# Systematic co-variation of monophthongs across speakers of New Zealand English

James Brand[a,*], Jen Hay[a,b], Lynn Clark[a,b], Kevin Watson[a,b], Márton Sóskuthy[c]

[a] *New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, New Zealand*
[b] *Department of Linguistics, University of Canterbury, New Zealand*
[c] *Department of Linguistics, The University of British Columbia, Canada*

**Abstract**

The study of phonetic change tends to concentrate on trajectories of particular variables in isolation, but it has proven challenging to move beyond an analysis of individual variables or small groups of variables, towards a better theoretical and empirical understanding of entire *vowel systems*. We introduce a large scale analysis of how elements of full sound systems co-vary across hundreds of speakers, demonstrating how constellations of vocalic variables operate together. Our focus is on changes to F1 and F2 for 10 monophthongs of New Zealand English. We first obtain estimates of how advanced each speaker is with respect to changes in each of the vowels, irrespective of known predictors of sound change (i.e. year of birth, gender, speech rate). This is done by extracting by-speaker intercepts from Generalised Additive Models. We then use Principal Component Analysis on these intercepts to investigate the underlying structural co-variation that exists across the vocalic variables. Our results demonstrate the inter-relatedness of production patterns and changes across different vowels. Within a large subset of vowels, we see 'leaders' and 'laggers'. However, there are also vowels which appear to be linked together through structural relationships, and some sets of vowels which appear to carry opposing social meaning, such that if you are innovative in one, you tend to be conservative in the other. Our results offer a means to overcome long-standing methodological challenges in the study of phonetic co-variation, offering novel insights into the structure of sound systems across large groups of speakers.

*Keywords:* Sound change, Co-variation, Monophthongs, chain-shift, New Zealand English

*Corresponding author:
*Email address:* james.brand@canterbury.ac.nz (James Brand)

## 1. Introduction

The study of phonetic variation and change has been built on a large number of increasingly sophisticated analyses of how specific linguistic variables are used across different speakers and contexts, and how they change over time. The vast majority of contemporary studies in this literature focus on the patterning of a single variable. From this, we have learned that the phonetic implementation of phonological variables can change radically across generations (Labov, 2001; Trudgill, 1974), across contexts (Love & Walker, 2013; Foulkes & Hay, 2015), and across social groups (Eckert, 2000; Mendoza-Denton, 2014; Dickson & Hall-Lew, 2017). 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 hypothesised to be causally related (e.g. chain-shifts, see Gordon 2002; Maclagan & Hay 2007; Hay et al. 2015).

However work from a variety of literatures predicts that individual variables would not vary in independently of each other, but rather that they should operate as part of an overall sound-system. Work in the so-called 'third-wave' of sociolinguistics (Eckert, 2016, 2018), for example, 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. Becker 2014; Podesva 2007).

The third-wave literature is explicit about the idea that variables do not operate in isolation of each other, and thus might predict that some speakers would be more likely to use certain clusters of variants together than other speakers. Indeed, such clustering is also predicted from a variety of other perspectives. Apart from being involved in the co-construction of related social meaning (Eckert, 2018), or because they are involved in a chain-shift (Gordon, 2002), we might also expect to see cases of parallel shifts – sound changes that are non-randomly associated with each other, perhaps because they share a phonological feature (Fruehwald, 2013). Longer term evolutionary pressures on the language system may also lead us to expect relationships between different sound changes. For instance, the interplay between constraints on articulation and pressures towards the maintenance of contrast may force sound categories to arrange themselves into systematic constellations (Liljencrants & Lindblom, 1972) and to evolve in tandem (Sóskuthy, 2015). Any one of these forces might lead there to be a systematic patterning across speakers, whereby so tend to be advanced in one or more other variables.

Indeed, central in the current language variation and change literature, is the ongoing debate relating to the leaders of change - who leads sound change, and who accelerates its momentum? (Labov, 2001; Tamminga, 2019). And crucially, with respect to the current paper, if one is a 'leader of sound change', does that mean that this person is advanced in all sound changes? (Guy & Hinskens, 2016). In sum, despite the preponderance of single-variable analyses, there are many reasons to expect that we might find sub-systems of variables which pattern together. If so, then it is crucial to investigate patterns of co-variation across linguistic variables so that we can move towards a fuller understanding of the range of factors that influence and constrain language variation and change.

In this paper, we examine the systematic co-variation of phonetic realisations in the monophthongs of New Zealand English (NZE) across speakers born over a 118 year time period. Controlling for known sources of co-variation (e.g. speech rate, year of birth and gender), can we find clusters of vowels that are non-independent of each other across the history of NZE? And, if so, what can these patterns of co-varying vowels tell us about the nature of chain-shifts, the spread of sound change, the phonological system and the social meaning of phonetic variants?

In the remainder of this section, we first outline two factors that might serve as trivial sources of co-variation in the vowel system: individual differences in anatomy and articulatory setting, and changes to different vowels that happen to occur during the same period. We seek to control both of these clear sources of co-variation in our data set, as outlined below. This is followed by a discussion of potential sources of non-trivial co-variation, such as chain shifts, socially or phonologically driven parallel changes and evolutionary pressures on sound systems. We conclude the section with a set of research questions that follow from the foregoing discussion.

### *1.1. Known Sources of Vowel Co-variation*

*Differences in anatomy and articulatory setting*

At the most basic level, it is well understood that the formant structure of vowels is systematically linked to anatomical features – most notably the length of the vocal tract (Fant, 1970; Stevens, 2000). Vowel space size can also differ through a combination of anatomical factors and articulatory habits, such as degree of hyperarticulation (Lindblom, 1990). A variety of 'normalisation' techniques have been developed to try to remove this source of variation. In this paper, we adopt a modified version of the Lobanov normalisation procedure (Lobanov, 1971), which is intended to ensure that the F1 and the F2 measurements for different speakers cover approximately the same space. This should remove effects of vocal tract length and vowel space size. Note, however, that there may be effects related to anatomy(Johnson, 2018) and/or articulatory setting (Honikman, 1964) which affect some parts of the vowel space more than others (Johnson, 2018). If this is the case, then we might expect the vowels most affected by such variation to systematically co-vary across speakers.

Moreover, the question of whether articulatory setting is a truly trivial source of variation is by no means straightforward. Changes within the Californian Vowel System, for example, have been attributed to variation in jaw setting, which may be more typically associated with certain personas (Pratt & D'Onofrio, 2017).

Recent work on Glasgow English by Sóskuthy & Stuart-Smith (to appear) shows that articulatory setting itself – in their case, tongue body position as cued by F3 – can be the subject of diachronic change, resulting in shifts in formant values that manifest across all sonorants. Sóskuthy & Stuart-Smith (to appear) argue that the distinction between segmental and voice quality changes is necessarily fraught, insofar as they often operate on the very same articulatory features and acoustic cues. This can make it difficult to distinguish between parallel shifts across different sounds and changes in articulatory setting. We take a conservative approach in this paper by factoring out overall changes of the type observed in Sóskuthy & Stuart-Smith (to appear), but acknowledge that such processes are potentially important in shaping vowel systems.

*Parallel time-course of change*

Our data covers 118 years of time, during which NZE, a relatively new variety of English, was being formed. The speakers with the earliest years of birth, of course, have vowel spaces that resemble each others' more than they resemble the vowel spaces of speakers born several generations later. If we simply take a large number of speakers and look at whether the realisation of one vowel is predictive of the realisation of a second, we would find many significant patterns of 'co-variation' across vowels that are statistically linked simply due to the fact that they have both undergone substantial change during the same period. This is inevitable given the time-depth of our data. Such correlations would not necessarily reveal any systematic link between the sounds, apart from the fact that they happen to be changing simultaneously. This, of course, can be meaningful in and of itself, but it is difficult to establish causal links across such time series. Thus, this type of cross-speaker correlation is not the focus of this paper. What we are interested in is whether the realisations of two vowels are systematically related within speakers, over and above the fact that the sound changes happen to have unfolded over the same time period.

Our methodology attempts to remove this source of co-variation by controlling for year of birth and gender in our statistical models. Holding constant year of birth and gender, can we still see relationships between vowels, such that, at any given time, speakers who are advanced in one vowel change are also likely to be advanced (or perhaps conservative) in one or more other vowel changes?

### *1.2. Hypothesised Sources of Vowel Co-variation*

If we are able to control for the clearly predicted sources of co-variation outlined above, then there are a number of other hypothesised mechanisms through which vowels may nonetheless co-vary.

#### *Chain-shifting relationships*

Vowel chain-shifts are cases of systematic relationships between adjacent vowels. The relationship may be a 'pushing' relationship, in which a vowel encroaches on another's space, leading to a retreating movement by the encroached-upon vowel, or it may be a 'pull' relationship, in which a vowel's movement leaves a vacuum, into which another vowel may move Gordon (2002). In NZE, the short front vowels have been argued to be in a chain-shifting relationship, with TRAP and DRESS raising and KIT centralising. Hay et al. (2015) have argued that the lexical frequency effects observed in the changes support a chain-shifting interpretation, in which the system evolves away from configurations leading to perceptual ambiguity.

Gordon et al. (2004) show that degree of innovation in TRAP (auditorily assessed) was statistically correlated with their degree of innovation of DRESS, whilst DRESS-raising was also correlated with KIT-centralisation. Although this may suggest that the chain-shifting relationship exists at the level of the individual, their analysis does not completely rule out the possibility that the correlations arise because the sound changes are unfolding over the same time period. Speaker year of birth is not controlled in the correlation analysis.

In Sóskuthy et al. (2017)'s analysis of changes in NZE diphthongs, they attempt to control for year of birth and known social effects. They fit a mixed-effects regression model to the vowel formant values, extracting the speakers' random intercepts as an index of how advanced they are in the change. They find a correlation between random intercepts from the models for PRICE and FACE, suggesting that these changes are causally linked in a sound change.

Boberg (2019) also looks at pairwise correlations, this time between normalised F1 and F2 measurements of front vowels implicated in a chain-shifting relationship: the Canadian vowel shift. He restricts the age of the participants, but both men and women are included. He finds pairwise correlations between front vowels participating in the shift, concluding this is evidence for a true chain-shifting relationship (although we note this could be carried by gender differences in participation in the shift, which are also demonstrated in the same paper). There are also some correlations between the movements of front vowels and those of back vowels, though Boberg finds it unlikely that this is 'causal', attributing it instead to coincidence and/or some shared social mechanism.

Chain-shifting, then, is one area where researchers have attempted to look for structural relationships between vowel positions across speakers. However, the focus has not been on a system-wide analysis, but rather on pair-wise correlations between neighbouring vowels. When non-adjacent vowels correlate, by contrast, (as with Boberg's (2019) front and back vowels), this is typically attributed to shared social meaning, as outlined below.

*Socially-driven parallel changes*

A growing literature dealing with potential co-variation amongst variables focuses on leaders in sound change, as reflected in the following quote from Guy & Hinskens (2016): "Are there socially identifiable leaders of change who tend to use all the innovative variants together, or are different innovations subject to differentiated social interpretations and individuated patterns of usage?" (2016: p.4).

Guy (2013) identifies "a dearth of research in the field that addresses whether clustering of variables actually occurs in the behavior of individuals" (p.64). He investigates pairwise correlations between 4 variables, and finds their realationships to be 'weak'. Since then, there has been a small flurry of studies investigating the question of whether variants of different variables cluster within speakers. The emerging answer to this appears to be that (when considered in a pair-wise fashion) some correlations can often be found, but that it is not the case that speakers tend to be uniformly 'leaders' or 'laggers' for all changes (see Tamminga (2019) for review). In other words, studies tend to find pairwise correlations between some, but not all, variables undergoing change.

