# SYSTEMATIC COVARIATION OF MONOPHTHONGS ACROSS SPEAKERS OF NEW ZEALAND ENGLISH

James Brand[1*], Jen Hay[1], Lynn Clark[1], Kevin Watson[1], Márton Sóskuthy[2]

[1]NZILBB, University of Canterbury, [2]University of British Columbia
*james.brand@canterbury.ac.nz

## ABSTRACT

The study of sound change tends to concentrate on trajectories of particular variables in isolation. We explore a methodology for testing whether individual sound changes cluster together across individual speakers. Adopting a 'big data' approach, we extracted speaker intercepts from mixed-effects models predicting normalised F1 and F2 for 12 monophthongs. Principal Components Analysis on the intercepts reveals three significant dimensions of covariation. PC1 separates speakers who vary in their production of boundary vowels. PC2 indicates covariation in a cluster of vowels that may be explained by stylistic or social factors. PC3 highlights the existence of 'leaders' and 'laggers' in sound change. This technique provides a means for finding structure in large data sets, moving beyond the analysis of isolated variables and towards one that considers entire systems.

**Keywords:** Sound change, Covariation, Princial Components, New Zealand English

## 1. INTRODUCTION

Foundational work in the field of 'first-wave' variationist sociophonetics ([19]) focuses on the sociolinguistic variable, and has repeatedly shown that variables can be socially and stylistically stratified, especially in cases where the variable is undergoing change. This work is usually large scale, and typically conducts a macro level analysis by aggregating across many individuals in several generations of speakers. The majority of such studies report on variation or change in an individual variable. Even if multiple variables are investigated in the same study, they are nearly always examined in isolation of each other, with the exception of a few specific sound changes, which are hypothesized to be causally related (e.g. chain-shifts, see [14]).

Work in the so-called 'third-wave' of sociolinguistics [6] has focused more explicitly on individual speakers, exploring how linguistic styles, as collections of different phonetic variants, unfold over the course of a conversation. It argues that speakers display stylistic variation through combinations of variants, and the meaning of a particular variant cannot be properly interpreted in isolation of the landscape of other variants with which it co-occurs. This sheds light on how sounds are used in actual conversation, but the focus is usually on micro-level variation, often within a single speaker (e.g. [3, 16]).

From first-wave studies, we have learned how individual variables pattern and change over time at the community level, and from third-wave studies, we now know about how variants might cluster together in individual interactions. However, the latter has thus far been mostly focused on individual conversations, and has not yet reaped the benefits of a 'big data' approach, in terms of understanding how speech communities might systematically cluster features, while the former lacks a precedent for studying how sound systems vary across speakers and time. When combined, the two literatures raise new questions about how *sound systems* - understood as the systematic covariation of a range of different phonetic variants, vary across speakers.

We investigate covariation between monophthongs over the history of New Zealand English. We are interested in two key questions. First, can we find systematic patterns of realizations of vowels, across speakers, that might relate to indexical meaning. As stated by Guy and Hinskens:

> How similar or different are the indexicalities of particular linguistic forms? Are there clusters of variables that coherently index, or are associated with, the dimensions or subdivisions of a community [10, p.4]

Second, can we identify individuals who 'lead' or 'lag' in sound change, in a way that can be observed across multiple ongoing changes?

> Are there socially identifiable leaders of change who tend to use all the innovative variants together, or are different innovations subject to differentiated social inter-

pretations and individuated patterns of usage? [10, p.4]

We analyse all available monophthongs from the Origins of New Zealand English corpus (ONZE) [9, 8]). We make use of the fact that the ONZE corpus has an extensive number of tokens available for analysis, from speakers who were born across a 130+ year time period. This provides an ideal resource which can be used to investigate the question of whether we can observe covariation of phonetic variants across different speakers.

Building on work in [18], we fit separate mixed effects regression models to each vowel category in the data set. As we explain further below, the models included year of birth as one of the predictors, and so are able to examine possible covariation among variants while controlling for speaker age. To do this, we extracted the by-speaker random intercepts (a measure of whether the speaker's formants are higher or lower than we would expect, given their year of birth and gender), then analysed these intercepts using a Principal Components Analysis (PCA). This method provides a way to expose the correlation structure of different phonetic variants across speakers, allowing us to identify whether certain subsets of the variables tend to cluster together in their usage.

## 2. METHODS

### 2.1. Materials

Our data are drawn from the ONZE corpus, which comprises transcribed recordings from speakers with birth years ranging from 1851 to 1988, providing data on accent variation and change across 130+ years. The full set of ONZE transcriptions has been force-aligned at the phoneme level using the HTK toolkit [20] available within the software package LaBB-CAT [7].

