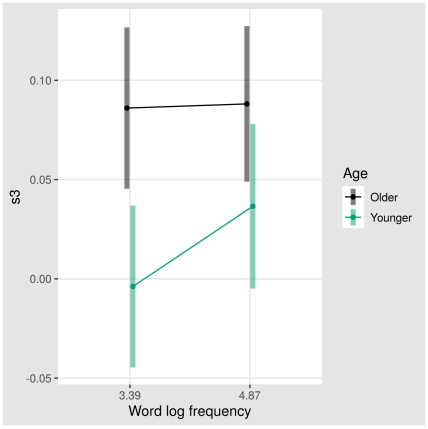

Figure 3: *Predicted variation in s_3 against log lexical frequency of /eə/-words, with 95% confidence intervals. Representative low/high frequency values correspond to mean ± st. dev.*

3. Lexical and coarticulatory effects

3.1. Method

We ran two mixed models in order to test for the effects of lexical frequency and for the influence of the preceding consonant's place of articulation on the sound change in progress. In both cases, the dependent variable was s_3 i.e. the FPCA parameter that had been shown to be most critical for the /eə/ → /ɪə/ sound change in Section 2.3. The test was applied to /eə/ only, since this was the diphthong in which the effects of lexical frequency as well as any influences of the place of articulation were predicted to be manifested [11], [19] in an /eə/ → /ɪə/ sound change. For the first mixed model, the fixed factors were log. Lexical Frequency (LF) which interacted with Age. LF was a numeric variable based on Zipf values i.e. standardized measures of word frequencies and calculated as $\log_{10}$(frequency per billion words) [29]. The random factors were (i) Word with slope calculated with respect to Age; and (ii) Speaker with slope calculated with respect to LF. If the sound change is conditioned by lexical frequency in the way predicted by [11], then the younger and older speakers' /eə/ should be closer together for high frequency words which should be manifested in the model as a significant interaction between the fixed factors Age and LF. For the second mixed model, the fixed factors were Age and Place of Articulation (PoA, two levels: [± cor]). The random factors were (i) Word with slope calculated with respect to Age; and (ii) Speaker with slope calculated with respect to PoA. The category [+cor] was defined to include /eə/ following velars /k,g/, post-alveolars /tʃ,dʒ,ʃ/, and alveolars /t,d,n,l/ given that these are the contexts with a high F2-locus in which the diachronic /eə/ → /ɪə/ shift is most likely to be advanced. [-cor] included /eə/ in all other contexts. Following [19], the prediction was that there should be an interaction between Age and PoA brought about by the closer approximation between younger and older speakers' /eə/ in the [+cor] context.

3.2. Results

The first mixed model showed a not quite significant interaction between Age and LF (F[1, 27.4] = 3.7, p = 0.07). However, as Figure 3 shows, the interaction came about because of a (not quite significant) shift in younger speakers' /eə/ towards /ɪə/ in low frequency words. There is therefore no evidence from these data that high frequency words lead this sound change by which /eə/ shifts towards /ɪə/. In the second model, the interaction between Age and PoA was not significant. The difference between older and younger speakers in /eə/ was about the same in both place of articulation contexts. There is therefore no evidence to suggest that the sound change is more advanced in a coronal than in other contexts.

4. Discussion and Conclusions

Independently of the merger in progress, one of the main findings in this study is that New Zealand English /eə/ differs from /ɪə/ not just because it is more central (i.e. with a lower F2 and higher F1) but also because it is more monophthongal as shown by its flatter formants compared with those of /ɪə/. Such findings about differences in the diphthongs' dynamic shapes have emerged naturally through the application of FPCA [18], a technique that processes time-varying, multi-dimensional signals without the need to consider static targets in either of the formants separately.

The present apparent-time analysis comparing younger with older speakers has also confirmed that there has been an ongoing sound change in New Zealand English by which /eə/ has shifted into the /ɪə/ space [4, 5, 8]. The new finding is that this approximation towards /ɪə/ involves not just raising and fronting of /eə/ in the first part of the diphthong, but also an increase from a relatively monophthongal /eə/ towards a markedly more diphthongal /ɪə/. That is, the sound change by which /eə/ has shifted towards /ɪə/ has also come about through an increase in which the falling diphthong is internally differentiated.

Finally, there is some indirect evidence from this study that the diphthongal shift in /eə/ could be linked to DRESS-raising [6] that has formed part of the New Zealand English front vowel shift [7, 8, 30]. First, consider the finding that place of articulation of the preceding consonant has no influence on the /eə/ → /ɪə/ approximation. This is exactly what would be expected if these sound changes are linked, given that DRESS-raising is not conditioned by phonetic context but instead linked to other shifts including in particular TRAP-raising and KIT-centralisation. Second, although we found no evidence that the /eə/ → /ɪə/ approximation was led by high frequency words (contrary to the predictions of the computational sound change model in [11]), we did find a (non-significant) tendency for this merger to be more advanced in younger speakers' low frequency words. This provides a further possible link to New Zealand English DRESS-raising in which, as Hay et al [7] have recently shown, changes are more advanced in low than in high frequency words.

The more general conclusion from this study is that FPCA provides a new way to quantify sound changes in a multi-dimensional space which for too long have been modelled by static snapshots of separately analysed formants at single points in time [31], even when, as in the raising of American English /eɪ, au, aɪ/ (FACE, MOUTH, PRICE) analysed in [31], the vowels are very obviously diphthongal involving changes to both formant frequencies.