Oushiro (2016) finds that a selection of structurally related and structurally non-related variables co-vary in their study of São Paulo Brazilian Portuguese. Watt (2000) notes parallel changes in the FACE and GOAT vowels of Tyneside English, which seem to be distributed in a similar way across speaker groups, attributing the co-variation between the two vowels to a shared social meaning. This argument is also used by Tamminga (2019), who conducts a series of pairwise correlations across vowels undergoing change in Philadelphia, and finds correlations between three out of six investigated changes. She controls for the parallel timecourse issue outlined above, by working with a homogeneous group of speakers, and statistically controlling for a number of social and linguistic factors. She argues that the co-varying vowel changes share some social meaning that is not shared by the others. Becker (2016) makes a similar claim, finding co-variation between lowering of the THOUGHT vowel and a decrease in non-rhoticity in New York. Becker proposes that young New Yorkers are trying to avoid association with a 'classic New Yorker' persona (p.97).

The degree of co-variation may also depend on the type of variable. Waters & Tagliamonte (2017) report no correlations between leaders of one variable and leaders of any other based on pairwise correlations of five spoken morphosyntactic variables. In another study of morphosyntactic changes in writing, over a longer time period, Nevalainen & Raumolin-Brunberg (2011) find that some individuals were innovative on a collection of variables. They note, however, that no-one was progressive on all changes underway and that "even among those predominantly progressive, the common pattern included not only in-between use but also conservatism with regard to some changes" (p. 33).

This is a topic of intense debate within sociolinguistics, but two methodological issues are pervasive. First, when variables are changing over time and are systematically distributed across groups, this needs to be controlled for in order for true individual effects to robustly emerge. Researchers who acknowledge this problem deal with it by looking for co-variation within a highly restricted sub-set of speakers who are socially more homogeneous (e.g. Tamminga (2019)). And secondly, the predominant statistical approach is to search for pairwise correlations between particular pairs of variables. While potentially locally revealing, it does not reveal any broader systematic structure that might extend beyond particular variable pairs.

*Shared Stylistic Meaning*

If speakers are 'leaders' in a constellation of changes (as outlined above) this may indicate that these changes carry related social meanings. However shared stylistic meaning between variants may also lead to co-variation independently of any patterns of ongoing change.

Extensive related evidence for the important stylistic role played by linguistic variables is provided by Eckert (2018). Eckert (2018, 2019) is clear in her interpretation that "Sociolinguistic variables do not occur independently, but as components of styles" (Eckert 2019:752). Individual speakers make a variety of choices with reference to their intended persona, and "Stylistic practice is a process of bricolage (Lévi-Strauss et al., 1962; Hebdige, 1984), or 'tinkering', in which people recombine things that are already at hand to create something new" (Eckert 2019:752). We thus expect that phonetic variants that share related social meanings, and contribute to related styles, should statistically co-occur when considered across a large sample of speakers.

*Phonologically-driven parallel changes*

Fruehwald (2013) outlines a variety of cases of 'parallel phonetic shifts' – shifts that are not linked to each other in a chain-shifting relationship, arguing that many such cases are actually phonological in nature. He includes in this diphthongs, that Watt (2000) claimed were socially linked, arguing that there may be a phonological link between them instead. In essence, Fruehwald (2013, 2017) proposes that co-variation can arise through the shifting phonetic implementation of an entire phonological class, possibly defined in terms of phonological features.

The proto-typical case of this phonologically-driven parallel shift is the case of back vowels, which have been found, across dialects, to undergo fronting. In such cases, Fruehwald claims, we may observe the back vowels fronting together – not primarily because of a shared chain-shifting relationship or shared social interpretation, but rather because speakers are shifting in their implementation of the feature 'back', which affects multiple vowels in parallel.

Chodroff & Wilson (2017) investigate a related case, although their variable of interest is not necessarily a feature undergoing change. They look at the Voice Onset Time (VOT) of English aspirated stops, finding correlations between VOT of the different stops, across speakers. They argue that "aspirated stops can be straightforwardly accounted for with a constraint that requires the talker-specific realisation of a phonetic property (e.g., glottal spreading) to be uniform across speech sounds. The uniformity constraint, which could extend to many other phonetic properties and sound classes, allows talkers to differ but imposes a common relational structure or pattern on their phonetic systems" (p. 31). The idea that there are phonetic uniformity pressures in language is previously argued for in Keating (2003).

The idea of phonologically driven parallels in the phonetic implementation of related phonemes might predict the existence of patterns of co-variation across speakers, for subsets of phonemes sharing a relevant phonological feature.

*Evolutionary pressures on the vowel space*

Finally, over longer time scales, it is well understood that there are evolutionary pressures on sound systems, and perhaps particularly on vowels. Thus, sound systems exhibit properties consistent with pressures towards symmetrical vowel spaces (Boersma, 1998) and vowel spaces that maximize dispersion and contrast (Liljencrants & Lindblom, 1972; Schwartz et al., 1997). Sóskuthy (2015) argues that such pressures interact to define stable states that determine the evolution of sound systems as wholes, as opposed to individual sound categories in a vacuum. Therefore, to the degree that such factors are indeed driving ongoing changes, we might expect to see co-variation between subsets of sounds that are, for example, jointly moving in a direction that increases overall symmetry.

### *1.3. Interim Summary*

Despite the fact that the study of language variation and change has focused largely on the isolated variable, there are a number of phenomena in the literature that might lead us to expect some co-variation across linguistic variables. We note that these are not at all mutually exclusive, and, indeed, it is possible (or perhaps even likely) that all of the above phenomena play a role in the evolution of sound systems. For this reason it is important to move beyond pairwise comparisons between variables hypothesised to be linked, and look simultaneously at the implementation of a range of sounds to gain a holistic view of patterns of co-variation across them.

### *1.4. Research questions*

This paper analyses the monophthongs of NZE. We aim to identify sub-systems within the vowel space – pairs or sets of vowels that are interconnected, when well-known sources of co-variation (anatomical and social) are controlled. We are particularly interested in the degree to which any patterns of co-variation might relate to patterns of ongoing sound change. Most of the NZE vowels have changed substantially over the 118 years of recordings we analyse. We ask whether there are any persistent systematic linkages between the realisations of these vowels. Controlling for a speaker's year of birth and gender, does the realisation of one monophthong predict the realisation of others, or are all the changes spreading independently through the population?

Can we identify individuals who generally ‘lead’ or ‘lag’ in sound change, in a way that can be observed across multiple ongoing changes (cf. the quote from Guy & Hinskens 2016 above)? Based on the work of others who have asked this question (see, e.g. Tamminga 2019 and references therein), we might expect some changes to be linked, but perhaps not all.

This leads to a further question: to the extent that patterns of co-variation exist, is it possible to isolate the mechanism underpinning it? Is it possible to distinguish between co-variation caused by chain-shifting relationships, shared socio-indexicality, shared phonology, or evolutionary biases? Whether this is possible will depend on which clusters of vowels are involved. If just the vowels in the NZE short front vowel shift co-varied, for example, and no other vowels, this would support the interpretation that these changes are being driven by factors at the individual level, not just at the community level, but would not provide any strong evidence in support of the other mechanisms discussed.

In order to address our research questions, we introduce a novel analysis procedure that aims to overcome the long-standing difficulties with the study of phonetic co-variation, and apply it to a suitably large corpus of NZE. The core steps are:

1. Compute estimates of how advanced speakers are in terms of their realisations of different vowels relative to others (based on their F1 and F2 values). This step involves statistically modelling formant frequencies and controlling for various predictors of change in the data. From these models we can extract intercepts for each speaker, which are used as a measure of how they differ from the rest of the population for each vowel (see Drager & Hay 2012, Fruehwald 2013).
2. Run a Principal Component Analysis (PCA) on the resulting data set of speaker intercepts. This will allow us to identify whether structural patterns exist in terms of speakers being advanced in groups of vocalic variables, i.e. is there evidence of certain vowels co-varying together in a theoretically meaningful way.
3. Once the resulting Principal Components (PCs) have been interpreted, we need to establish that there is no statistical relationship between the speakers who drive the components and any of the variables used in the initial modelling procedure, i.e. are the PCs driven by known sources of co-variation, such as year of birth, or are they instead representative of co-variation that exists independently of those sources?

Our data set is comprised of monophthongs from the Origins of New Zealand English corpus (ONZE) Gordon et al. (2007, 2004)). We make use of the fact that the ONZE corpus has an extensive number of tokens available for analysis, from speakers born over several generations. This provides an ideal resource for testing whether there is co-variation of phonetic variants across different speakers. In addition, ONZE's substantial time depth and well documented evidence of sound change will allow us to identify whether there are speakers who lead along the changes represented by these phonetic variants. Building on work in Sóskuthy et al. (2017), we fit separate mixed effects regression models to each vowel category in the data set. As we explain further below, the models include year of birth as one of the predictors, and so are able to shed light on co-variation among variants while controlling for diachronic changes.

## 2. Data

### *2.1. Corpus*

The data analysed in this paper come from the ONZE corpus (Gordon et al., 2007), which comprises three sub-corpora: the Mobile Unit recordings (MU), the Intermediate Archive (IA), and the Canterbury Corpus (CC). We have combined this with the Canterbury Regional Survey (D'Arcy, 2017). The ONZE corpus is unique in that it provides the longest record of transcribed NZE in existence, with speakers born over a 137 year time period (1851-1988). The speech samples were collected during interviews, with much of the speech being spontaneous. The combined corpora contain over 600 different speakers, who come from a range of social backgrounds and geographical locations within New Zealand's South Island.

### *2.2. Data processing*

The corpora are searchable through the LaBB-CAT database (Fromont & Hay, 2008), where large volumes of speech have been force-aligned at the phoneme level using the HTK-Toolkit (Young et al., 2002). We initially extracted all vowel tokens that contained any of the 12 monophthongs of NZE (DRESS, FLEECE, FOOT, GOOSE, KIT, LOT, NURSE, SCHWA, START, STRUT, THOUGHT, TRAP), querying all 636 available speakers. We then automatically extracted F1 and F2 values at the midpoint of each vowel token using Praat (Boersma & Weenink, 2018), resulting in a data set of over two million vowel tokens. In order to ensure sufficient quantity and quality from this data set, we applied a number of filtering and data processing steps. These are summarised below and in Figure 1, with a more detailed explanation included within the Supplementary Materials (including the code used at each step).

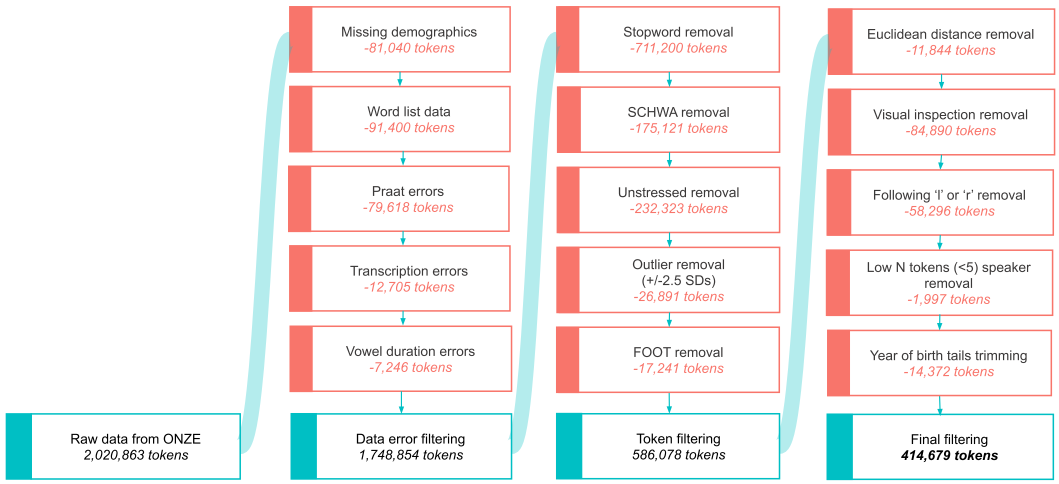


Figure 1: Steps taken to obtain the final data set used in the subsequent analyses

Firstly, we removed any speaker whose birth year or gender was unknown to us. We then removed any data that came from a word list reading style, or where Praat was unable to extract an F1/F2 measurement. As the data was force-aligned, we also ensured that all tokens had an F1 under 1000Hz, had a valid orthographic and phonological transcription and a vowel duration between 0.01 and 3.0 seconds. Next, we removed all tokens that occurred in a list of 82 stopwords (high frequency grammatical words, see the Supplementary Materials for the full list) as these words are prone to phonetic reduction (see Hay et al. (2015)). We also removed all unstressed tokens in the data set (including all tokens coded as SCHWA and tokens without any lexical stress coding in CELEX (Baayen et al., 1995). In order to remove outliers, we filtered out all tokens that had an F1 or F2 value that was $+/-2.5$ standard deviations from the mean formant value per speaker, per vowel. Upon inspecting the token counts for each speaker, we observed relatively low counts for FOOT. As the statistical modelling approach we use requires that all speakers should have sufficiently high token counts across all vowels, we decided to remove all FOOT tokens.