### 2.2. Data processing

In order to assess the covariation of vocalic variables across speakers, we first extracted all available tokens that contained monophthongs in the ONZE corpus, then automatically extracted F1 and F2 values at the midpoint of each vowel. This resulted in a data set of over two million tokens from 636 speakers. Next, we removed speakers with < 10 tokens for any single vowel, or with no demographic information in the corpus. Further to this, we removed tokens that were more than 3 standard deviations outside the mean F1 or F2 value (as these were likely inaccurate measurements), calculated per vowel for each speaker. Finally, we removed tokens from a list of 73 grammatical words which are often high frequency and occur in unstressed environments. We then normalised all F1 and F2 values using the Lobanov method from the *Vowels* package in R (see [1]). The final data set comprised 978,610 tokens, from 23,943 unique word forms across 572 speakers (283 female, 289 male). A summary of the tokens by vowel can be found in Table 1.

**Table 1:** No. of tokens per vowel in final data set.

| Vowel | N Tokens | % |
|---|---|---|
| DRESS | 116,993 | 11.96 |
| FLEECE | 73,806 | 7.54 |
| FOOT | 21,249 | 2.17 |
| GOOSE | 42,915 | 4.39 |
| KIT | 259,788 | 26.55 |
| LOT | 52,481 | 5.36 |
| NURSE | 25,804 | 2.64 |
| SCHWA | 162,566 | 16.61 |
| START | 44,842 | 4.58 |
| STRUT | 70,721 | 7.23 |
| THOUGHT | 58,676 | 6.00 |
| TRAP | 48,769 | 4.98 |

Some of these vocalic variables have received considerable attention in work on NZ English, particularly TRAP, DRESS and KIT which have been reported to be changing in tandem as part of a push-chain shift [14, 12]. In order to obtain an estimate of how far advanced a speaker is in each of these individual sound changes, irrespective of their gender or year of birth, we ran a series of 24 linear mixed-effects models, fitted using the *lme4* package in R [2, 17]. Each model predicted either the normalised F1 or F2 of each of the 12 monophthongs, with a fixed effect structure of gender*year of birth + stressed environment + speaker speech rate, in addition to random intercepts for speaker and word. Year of birth was modelled as a non-linear predictor using a restricted cubic spline with 4 knots (*rms* package [11]). From each of the models, we extracted the by-speaker random intercepts, thus providing us with 24 data points for each speaker corresponding to how their F1 or F2 values for each vowel differ compared to the group (see [5, 18]).

## 3. RESULTS AND DISCUSSION

As a first step, we attempted to explore whether we could find covariation of the vowels using a hierarchical agglomerative cluster analysis. This initial exploration strongly motivated the interpretation that intercepts were non-randomly associated with each other. Some of the covariation seemed

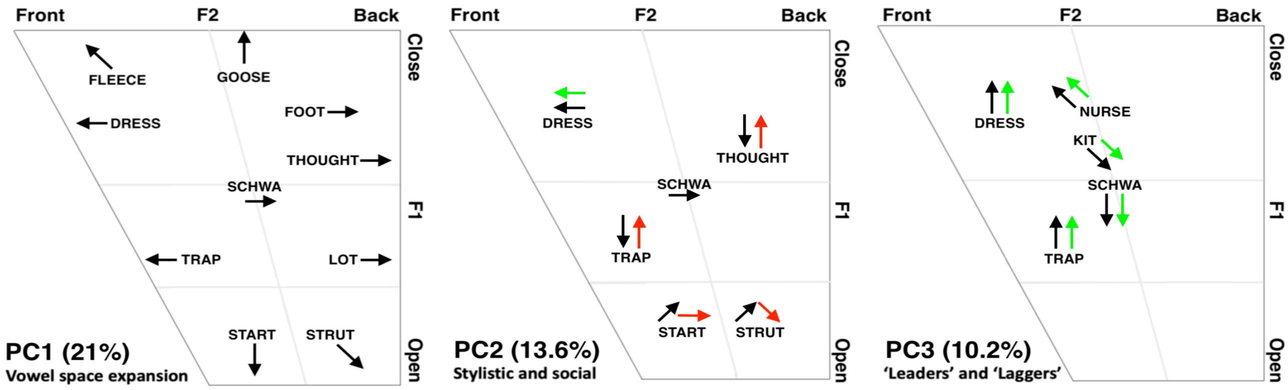


**Figure 1:** Vowel spaces capturing each of the 3 principal components, with variance explained given in brackets. Vowels are shown when the absolute value of their loading is >.2 on the relevant component, with loading values given in black text next to directional arrow. In PC2 and PC3, coloured arrows represent the direction of change the vowel is moving in, with green arrows indicating movement in the same direction with PC loading, whilst red arrows indicate opposite direction. All PC2 and PC3 loadings with an absolute value of >.2 are on formants that are significantly changing over time, with the exception of START F2, and SCHWA F2. Note that the exact position of the vowels is not data-driven, but schematic and based on past descriptions of NZE, see [14]