5. Acknowledgements

This research was supported by European Research Council Grant no. 742289 'Human interaction and the evolution of spoken accent (2017-2022).

6. References

[1] E. Gordon and M. Maclagan, "A study of the /ɪə/ ∼ /eə/ contrast in New Zealand English," *The New Zealand Speech-Language Therapists' Journal*, vol. 38, pp. 16–29, 1985.

[2] M. Maclagan and E. Gordon, "Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger," *Language Variation and Change*, vol. 8, no. 1, pp. 125–147, 1996.

[3] E. Gordon and M. Maclagan, "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*, vol. 21, no. 2, pp. 215–238, 2001.

[4] M. Rae and P. Warren, "Goldilocks and the three beers: sound merger and word recognition in NZE," *New Zealand English Journal*, vol. 16, p. 33, 2002.

[5] J. Hay, P. Warren, and K. Drager, "Factors influencing speech perception in the context of a merger-in-progress," *Journal of Phonetics*, vol. 34, no. 4, pp. 458–484, 2006.

[6] J. Holmes and A. Bell, "On shear markets and sharing sheep: The merger of EAR and AIR diphthongs in New Zealand English," *Language Variation and Change*, vol. 4, no. 3, pp. 251–273, 1992.

[7] J. Hay, J. Pierrehumbert, A. Walker, and P. LaShell, "Tracking word frequency effects through 130 years of sound change," *Cognition*, vol. 139, pp. 83–91, 2015.

[8] C. Watson, M. Maclagan, and J. Harrington, "Acoustic evidence for vowel change in New Zealand English," *Language variation and change*, vol. 12, no. 1, pp. 51–68, 2000.

[9] P. Warren and J. Hay, "Using sound change to explore the mental lexicon," *Cognition and Language: Perspectives from New Zealand*, p. 105, 2006.

[10] P. Warren, J. Hay, and B. Thomas, "The loci of sound change effects in recognition and perception," *Laboratory Phonology*, vol. 9, no. 87-112, 2007.

[11] S. Todd, J. Pierrehumbert, and J. Hay, "Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model," *Cognition*, vol. 185, pp. 1–20, 2019.

[12] C. Watson, J. Harrington, and Z. Evans, "An acoustic comparison between New Zealand and Australian English vowels," *Australian Journal of Linguistics*, vol. 18, no. 2, pp. 185–207, 1998.

[13] M. Kennedy, "Prince Charles has two ears/heirs: semantic ambiguity and the merger of NEAR and SQUARE in New Zealand English," *New Zealand English Journal*, vol. 18, p. 13, 2004.

[14] C. Langstrof, "The centring diphthongs of New Zealand English in the Intermediate Period: an acoustic analysis," in *Proceedings of the 10th Australian International Conference on Speech Science & Technology*, 2004, pp. 207–212.

[15] J. Harrington and F. Schiel, "/u/-fronting and agent-based modeling: The relationship between the origin and spread of sound change," *Language*, vol. 93, no. 2, pp. 414–445, 2017.

[16] M. Wieling, "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*, vol. 70, pp. 86–116, 2018.

[17] J. Ramsay and B. Silverman, *Functional Data Analysis*. Springer Series in Statistics, 2005.

[18] M. Gubian, F. Torreira, and L. Boves, "Using functional data analysis for investigating multidimensional dynamic phonetic contrasts," *Journal of Phonetics*, vol. 49, pp. 16–40, 2015.

[19] P. Warren, "Word recognition and sound merger," in *Cognitive linguistics investigations*. John Benjamins, 2006, pp. 169–186.

[20] J. Garofolo, "TIMIT acoustic phonetic continuous speech corpus," *Linguistic Data Consortium, 1993*, 1993.

[21] J. Millar, J. Vonwiller, J. Harrington, and P. Dermody, "The Australian national database of spoken language," in *Proceedings of ICASSP'94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1. IEEE, 1994, pp. I–97.

[22] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326–347, Sep. 2017.

[23] R. Winkelmann, J. Harrington, and K. Jänsch, "EMU-SDMS: Advanced speech database management and analysis in R," *Computer Speech & Language*, pp. 392–410, 2017.

[24] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," http://www.praat.org/, 2019, [Version 6.0.49, retrieved 2 March 2019].

[25] B. Lobanov, "Classification of russian vowels spoken by different speakers," *The Journal of the Acoustical Society of America*, vol. 49, no. 2B, pp. 606–608, 1971.

[26] B. West, K. Welch, and A. Galecki, *Linear mixed models: a practical guide using statistical software*. Chapman and Hall/CRC, 2014.

[27] S. Nakagawa and H. Schielzeth, "A general and simple method for obtaining R2 from generalized linear mixed-effects models," *Methods in Ecology and Evolution*, vol. 4, no. 2, pp. 133–142, 2013.

[28] P. Johnson, "Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models," *Methods in Ecology and Evolution*, vol. 5, no. 9, pp. 944–946, 2014.

[29] W. Van Heuven, P. Mandera, E. Keuleers, and M. Brysbaert, "SUBTLEX-UK: A new and improved word frequency database for British English," *The Quarterly Journal of Experimental Psychology*, vol. 67, no. 6, pp. 1176–1190, 2014.

[30] P. Warren, "Quality and quantity in New Zealand English vowel contrasts," *Journal of the International Phonetic Association*, vol. 48, no. 3, pp. 305–330, 2018.

[31] W. Labov, I. Rosenfelder, and J. Fruehwald, "One hundred years of sound change in Philadelphia: Linear incrementation, reversal, and reanalysis," *Language*, pp. 30–65, 2013.