In order to further ensure that our data set contained reliable F1/F2 measurements, we implemented three additional filtering steps which were applied on a speaker-by-speaker basis (rather than token-by-token). Firstly, we implemented a minimal vowel-space dispersion requirement. If, for some reason, automatic formant tracking failed for most tokens for a single speaker, the vowel-based outlier-removal processes (described above) would not necessarily remove the mis-measured tokens. If the distributions of all vowel categories overlap each other, then no individual token is going to be flagged as an outlier for any specific vowel category.

In order to identify speakers with vowel spaces that do not contain distinguishable categories, we calculated the mean Euclidean distance between all of the vowel categories produced by a speaker. This was done to obtain a measurement of dispersal for each speaker. We identified 14 speakers who had particularly small values for this measurement ($< -2$ standard deviations from the population's mean Euclidean distance), which indicated that there was little variation in the location of the formant measurements across all the vowels, i.e. their vowel realisations overlapped across lexical sets. As a consequence, these speakers were removed from the data set. This led us to our second speaker filtering step, which involved manually inspecting the vowel spaces and density distributions of the F1/F2 values for each of the remaining speakers. Following this, we decided to remove a further 76 speakers from the data set, where the distribution of the vowels was clearly atypical. Finally, 13 additional speakers were removed who had year of birth values at the tails of the distribution. They were removed as we found that the distribution of speakers with the latest or earliest year of birth values was very sparse, and these speakers ended up exerting an unduly large influence on our statistical models (resulting in overfitting). Note that while exclusion of these speakers improves our models, it does not affect our main findings.

Next, we removed all tokens where the vowel was followed by /l/ (because NZE has some vowel mergers before /l/ (Thomas & Hay, 2005; Bauer et al., 2007)) or /r/ (because rhoticity has been variable over the history of NZE (Hay & Sudbury, 2005), and will affect vowel quality). Finally, we removed any speaker with fewer than 5 tokens for any of the vowels. This was done in order to ensure every speaker had sufficient data for the analyses (token counts for each speaker can be found in the Supplementary Materials). The final data set comprises 414,679 tokens across 10 monophthongs, from 481 speakers (225 female, 256 male), born between 1864–1982. A summary of the token counts per vowel can be found in Table 1.

### *2.3. Normalisation*

We normalised the formant frequencies using an adapted version of the Lobanov method (Lobanov, 1971). Lobanov's formula normalises the formant frequencies on a speaker intrinsic basis and has been widely reported and used as the optimal method for normalisation (Adank

Table 1: Number of tokens per vowel and their percentage in the final data set.

| Vowel | N Tokens | % |
|---|---|---|
| NURSE | 16,891 | 4.1 |
| START | 21,217 | 5.1 |
| GOOSE | 26,432 | 6.4 |
| THOUGHT | 28,201 | 6.8 |
| TRAP | 32,284 | 7.8 |
| LOT | 35,228 | 8.5 |
| FLEECE | 49,757 | 12.0 |
| STRUT | 50,907 | 12.3 |
| DRESS | 69,925 | 16.9 |
| KIT | 83,837 | 20.2 |
| **Total** | **414,679** | **100** |

et al., 2004). Yet, when we consider the data that we intend to normalise, a number of issues arise when applying the Lobanov method. The formula normalises by subtracting the raw formant value from the mean of all formant values, then dividing by the standard deviation of all formant values, see Equation 1 (where $i$ is an index denoting a given speaker, $\mu$ is the mean and $\sigma$ is the standard deviation). This is convenient when the data set is balanced, i.e. each speaker has roughly equal token counts per vowel, as well as roughly equal token counts per speaker.

$$F_{lobanov_i} = \frac{(F_{raw_i} - \mu_{raw_i})}{\sigma_{raw_i}} \tag{1}$$

However, our data set contains spontaneous speech from hundreds of speakers, and as a consequence, has large variability in the token counts per vowel and per speaker. When using Lobanov's formula in this situation, the normalisation procedure is prone to bias towards vowels with larger token counts, and thus instability when different speakers produce different counts across vowels (e.g. in our data set KIT is more common than NURSE, so the calculation of $\mu$ and $\sigma$ will be weighted towards vowels with higher counts). To resolve this issue, we introduce a formula that is better suited to imbalanced data sets, which we will refer to as *Lobanov 2.0*, see Equation 2. In this adapted version, the primary difference is that means are calculated per vowel category, then a mean of means (i.e. the mean value calculated across the individual vowel means) is subtracted from the raw formant value. This value is then divided by the standard deviation of the mean of means, giving the normalised formant value.[1]

$$F_{lobanov2.0_i} = \frac{(F_{raw_i} - \mu_{(\mu_{vowel_1},\cdots,\mu_{vowel_n})})}{\sigma_{(\mu_{vowel_1},\cdots,\mu_{vowel_n})}} \tag{2}$$

## 3. Analysis and Results

As outlined earlier, we adopt a structured analysis procedure to address our research questions. These are described in more detail in the subsections below. Following guidance in Roettger et al. (2019), all data and code are provided in the Supplementary Materials or accessible directly at `https://github.com/jamesbrandscience/Covariation_monophthongs_NZE_2020`. All code was written in R version 3.6.3 (R Core Team, 2018).

### *3.1. Speaker intercepts*

In order to analyse co-variation in the data set, we must first extract a measure of how each speaker's realisations of the different vocalic variables compare with those of the other speakers. The goal is to assess how deviant each speaker is for each vowel, given where we would predict them to be for their year of birth and gender, and also controlling for variability in speech rate. To obtain this measure, we exploit the random intercepts produced by mixed models, which provide estimates of speaker variability, whilst keeping other variables constant (see Drager & Hay (2012)). This approach runs counter to traditional applications of mixed-effects modelling, where fixed-effects are the main focus of interest. Instead, it is the random effect of speaker in the models that we are interested in. The fixed-effects are included primarily as variables we wish to control for.

Take for example the vocalic variable of F1 for DRESS, which has been widely reported to have undergone sound change in NZE (Maclagan & Hay, 2007). If we were interested in testing whether year of birth can predict changes in F1 DRESS, we could use a mixed model with fixed effects of *year of birth* and a random effect for *speaker*, where the fixed-effect is the variable of interest and the random effect is included to account for variation between speakers. Such

[1] See the Supplementary Materials for the implementation of this formula in R, with further discussion on the differences in normalised values produced by the two methods.

a model may tell us that there is a significant effect for year of birth, and thus evidence for a sound change. If we inspect the random intercepts for the speaker variable, we can assess how advanced each speaker is in relation to other speakers with similar year of birth values.

Intercept values from mixed models are typically normally distributed[2] as a continuous variable, with values centered around 0, making them comparable across different models. If we take the F1 DRESS model example, a positive speaker intercept indicates a deviation from other speakers born around the same time in a positive direction (which corresponds to an articulatorily lower vowel). On the other hand, a negative intercept indicates a deviation in a negative direction from other comparable speakers. In this case, this is also the direction of the sound change, as DRESS has undergone raising in NZE. Thus, a negative intercept would mean that our speaker is more advanced than others in terms of F1 DRESS. The further the intercept is from 0, the more the speaker deviates from the rest of the population. Thus, the speaker intercepts provide an estimate representing variability across speakers, but crucially keeping all other predictive variables constant. See Figure 2 for a visual demonstration of how these intercepts can identify speakers who are advanced in the NZE short front vowel shift.

An appropriate modelling technique to obtain the speaker intercepts from our data is generalised additive mixed modelling (GAMMs). This is because we want to model the normalised F1 and F2 of each of the 10 monophthongs separately, whilst also reliably capturing non-linear changes across time (see Winter & Wieling 2016, Sóskuthy 2017, Wieling 2018 and Chuang et al. 2020 for introductory tutorials on GAMMs). We fitted separate models to normalised F1 and F2 for each of the 10 vowels in the data set, giving 20 models in total. All models were fitted using the same formula via the *mgcv* package in R (Wood, 2017), with fixed-effects comprising separate smooths over year of birth for each gender (using an adaptive smooth basis with 10 knots), as well as a smooth over speech rate. Random intercepts were included for speaker and word form. We then combined all of the speaker intercepts from the separate models, providing us with a final 481 $\times$ 20 data set, where each of the 481 speakers had a separate intercept for each of the 20 vocalic variables.

### *3.2. Principal Component Analysis*

Using our data set of speaker intercepts, we now have a suitable set of estimates to explore how the vocalic variables may co-vary. How can we find interlinked variables within this data-

[2]The intercepts from the modelling procedure described below are indeed normally distributed as shown in the Supplementary Materials.

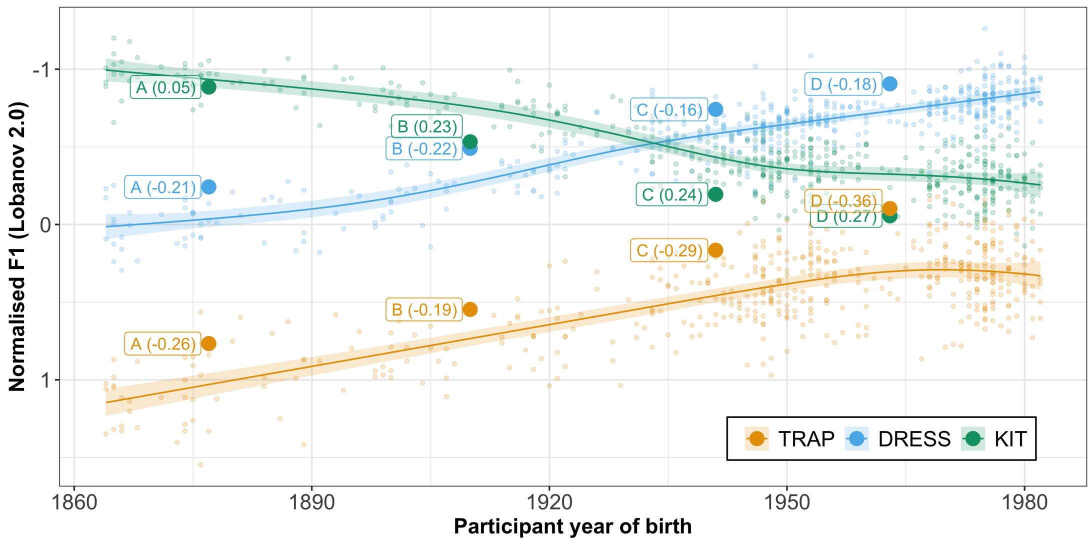


Figure 2: Visualisation of change in F1 (normalised) for TRAP, DRESS and KIT over the period of change in ONZE. GAMM smooths are added to show direction of change, whilst individual points represent each of the 481 speaker's mean F1 values for the three vowels. Four speakers (A, B, C and D) have been highlighted on the plot as they all have random intercepts (given in brackets) which indicate they are advanced in the direction of change (negative intercepts for TRAP and DRESS and positive for KIT). Typically, the larger the distance from the smoothed line, the larger the absolute value of the intercept.

set? Which subsystems of vowels appear to be co-varying? If we know how extreme a speaker is for one particular vowel-formant, can this help us predict how extreme they are for others?