to be driven by differences in overall vowel space, whereas some of it reflected covariation of structurally related changes. In order to try and disentangle possible dimensions of covariation that might be underpinning these stable clusters, we turned to PCA. PCA was used with the main objective of identifying whether and to what extent specific vocalic variables co-vary over time in NZE. The result of the PCA motivated three main components, which in combination explained 44.8% of the total variance. The loadings on the three principal components (hereafter referred to as PC1, PC2 and PC3) are shown in Figure 1, for predictors with loadings with an absolute value greater than 0.2.

### 3.1. PC1: Vowel Space Expansion

The first component, PC1, accounts for 21% of the variance. Inspection of the loadings (left, Figure 1) reveals that it separates speakers who produce extreme values for boundary vowels, and those that do not. Speakers with high PC1 have high F1 for high vowels, high F2 for back vowels, low F1 for low vowels, and low F2 for front vowels. Thus, speakers with high PC1 tend to produce less extreme vowels than speakers with low PC1. This operates, in a sense, as an extra layer of speaker normalization, whereby speakers are separated by the size of their vowel space, rather than applying a uniform vocal tract length adjustment which shunts F1/F2 uniformly up or down.

Reduction in overall vowel space is known to correlate with casual speech [13]. Overall vowel space size has been shown to vary across speakers, in a way that correlates with speaker intelligibility [4]. This dimension is also interesting for social reasons, differentiating speakers according to the overall degree of hyper/hypo-articulation employed.

### 3.2. PC2: Stylistic Variation and Social Meaning

PC2, accounting for 13.6% of the variance, represents a further group of vowels that are not operating independently. This group is clearly not related to speaker normalization issues, as they contain negative correlations as well as positive ones. For example, speakers with high PC2 have a higher than expected F1 for THOUGHT, and a lower F1 for START. Moreover the direction of these loadings is not uniformly aligned with directions of sound change. Coloured arrows show the direction of change as indicated by year of birth coefficients from simple linear regressions of each vowel. In some cases (e,g, DRESS) the principal component is aligned with the direction of change, and in others (e.g. THOUGHT) it is in opposition with it. Furthermore, there are several vowels in this group which are not known for rapid change, and in two cases (SCHWA F2 and START F1), the coefficient for year of birth does not reach significance. Here, then, we have some significant covariation of vowel production that is neither related to vowel space expansion nor to sound change. This seems likely to indicate a cluster of vocalic variables not undergoing vigorous change. It is possible that these vowels collectively cohere to form a particular style or to index similar social meanings in this community. The social dimension of these clusters of variables will be explored in future work.

### 3.3. PC3: Leaders and Laggers

As can be seen from the right panel of Figure 1, the loadings in PC3, accounting for 10.2% of the variance, are directly aligned with the direction of change for all vowels. All vowels with high loadings are involved in significant change over this time period. This component also captures the short front vowel shift, with speakers who have higher TRAP and DRESS, also having more central KIT. That is - these speakers can be seen to be leading in the short front vowel shift. Interestingly, these speakers also show some patterns in other vowels which are not known to be part of the chain shift, but have changed over the same time period. For example, speakers leading in the chain shift also have a higher, fronter NURSE vowel, and thus are leading in the sound change involving NURSE raising and fronting ( [15])

These patterns of covariation suggest that the previously reported short front vowel shift exists within individual speakers rather than just across generations. Moreover, they suggest that there is non-random patterning of other vowels that are moving over the same time period. The existence of PC3 points toward individuals being overall 'leaders' and 'laggers' in sound change. Without this particular methodology, which accounts for the - much more substantial - covariation across speakers due to vowel space size - this sound-change related covariation would be much harder to capture.

In order to further explore the degree to which these patterns can really be interpreted as existing within individuals, as opposed to reflecting time-based differences that we have not adequately captured in our regression models, we inspected the relationship between each component and speakers' social characteristics. Figure 2 shows this relationship for PC3. Each point represents a speaker. Their year of birth is plotted against their loading for PC3. Horizontal lines have been added at $+2$ and $-2$ to visually separate speakers with more extreme positive or negative values. We can see that there is no strong relationship with year of birth or with gender. Indeed these are not significant predictors of the component when a linear regression is fitted to the speaker loadings. Speakers exist at all time periods who are substantial 'laggers' (above the top line) or 'leaders' (below the bottom line) with respect to this cluster of vowels, given where we would expect them to be for their year of birth and gender. This provides evidence in support of the interpretation that what we are seeing relates to individuals. At any given time, over the history of NZ English, there have been people who are 'leaders' and people who have been 'laggers' in this constellation of vowel changes. The vowels are not operating independently, but are related - across individuals - in a non-random way. Plots for PC1 and PC2 tell a similar story, with no clear pattering with respect to year of birth or gender.