Figure 3 illustrates the distribution of Pearson's correlation co-efficients (r) when all pairwise correlations in the intercepts data are calculated (red line). We can see that there are a number of correlations that appear to be quite strong, indicating that some vowel-formants may be tightly linked to one another - some positively, some negatively. Such a distribution of correlations is not likely to arise by chance (see blue lines, where the distribution of r values are given from intercepts that have been permuted 100 times), but it still remains difficult to understand which variables are forming cohesive clusters of co-variation.

An appropriate analysis technique to find structure amongst a large set of apparent correlations is Principal Component Analysis (PCA). We explicitly decided not to rely on simple correlation analyses (as represented in Figure 3) given the large number of variables we are exploring and because we had no strong prior hypotheses regarding which variables will co-vary, making inferential statistics unsuitable for our exploratory analysis. Similarly, we chose not to use clustering algorithms because although they traditionally identify groups of variables that

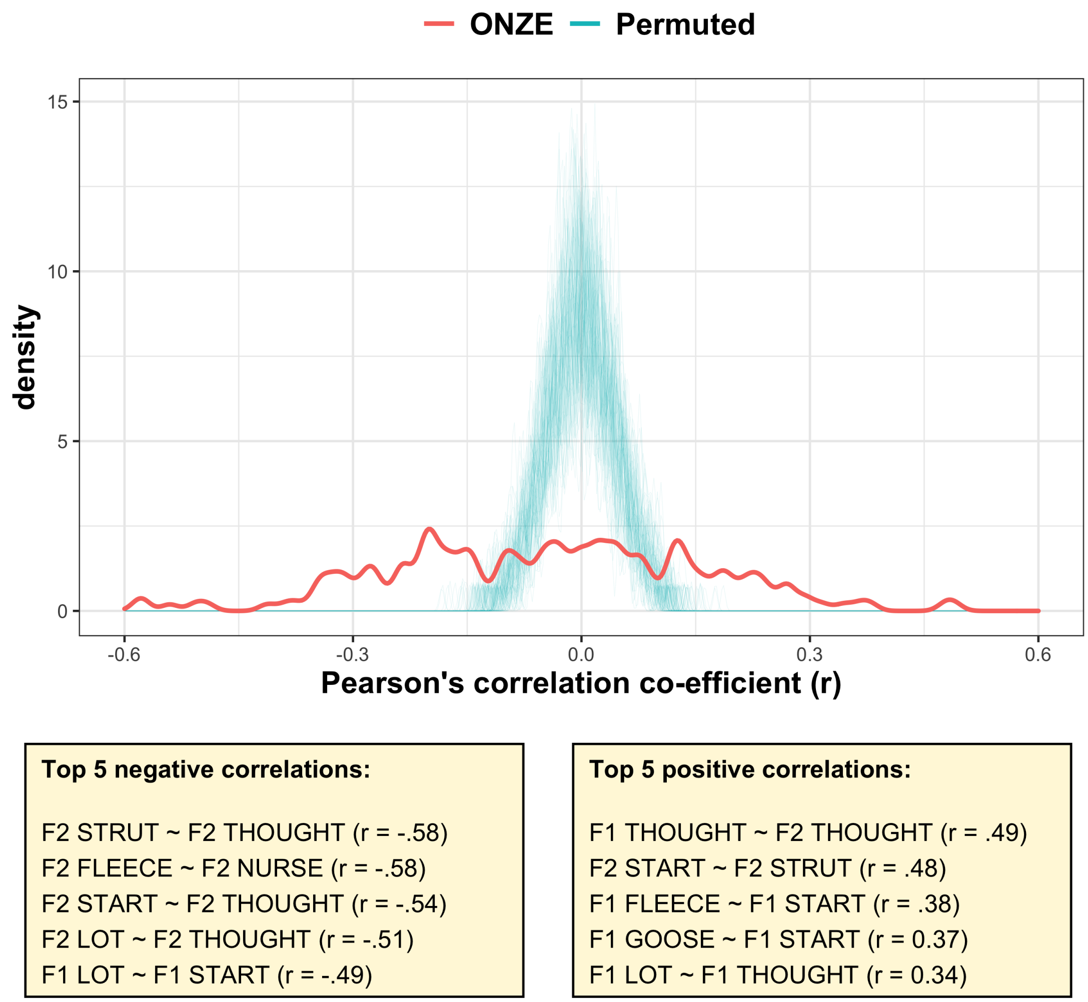


Figure 3: Distribution of Pearson correlation co-efficients (r) for the random speaker intercepts extracted from the GAMMs run on the ONZE data (red line), with the five strongest positive and negative correlations given in the boxes below. The distributions where the intercepts have been permuted 100 times are also given (blue lines), which represents what we would expect to find when any underlying structure has been removed.

cluster together, the clusters do not necessarily provide insights into the identification of leader and laggers. Moreover, they also do not offer the multilayered results that PCA does, where independent sources of co-variation can be inspected. Thus, PCA allows for a more fine grained exploration and theoretically driven interpretation of any co-variation present in the data set.

In addition to the results presented in the following sections, we also offer an interactive web application to explore the results further, accessible at `https://nzilbb.shinyapps.io/Covariation_shiny/`. This application allows users to inspect the vowel spaces of each of the 481 speakers, with their individual scores on each of the three PCs, in addition to their demographic information.

### *3.3. Patterns of Sound Change*

Before going through the patterns of co-variation revealed by the PCA, we briefly orient the reader to the primary patterns of sound change as shown by the GAMMs. While our primary purpose in fitting the GAMMs is to control for speaker factors such as year of birth, understanding the major patterns of sound change occurring over this time period is also a necessary precursor to interpreting the patterns of co-variation. In particular, in determining whether realisations of vowels which co-occur are all innovative, we need to establish the direction of major sound changes. Figure 4 shows the trajectories of vowel change, based on GAMM smooths over year of birth[3]. An animated and interactive version of the changes is available in the Shiny application.

The figure shows the well-documented changes of NZE: raising of TRAP, DRESS and centralisation of KIT, which are linked together in a chain-shift (Hay et al., 2015). The fronting of GOOSE, as well as the fronting and raising of NURSE are also well-documented (Maclagan et al., 2009, 2017). We also see significant lowering and backing of STRUT, and that the well-documented early fronting of START (Gordon et al., 2004) was followed by a period of considerable backing. Finally, we see considerable raising of THOUGHT and backing of LOT.

It is important to note that while such trajectories are useful for showing the main direction of change of individual vowels, they conceal considerable individual variation. Figure 5 shows the same trajectories in vowel-specific panels, this time including points representing the individual speakers alongside the main trend. This distribution is sometimes very broad, reflecting considerable speaker-based variation around the overall trajectory of change.

### *3.4. Interpreting the PCA*

We have seen that the random intercepts estimate the degree of deviation that individuals show from the main trends of the fixed effect GAMM smooth. We then use these intercepts

[3]Although this was an exploratory analysis, we note that all vowels have undergone some change, either linear or non-linear, based on the year of birth smooth effect (all $p$-values $< 0.001$), see the Supplementary Materials for the model summaries.

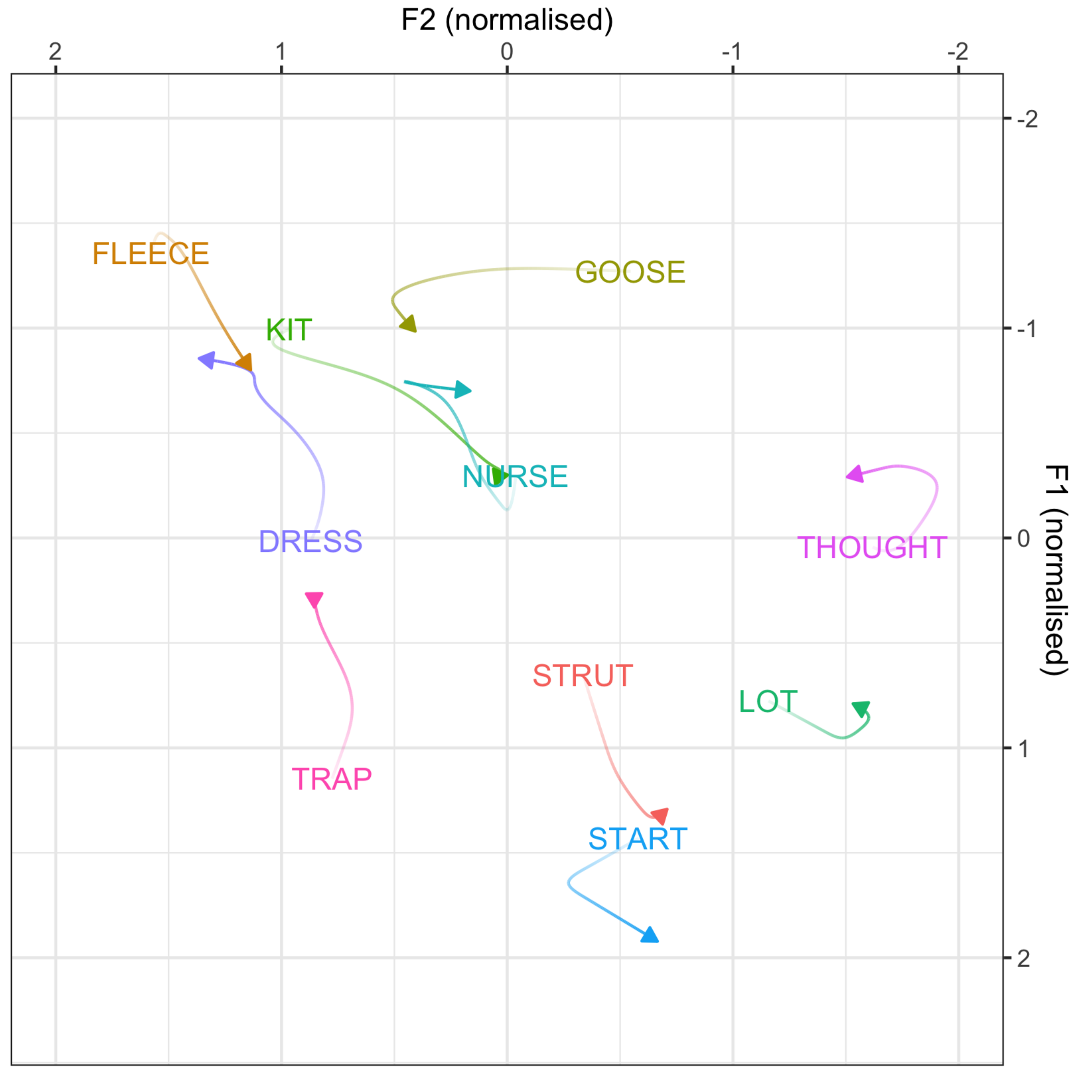


Figure 4: Trajectories of the 10 vowels based on a GAMM smooth over year of birth. The vowel label represents realisations for speakers born at the start of the ONZE corpus (1864), whereas the end point represents the most recent speakers (1982).

as the data in the PCA, allowing us to identify patterns of co-variation that exist across the vowel-formants. We interpret the three most informative principal components (PCs) returned by the PCA, referred to individually as PC1, PC2 and PC3, each of which explain $> 10\%$ of the variance in the data, accounting for 43.1% of the variance overall. Each PC will identify distinct sources of co-variation in the data set, which can be understood by assessing which vowel formants are explaining the most variation in the PC.

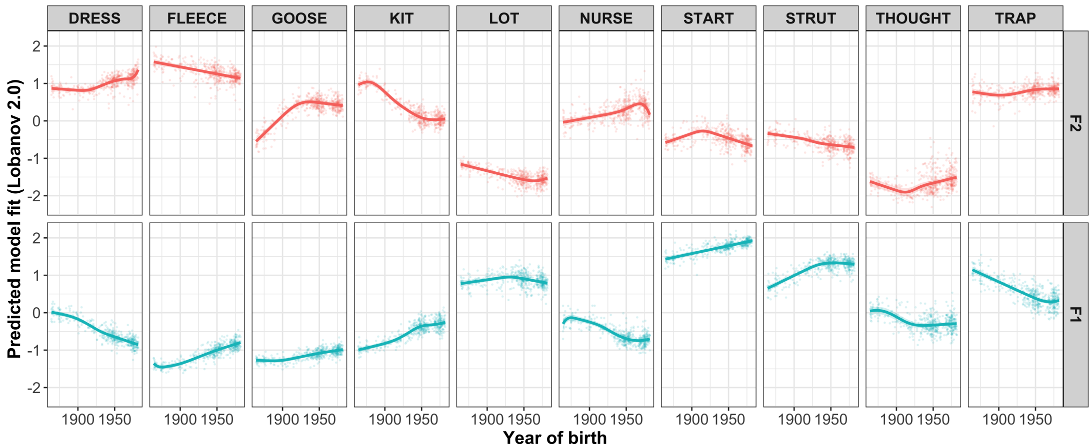


Figure 5: Trajectories of vowels based on a GAMM smooth over year of birth (as in figure 4), with each individual speaker plotted as points around the smooth (calculated by adding their random intercept values to the model fit.

We will describe for each PC:

- *Which* specific vocalic variables load meaningfully onto the PC (determined by the contributions made by each variable).
- *What* this tells us about the way certain vocalic variables are realised by different speakers in terms of the F1∼F2 space.
- *If* the PC can be explained by demographic factors such as year of birth or gender.

We first note that the signs of the PCs in a PCA solution are completely arbitrary. For instance, if there was a PC corresponding to leaders *versus* laggers in multiple changes, it could be coded such that leaders are assigned positive scores and laggers negative scores; but it could also be coded the opposite way, with leaders receiving negative scores and laggers positive scores. The two PCA solutions are mathematically equivalent.

With this in mind, we now turn to the loading values that the PCA assigns to each of the vowel-formants. These not only indicate how important each vowel-formant is to each PC, but also provide information about their directionality. The directions of the loadings can be interpreted in terms of their sign (+/−). Given the arbitrariness of the signs of the PC scores, the loading directions in themselves are not meaningful. However, the way these signs pattern together across different vowel formants is potentially meaningful. If the signs are the same for any pair of formants (e.g. both are positive) then it indicates that they are co-varying in the same direction, whereas if the signs contrast (e.g. a positive and a negative) then it indicates that the formants are co-varying in opposite directions.

Similarly, each of the speakers in the dataset will be assigned a PC score. This value indicates the extent to which a given speaker exhibits the pattern of co-variation in a given PC. Again, these can be interpreted in terms of +/− signs, which are arbitrarily assigned. Speakers with large absolute PC scores are at the margins of the co-variation. If a speaker has a large positive score, then they will represent co-variation in one direction, while a speaker with a large negative score will be co-varying in the opposite direction. It is through the combination of the PC loadings and the PC scores that we can determine whether we can observe 'leaders' and 'laggers' of sound change, or other forms of co-variation.

### *3.4.1. PC1: The restructuring back vowels*

PC1 accounted for 17.2% of the total variance in the PCA. Figure 6A shows the contributions made by each vowel-formant to the PC. To help with the interpretation of which vowel-formants are most important to the PC, we present the contributions as a contribution percentage. These are calculated by squaring the individual loadings and then multiplying them by 100, to give a value that represents how much the vowel-formant contributes to the variance explained by the PC. The vowel-formants are ordered left to right as a function of their contribution percentage. There is a clear cluster of four variables which drive this PC: THOUGHT F1 and F2, and the F2 of START and STRUT, all of which collectively contribute over 50% to the PC (i.e. are to the right of the red dashed line). Each vowel-formant has a positive or negative sign, which indicates the directional relationship of the co-variation between all of the other vowel-formants. For instance, THOUGHT F2 is negative (red '−' sign in Figure 6A), indicating that speakers with a low PC1 score will have a more backed THOUGHT realisation, i.e. their F2 values are small. Whereas, STRUT is positive (black '+' sign), indicating that the co-variation is operating in the opposite direction, i.e. speakers with a low PC1 score will have a fronter STRUT realisation.[4]

Before trying to understand what the PC is telling us, and how it might (or might not) relate to sound change, it is useful to revisit the sound changes for these particular vowels, as shown in Figure 4 and the animation available in the Supplementary Materials. As outlined in Section 3.3, all vowels are undergoing change. To what degree do the changes in these vowels look like they might be interlinked? We focus on the bottom right quadrant of the vowel space,

[4]Note again that the +/− signs are arbitrary in the sense that if all of the signs are reversed, the interpretation will remain the same.

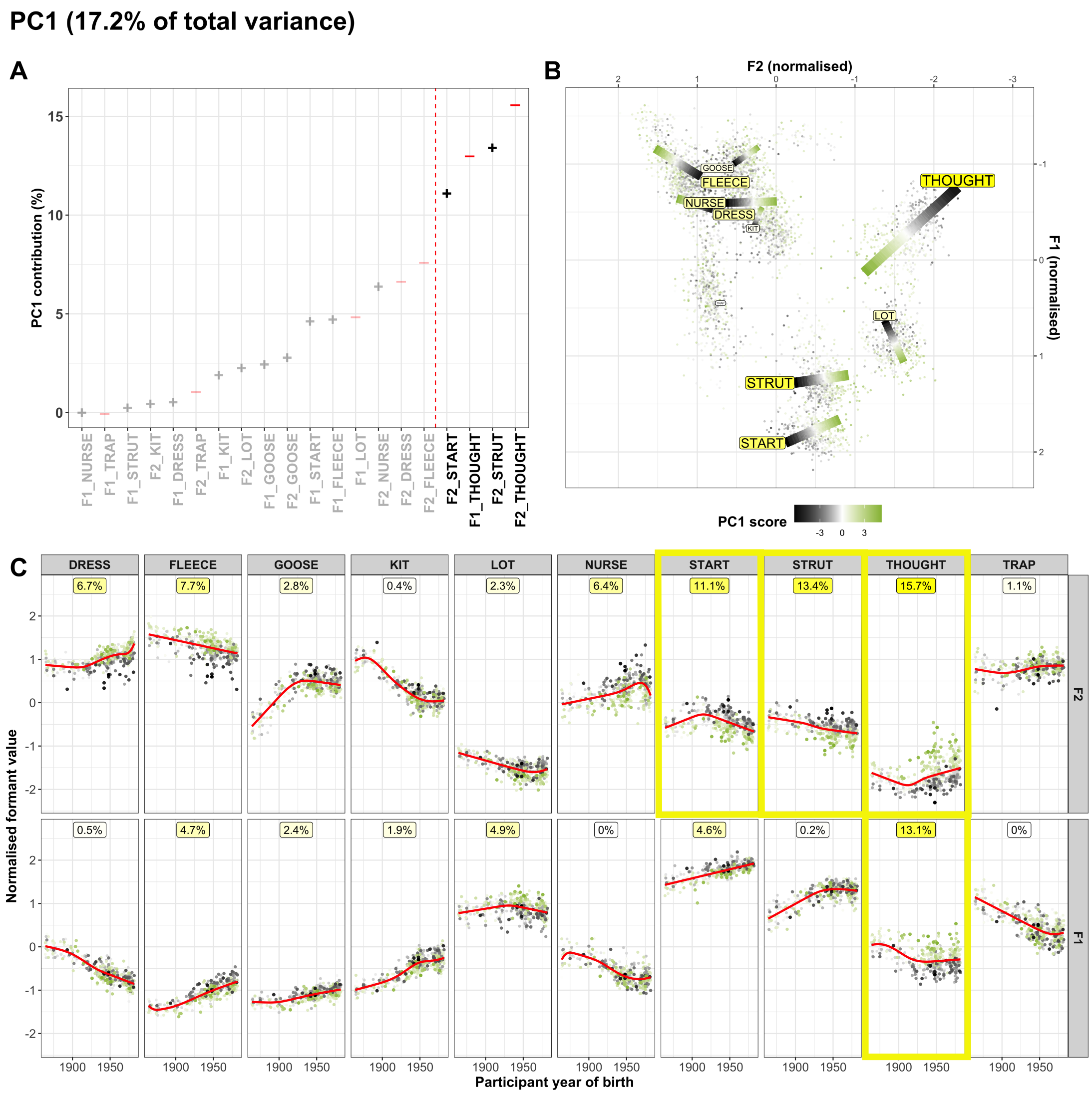


Figure 6: **A** Dot plot of the contributions of each vowel-formant from PC1, the larger the contribution the more variance explained and therefore the more important that vowel-formant is to the PC. Vowel-formants that are to the right of the red dashed line collectively contribute over 50% to the PC. The +/− signs indicate the direction of the loadings. **B** Visualisation of how the loadings can be interpreted in terms of F1∼F2 space. The location of the vowel label indicates the position of realisations made by speakers with the lowest PC1 scores, with the trajectory ending at the position where speakers with the highest scores are. The size and colour of the vowel label is indicative of the importance of the vowel to PC1. **C** The trajectories of change for each vowel-formant by year of birth (red line), with the speaker intercepts plotted as points around the trajectory. The colour of the points indicates the speaker's PC score, where green points are positive scores and black are negative. The contribution percentages for each vowel-formant are given at the top of the panels and the most important vowel-formants are highlighted by yellow boxes.

where the four back vowels start in a diamond shape, with THOUGHT and START forming the top and bottom points, and STRUT and LOT sitting adjacent to each other in the middle.

Over time, we then see a gradual rearrangement, into a 'linear' reorganisation in which START, STRUT, LOT and THOUGHT are arranged into an elongated linear pattern along the back of the vowel space. This is achieved, in F1 space, through the lowering of STRUT and START and the raising of THOUGHT. LOT initially slightly lowers, and then raises again. In terms of F2, LOT and STRUT consistently back, and we see initial fronting of START, followed by considerable backing. THOUGHT shows some initial backing, and then moves frontwards.

Note that THOUGHT and STRUT in particular, begin fairly close to each other, and then undergo quite dramatic changes in opposing directions. Impressionistically, as STRUT approaches START, these two vowels look like they become linked in terms of F2 and move back in tandem. It is also notable that both START and THOUGHT both appear to change direction in F2 movement at about the same time (between 1910 and 1920). One interpretation is that the two long vowels were initially repelled from each other (in both F1 and F2), but then once they were sufficiently separated in terms of height, they retreated back to their original front/back position. The end result of this complex series of changes is a reorganisation in which START, STRUT, LOT and THOUGHT are arranged into an elongated linear pattern along the back of the vowel space.

Holding this set of vowel changes in mind, we now return to PC1. Our focus is on trying to understand what PC1 tells us about links among these vowels within speakers. PC1 relates to the overall position of THOUGHT, and to the F2 of START and STRUT. A number of different visualisations help us interpret the PC. First, in Figure 6B, we see the trajectory from speakers with the lowest PC1 score to those with the highest PC1 score. These trajectories are the fitted values from linear regressions predicting the speaker intercepts by the speaker's PC1 scores. These are used for visualisation purposes, in order to give an impression of the degree and directionality of the loadings in terms of F1$\sim$F2 space. They illustrate the distribution of speakers from low PC1 (in black) to high PC1 (in green), and show that high PC1 speakers have backer START and STRUT vowels, and lower, fronter THOUGHT.