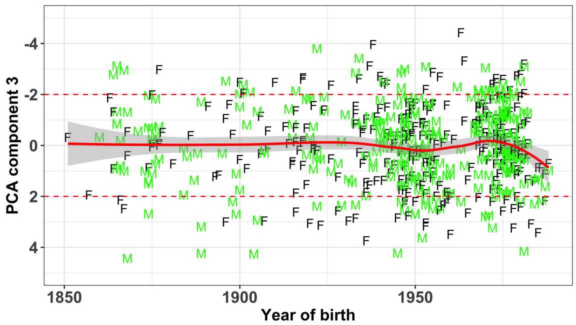


**Figure 2:** Speaker loadings for PC3, as predicted by year of birth. Horizontal lines are added to visually separate substantial 'leaders' (below the bottom line) and 'laggers' (above the top line), with a loess smooth fitted (solid red line).

## 4. CONCLUSION

By modelling individual monophthong formants and extracting the by-speaker random intercepts, we can assess whether each speaker's vowel formant is higher or lower than predicted given their year of birth and gender. PCA over these intercepts reveals considerable structure. First, we can distinguish a significant pattern of covariation relating to individual vowel space size. Second, a collection of vowels co-vary systematically in a manner that is independent of articulatory pressures, vowel space size, or direction of change. This suggests that they may be operating together in the construction of some shared social or stylistic meaning. Finally, a set of vowels which have undergone substantial change show significant patterning across individuals. We can identify individuals who are 'leaders' or 'laggers' across a set of changing vowels.

Of course, it is unlikely that variables covary in the same way for all speakers and at all times. Such changes will affect parts of the covariance structures, as systemic meanings and styles change. In future work, we will be investigating the degrees to which different aspects of the covariation we have uncovered are stable across time. While exploratory, we believe that this technique lays down a very promising path for moving beyond the study of isolated variables, toward a fuller understanding of how individuals are situated within overall sound systems.

## 5. REFERENCES


[1] Adank, P., Smits, R., Van Hout, R. 2004. A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America* 116(5), 3099–3107.

[2] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.

[3] Becker, K. 2014. Linguistic repertoire and ethnic identity in new york city. *Language & Communication* 35, 43–54.

[4] Bradlow, A. R., Torretta, G. M., Pisoni, D. B. 1996. Intelligibility of normal speech i: Global and fine-grained acoustic-phonetic talker characteristics. *Speech communication* 20(3-4), 255–272.

[5] Drager, K., Hay, J. 2012. Exploiting random intercepts: Two case studies in sociophonetics. *Language Variation and Change* 24(1), 59–78.

[6] Eckert, P. 2018. *Meaning and Linguistic Variation: The Third Wave in Sociolinguistics*. Cambridge University Press.

[7] Fromont, R., Hay, J. 2008. ONZE Miner: the development of a browser-based research tool. *Corpora* 3(2), 173–193.

[8] Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., Trudgill, P. 2004. *New Zealand English: Its origins and evolution*. Cambridge University Press.

[9] Gordon, E., Maclagan, M., Hay, J. 2007. The onze corpus. In: *Creating and digitizing language corpora. Volume 2: Diachronic Corpora*. Palgrave 82–104.

[10] Guy, G., Hinskens, F. 2016. Linguistic coherence: Systems, repertoires and speech communities. *Lingua* 172(173), 1–9.

[11] Harrell Jr, F. E. 2016. rms: Regression modeling strategies. 2016. *R package version*.

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

[13] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: *Speech production and speech modelling*. Springer 403–439.

[14] Maclagan, M., Hay, J. 2007. Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language variation and change* 19(1), 1–25.

[15] Maclagan, M., Watson, C. I., Harlow, R., King, J., Keegan, P. 2017. Investigating the sound change in the New Zealand English Nurse vowel /ɜ/. *Australian Journal of Linguistics* 37(4), 465–485.

[16] Podesva, R. J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona. *Journal of Sociolinguistics* 11(4), 478–504.

[17] R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[18] Sóskuthy, M., Hay, J., Maclagan, M., Drager, K., Foulkes, P. 2017. The closing diphthongs in early New Zealand English. In: Hickey, R., (ed), *Listening to the past: Audio records of accents of English*. Cambridge University Press.

[19] Trudgill, P. 1974. *The social stratification of English in Norwich*. Cambridge University Press.

[20] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., et al., 2002. *The HTK book*. Cambridge University Engineering Department.