In Figure 6C we can see how the speaker intercepts from the sound change models relate to the speaker PC scores . When the positive and negative scores (green and black points respectively) are clearly distinguishable above and below the sound change smooth (red line), this indicates that the degree of co-variation is large (with the yellow outlines used to highlight which vowel-formants contribute most to the PC). This also aids in the assessment of whether the loadings are aligned in the same direction of sound change (separating, for example 'leaders' and 'laggers' of a constellation of changes) or whether they are pushing in somewhat separate directions. In a sound change trajectory showing a positive slope for example, 'leaders' will appear above the smooth line. If these points appear systematically green, then high PC1 speakers are systematically leading for that change.

In this visualization we see again that speakers with high PC1 scores have considerably larger F1 and F2 intercepts for THOUGHT, that is, their THOUGHT is lower and fronter in comparison to the rest of the population. Speakers with high PC1 scores have lower F2 intercepts (are backer) for START and STRUT than the rest of the population.

How does this compare to the direction of sound change? The answer is straightforward for STRUT F2 and THOUGHT F1, both of which show a consistent direction of change. Speakers with high PC1 scores are innovative for the backing of STRUT and conservative for the raising of THOUGHT. In other words, speakers who are leaders in one of these sound changes, tend to be laggers in the other.

START and THOUGHT F2 are more complex to interpret because they begin by moving away from each other (with START fronting and THOUGHT backing), and then start to move back towards each other for speakers born between 1910 and 1920. Speakers with high PC1 scores have a backer START F2, and thus would be conservative for early speakers, but innovative for later speakers. Similarly, those same speakers have fronter THOUGHT F2, and thus would be conservative for early speakers, but innovative for later speakers.

Innovation in F2 for START and THOUGHT does not mean the same thing for early and late speakers, but in both cases the vowels are linked together. For earlier speakers, innovation in START/THOUGHT F2 is associated with speakers who are less innovative in STRUT-backing and more innovative in THOUGHT-raising. For later speakers, it is associated with those speakers who are more innovative in STRUT-backing, and less innovative in THOUGHT-raising.

To summarise, as STRUT backs, THOUGHT raises. These changes are linked, such that if an individual speaker was leading in one change, they were likely to be lagging in the other. In early NZE, those who were advanced in THOUGHT raising were also backing THOUGHT, while fronting START. By around 1910-1920, THOUGHT was relatively high, and started to front again, while START appears to have become linked with THOUGHT, and backed in parallel with it. The same speakers led both THOUGHT-fronting and START-backing, and they were the same speakers who were advanced at this time in the backing of STRUT.

In sum, we have a set of three vowels that move in a complex but co-ordinated fashion

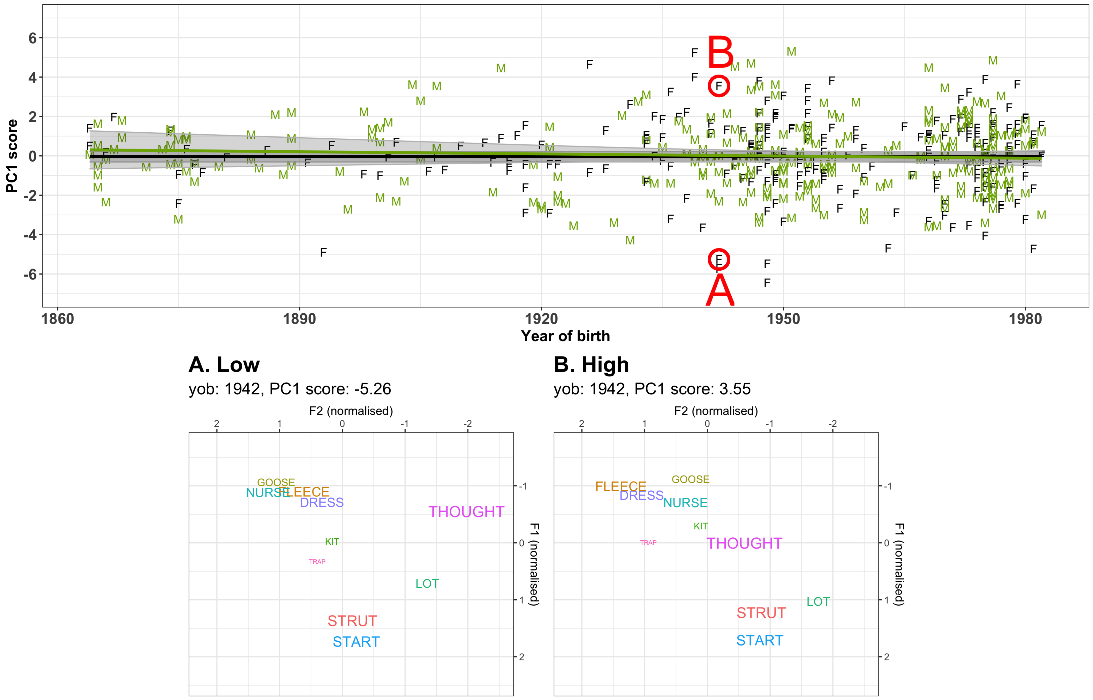


Figure 7: **Top panel** Relationship between year of birth and speaker PC scores for PC1. Each black F and green M (for female and male speakers respectively) indicates an individual speaker in the data set. Two GAM smooths are fitted (green and black lines), highlighting the lack of relationship between year of birth and PC1 score for both genders. **Bottom panel** Two speakers with similar year of births and gender have been circled to represent a negative (speaker **A**) and a positive PC score (speaker **B**), with their vowel spaces given in the bottom panel.

over the time-period. At any given time, if we know the position of one of the vowels for a speaker, we can make a reasonable estimation of the position of the other two. Vowels in the system appear to be linked both through repulsion (e.g. THOUGHT and START/STRUT for early speakers), and through potential structural association (in the case of START/STRUT for the later speakers).

Pairwise correlations (see Figure 3 and the Supplementary Materials), show that the strongest 'link' amongst the top 4 variables in PC1 are a negative correlation between STRUT and THOUGHT F2 ($r = -.58$). There is also a strong positive link between STRUT and START F2 ($r = .48$). This reinforces the interpretation that PC1 is carried by a repulsive relationship between THOUGHT and STRUT, together with a link between the front/back dimension of STRUT and START. Likewise, we observe a strong negative link between THOUGHT F1 and F2 ($r = -.54$), highlighting that when THOUGHT is lower, it is also likely to be backer.

As shown in Figure 7 (top panel), these patterns of co-variation exist across speakers even when we control for year of birth and gender. At any given time there are speakers who are ‘high PC1’ speakers, and those who are ‘low PC1’ speakers, with the PC score indicating where they sit for this set of relationships, given their year of birth. In Figure 7 (bottom panel) we also see examples of vowel spaces from speakers born at the same time - in the later part of the sound change - with low and high PC1 scores. Their vowel spaces are plotted in A (a low PC1 speaker) and B (a high PC1 speaker). In terms of the linear reorganisation of the back vowels, speaker A is advanced in the raising of THOUGHT, but still has relatively front STRUT/START, and THOUGHT has not yet fronted. Speaker B has more innovative backer STRUT/START, and THOUGHT is front, but is correspondingly less raised. Comparing across the speakers, we see that the degree of backness of STRUT/START is comparable. In the most general terms, we can say that speakers with high PC1 scores are innovative in F2 for this cluster of three vowels, and conservative in F1 (for THOUGHT) . Speakers with low PC1 scores show the opposite pattern.

#### *3.4.2. PC2: ‘leaders’ and ‘laggers’ of sound change*

The second PC accounted for 15.8% of the total variance. Figure 8A gives the contributions for each of the vowel-formants, in addition to a visualisation of how these can be interpreted in terms of F1∼F2 space (8B). We see that the highest contributor is TRAP F1. This is plotted with a black ‘+’ sign, indicating it is positively loaded onto PC2, meaning that speakers with high PC2 scores have a raised TRAP. FLEECE F1, on the other hand, is negatively loaded onto PC2, which is plotted with a red ‘− sign, indicating a negative loading. In other words, speakers with a high PC2 score have a low F1 FLEECE. While TRAP F1 is particularly strongly implicated in PC2, unlike for PC1, there is no leading cluster of variables driving the PC. We concentrate on the set of variables with the highest contribution to PC2 which cumulatively explain > 50% of co-variation (i.e. to the right of the red dashed line). Figure 8B gives a visual representation of the directionality of the vowel-formants in F1∼F2 space, as we move from speakers with low to high PC scores. Figure 8C again shows the trajectories of sound change for each of the vowel-formants, with the individual speaker intercepts plotted around the trajectory, speakers with high/low PC2 scores are plotted in green/black respectively.

In the case of PC2, there is complete alignment with sound change. We therefore interpret PC2 as highlighting co-variation in terms of vowels known to have undergone significant sound changes in NZE. If speakers are ahead in one of these changes, they are likely to be ahead in all of them. We can see that the well documented short front vowel shift is reflected in the

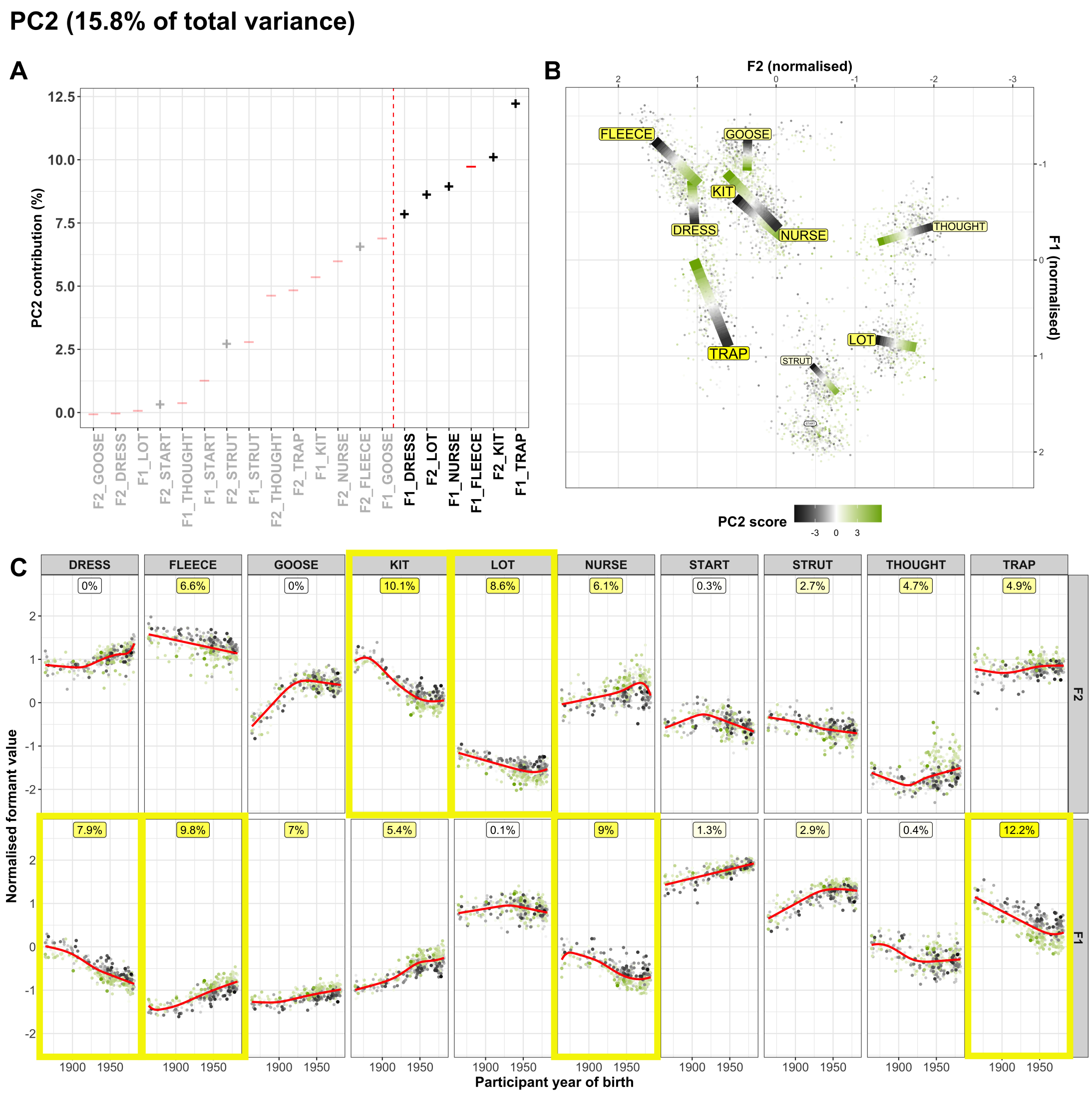


Figure 8: **A** Dot plot of the contributions of each vowel-formant from PC2, the larger the contribution the more variance explained and therefore the more important that vowel-formant is to the PC. Vowel-formants that are to the right of the red dashed line collectively contribute over 50% to the PC. The +/− signs indicate the direction of the loadings. **B** Visualisation of how the loadings can be interpreted in terms of F1∼F2 space. The location of the vowel label indicates the position of realisations made by speakers with the lowest PC2 scores, with the trajectory ending at the position where speakers with the highest scores are. The size and colour of the vowel label is indicative of the importance of the vowel to PC2. **C** The trajectories of change for each vowel-formant by year of birth (red line), with the speaker intercepts plotted as points around the trajectory. The colour of the points indicates the speaker's PC score, where green points are positive scores and black are negative. The contribution percentages for each vowel-formant are given at the top of the panels and the most important vowel-formants are highlighted by yellow boxes.

PC, with F1 for TRAP and DRESS both lowering (vowels raising), and KIT F2 lowering (vowel backing). Further to this, the PC also captures movement of NURSE and FLEECE in the direction of change over time, as well as the back vowel LOT. This indicates that there is co-variation of a specific constellation of vowels that represent differences in speakers being advanced, or behind, in the evolving sound system. We see this as identifying 'leaders' and 'laggers' of sound change – at any point in time there are speakers whose realisations are already advancing towards the direction of change. Moreover, this also shows that there are certain speakers who are in the opposite direction of change, those who are conservative in this cluster of sound changes and who are lagging behind. Indeed, if we continue down below the 50% contribution line, we find that the next five vowels are also loaded in the same direction as sound change – we need to go down to F2 THOUGHT (the 12$^{th}$ highest contributor to the PC) to find a vowel that is not aligned.

An inspection of the speaker's PC2 scores in the top panel of Figure 9 reveals that there is no significant relationship between year of birth and whether you are a 'leader' (high positive PC2 score) or a 'lagger' (large negative PC2 score), with the same being true for gender. Crucially, this shows that our GAMMs were able to control for the known predictors of sound change included in the models (specifically, year of birth and gender), giving us confidence that this modelling step has factored out such sources of known co-variation. Furthermore, the fact that there are speakers with large absolute scores (both positive and negative) across the time period demonstrates that at any point in time there will likely be speakers who are leading and lagging in sound changes, not just for a single variable, but for a sub-system of variables.

In the top panel of 9, we can see the 'leaders' and 'laggers' (those who have large positive/negative PC2 scores), but we can also see those speakers with scores close to zero, these are speakers who are exactly where we expect them to be with respect to this set of vowels given their birth year, in order words they are the speakers who are typical of the population for that point in time. Shown in the bottom panel of Figure 9 are vowel spaces from four example speakers, which illustrate the type of co-variation that the PC captures. Speakers A and D are two typical speakers born 40 years apart (both males with PC2 scores close to zero). As would be expected, their vowel spaces are substantially different. Of particular note is D's raising of NURSE, backing of LOT, and general progression in the short front vowel shift. Speakers B and C are both males who are born at the midpoint between speakers A and D, with birth dates that are very close together. C has a positive PC2 score, and so is more advanced than we would expect for the vowels implicated in PC2, so are considered a 'leader'. B has a negative

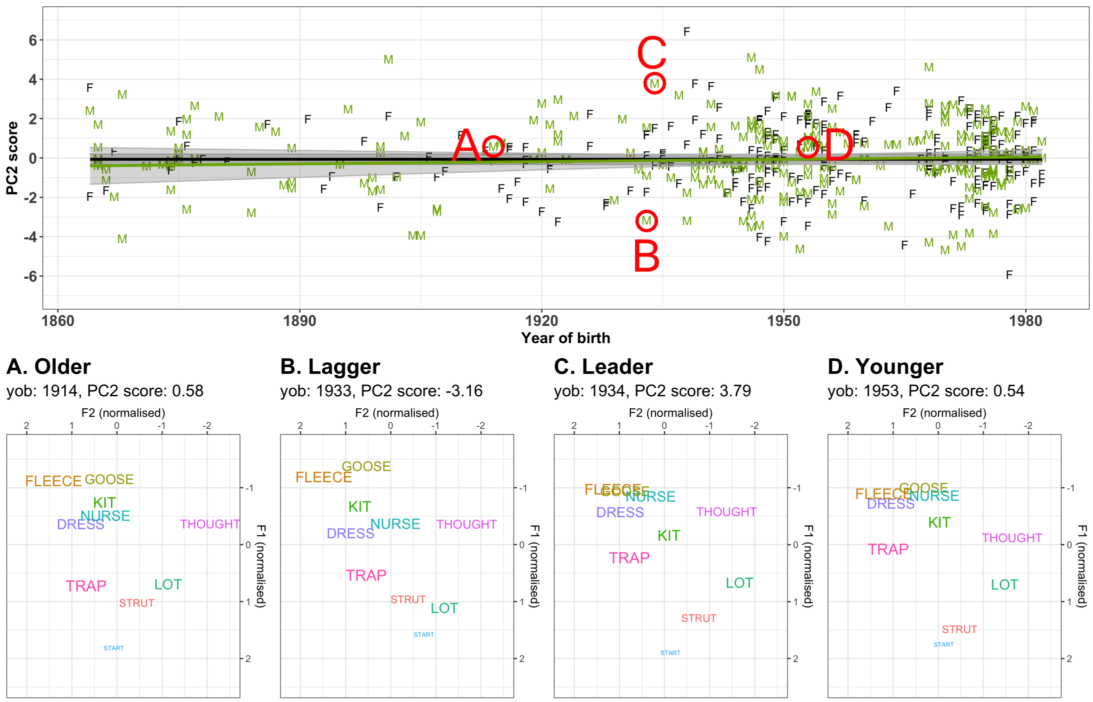


Figure 9: **Top panel** Relationship between year of birth and speaker scores for PC2. Each black F and green M (for female and male speakers respectively) indicates an individual speaker in the data set. Two GAM smooths are fitted (green and black lines), highlighting the lack of relationship between year of birth and PC2 score for both genders. **Bottom panel** Two speakers with similar years of birth and gender have been circled to represent a negative and a positive score, with their vowel spaces given in **B** and **C** respectively. Two speakers born 40 years apart have also been circled, who represent an older and younger speaker with PC2 scores close to zero for comparison, with their vowel spaces given in **A** and **D** respectively.

PC2 score, and so is less advanced than we would expect, so are considered to be a 'lagger'. Visual comparison of the vowel spaces reveals that, although speakers B and C share similar year of births, the 'lagger' looks a lot like the older speaker (A), and the 'leader' looks a lot like the younger speaker (D). In other words, moving from a speaker with a low PC2 score to a speaker with a high PC2 score appears to be roughly equivalent to moving between speakers (with a PC2 score of ∼0) who are born 40 years apart.[5]

[5] Note that comparing speakers with high and low PC2 scores to speakers born earlier and later is relatively straightforward for this PC, because all of the vowels are loaded in a direction that is aligned with the directions of ongoing sound change. It would be more difficult for the other PCs, neither of which show this complete alignment with ongoing sound changes. The online Shiny application allows exploration of the vowel spaces for all speakers in the data set for the three main PCs.

### 3.4.3. PC3: Adjacent vowel interactions at the front and the back

The third PC accounted for 10.1% of the total variance. Figure 10A gives the contributions for each of the vowel-formants, in addition to a visualisation of how the PC can be interpreted in terms of the F1∼F2 space (Figure 10B). There are four variables which appear to carry this co-variation: LOT and START F1 and - to a slightly lesser degree - DRESS and GOOSE F2.

Inspecting Figure 4 and/or the sound change animation, we are reminded that LOT lowers and then raises, whereas START consistently lowers. In terms of DRESS and GOOSE F2, GOOSE fronts dramatically and DRESS fronts slightly. The fronting of GOOSE is most vigorous amongst the earlier born speakers, whereas the fronting of DRESS is more vigorous in the later born speakers.

If we look at the pairwise correlations (shown numerically in the Supplementary Materials), the strongest correlation amongst these vowels is a negative correlation between START F1 and LOT F1 ($r = -.49$), whilst there is also a relatively strong correlation between GOOSE F2 and DRESS F2 ($r = -.37$). That is, this PC is driven by two particularly strong relationships: one at the front of the vowel space and one at the back.

The relationship with sound change can be seen in Figure 10C, where we see where speakers with high/low PC3 scores are positioned in relation to the sound change trajectories. From this we see that speakers with high PC3 scores tend to have more innovative START F1. LOT F1 doesn't move in a consistent direction; thus speakers with high PC3 scores would vary in terms of innovativeness/conservativeness depending on their year of birth - earliest born speakers are conservative, but later born speakers are innovative.. Whilst there is very little change in LOT F1 over time, there is considerably more variation in the PC scores, suggesting that – if this is linked to sound change – the link is more likely to be via changes in START than LOT, as START lowers in a linear direction over time. Speakers with higher PC3 scores have an overall more innovative configuration, with these two vowels being more separated - a lower START and a higher LOT.

A reasonable interpretation is that, in the back vowel space, PC3 captures the linear re-organisation of the space – speakers with low PC3 scores have LOT and START close together, speakers with high PC3 scores have them far apart. This interpretation is reinforced by the inspection of slightly lower contributors to PC3 - we see that THOUGHT and LOT are loaded in the same direction (Figure 10A). Speakers with high PC3 scores are overall more innovative with the back vowels. Recall that in PC1, we saw a repulsive force between THOUGHT F1/F2 and START/STRUT F2, with individual speakers either having THOUGHT raised and backed,

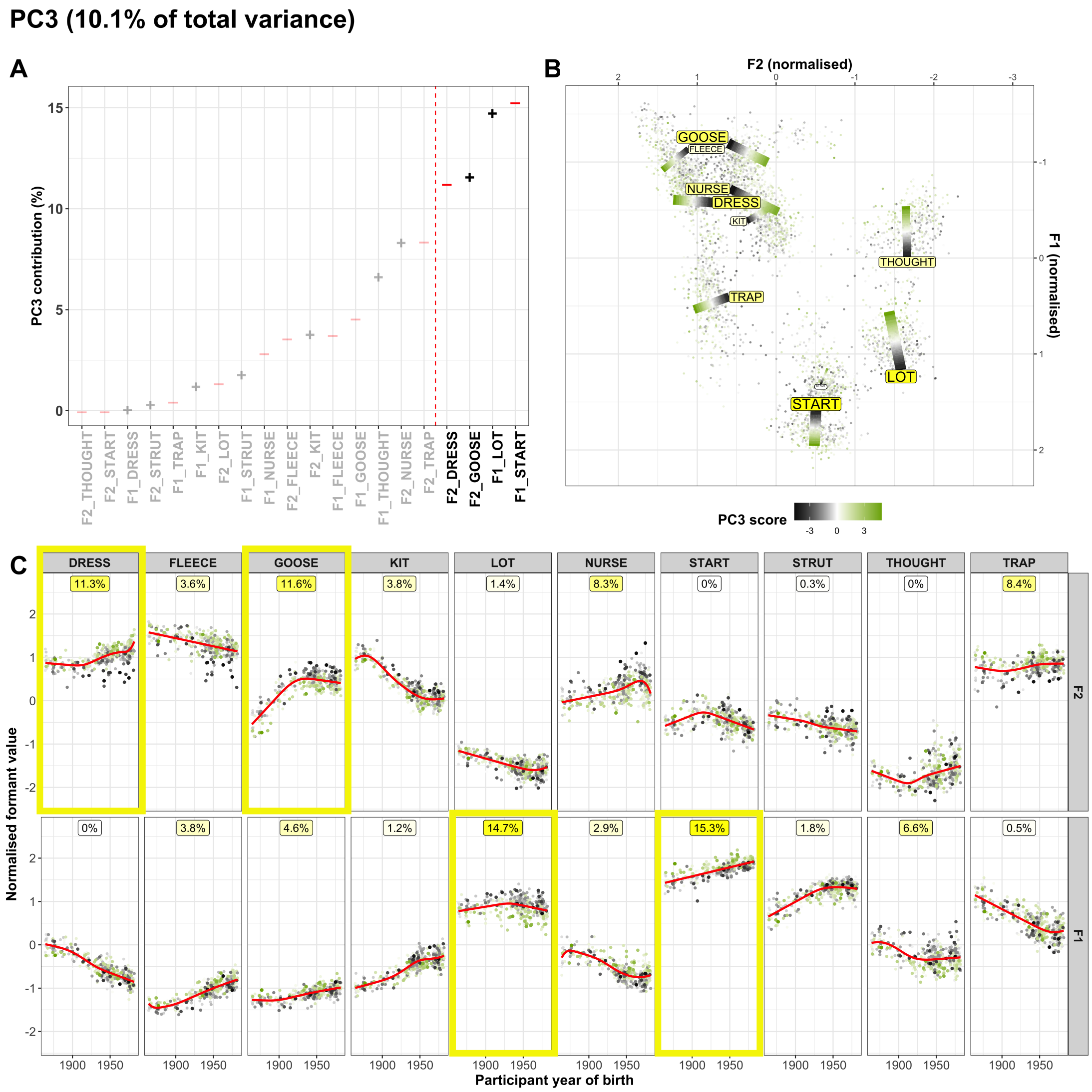


Figure 10: **A** Dot plot of the contributions of each vowel-formant from PC3, the larger the contribution the more variance explained and therefore the more important that vowel-formant is to the PC. Vowel-formants that are to the right of the red dashed line collectively contribute over 50% to the PC. The +/− signs indicate the direction of the loadings. **B** Visualisation of how the loadings can be interpreted in terms of F1∼F2 space. The location of the vowel label indicates the position of realisations made by speakers with the lowest PC3 scores, with the trajectory ending at the position where speakers with the highest scores are. The size and colour of the vowel label is indicative of the importance of the vowel to PC3. **C** The trajectories of change for each vowel-formant by year of birth (red line), with the speaker intercepts plotted as points around the trajectory. The colour of the points indicates the speaker's PC score, where green points are positive scores and black are negative. The contribution percentages for each vowel-formant are given at the top of the panels and the most important vowel-formants are highlighted by yellow boxes and the most important vowel-formants are highlighted by yellow boxes.

or START/STRUT backed. In PC3, by contrast we see speakers with high PC3 scores being more advanced in the separation between LOT/THOUGHT F1 and START F1. As LOT (and to a lesser degree THOUGHT) raise, START lowers. PC3 separates speakers who are more *versus* less advanced in this process.

Moving to the front vowels, we see that GOOSE and DRESS F2 also contribute to the PC. Speakers with high PC3 scores have a backer GOOSE, indicating that they are conservative in relation to the direction of sound change. This trend is also seen in NURSE, which although it contributes slightly less than GOOSE, also fronts over time, but speakers with high PC3 scores produce more backed variants. In contrast to this, we see that those same speakers with high PC3 scores have fronter DRESS. Over time DRESS F2 is fairly stable for earlier born speakers, and then fronts. Construed in terms of sound change, speakers with high PC3 scores are more innovative in the fronting of DRESS. However, the change in DRESS over time is a much smaller effect than the change captured by the PC scores (i.e. there is much more variation in terms of the scores than there is from changes predicted by year of birth), suggesting that there may be some co-variation occurring here that is not solely due to the fact that DRESS is undergoing change. Indeed, looking further down the vowel-formant contributions, we see TRAP loaded in the same direction, and it shows even less movement than DRESS in F2 over time, but similar variation captured by PC3, with speakers who have high PC3 scores having fronter TRAP. Thus, the main patterns observed in the front vowels for this PC are that speakers with backer GOOSE/NURSE have fronter DRESS/TRAP and speakers with fronter GOOSE/NURSE have backer DRESS/TRAP.

When we inspect individual vowel spaces, speakers with a high PC3 score uniformly have DRESS fronter than GOOSE. Many speakers with a low PC3 score actually have GOOSE fronter than DRESS (see the Shiny app to explore these trends). It appears that there is only room for one set of very front vowels. If DRESS (and TRAP) are already occupying that space, there are limits to how far forward GOOSE (and NURSE) can move. But for speakers with somewhat backer DRESS/TRAP (either because they are conservative in the mildly fronting sound change, or because their DRESS/TRAP vowels just happen to be a bit backer than the average population) there is an opportunity for GOOSE and NURSE to overtake them, and move to be the vowels at the very front edge of the vowel space.

Indeed, although the overall trajectories of GOOSE and NURSE show fronting, followed by very recent backing, consideration of the distribution around this trajectory (Figure 10C) shows that, even as these vowels back, a number of speakers maintain very front vowels. Even when

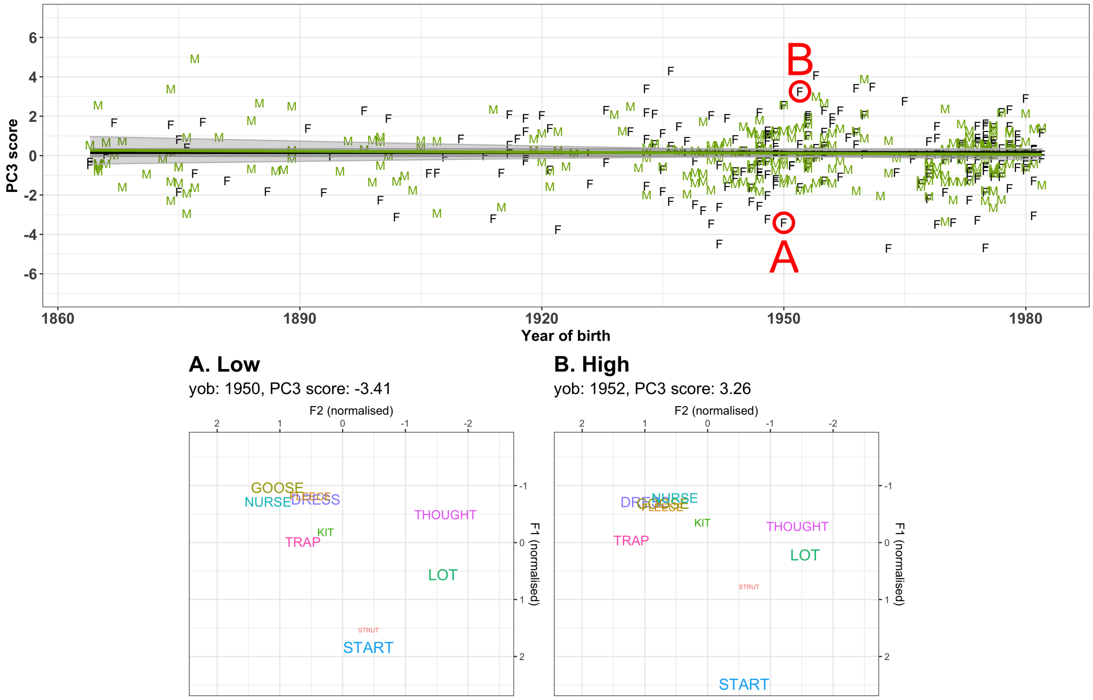


Figure 11: **Top panel** Relationship between year of birth and speaker scores for PC3. Each black F and green M (for female and male speakers respectively) indicates an individual speaker in the data set. Two GAM smooths are fitted (green and black lines), highlighting the lack of relationship between year of birth and PC3 score for both genders. **Bottom panel** Two speakers with similar year of births and gender have been circled to represent a low and a high PC score, with their vowel spaces given in **A** and **B** respectively.

we have a trajectory of lowering of F2, there are still speakers who maintain a very high F2. These are speakers with a low PC3.

Figure 11 shows two representative speakers with scores in opposite directions. The speaker with a high PC3 score (speaker B) has DRESS (and TRAP) as the front-most mid-vowels and GOOSE (and NURSE) as more central. For the speaker with a low PC3 score (speaker A), these vowels swap positions in terms of frontness - GOOSE and NURSE are the fronter vowels, and DRESS and TRAP are somewhat more centralised. At the same time, speaker A (with the fronted GOOSE/NURSE system) also has higher START and lower LOT (i.e. they are more conservative in the position of START/LOT). This is also true – to a lesser degree – for THOUGHT. Whereas Speaker B has START and LOT distinctly far apart in terms of F1.

In sum, PC3 tells us that, at any given time, those who are most advanced in the overall fronting of GOOSE (and NURSE), are those whose DRESS (and TRAP) vowels are slightly more central (and thus more conservative). In other words, there is only space at the front of the vowel space for one pair of extreme front vowels. The speakers who are most advanced in the fronting of GOOSE and NURSE are most conservative in the elongation of the back vowels along the F1 space.

This may suggest that these two subsystems carry opposing social meaning, and/or that innovative speakers aren't necessarily 'leaders' across the whole vowel space, but different groups of speakers may concentrate their innovation in different subsystems of the vowel space.

### *3.5. Other pairwise correlations*

Finally, it is worth considering that the PCA is most suited to identifying patterns of co-variation that are carried by multiple co-varying vowels. Any relationship between just two vowels, even if very strong, is unlikely to outrank a set of correlations carried across many vowels. While most strong pairwise correlations are captured in the above analysis, one of the very strongest is not. There is a strong negative relationship between F2 FLEECE and F2 NURSE ($r = -.58$). Speakers with backer FLEECE have fronter NURSE, and vice-versa. This relationship does not appear above the 50% threshold that we have used for interpreting any of the PCs, but both vowel's F2 feature somewhat below the 50% threshold loaded in opposite directions in all three PCs. This is a strong negative correlation, and our interpretation is the same as the negative relationship we saw between DRESS and NURSE F2 in PC3. There is only space for one set of very front vowels. In each of Figures 7 and 11 for example, we see two vowel spaces, one with NURSE the fronter vowel and one with FLEECE the fronter vowel.

## 4. Discussion

The PCA has revealed three distinct clusters of co-varying vowels. It is important to note that these do not exhaust the sets of inter-relationships in the vowel space. In particular, if there are two vowels that are very tightly linked, but work independently of all other vowels, this relationship is unlikely to have surfaced in the top 3 components.

PC1 is related to a rearrangement of the back vowels, showing that individual speakers may have been innovative either in F1 for THOUGHT or in F2 for STRUT/START/THOUGHT, but they are unlikely to have been innovative in both at the same time. PC2 captures a set of vigorous vowel changes in NZE, including the NZE short front vowel shift. Speakers who lead in one of these changes are likely to be leading in all of them. The front vowels predominate in this cluster of vowel changes, but it is not exclusive to them – the back vowel LOT, for example, is also involved. PC3 distinguishes between leaders in the elongation of the back vowel space, and in the fronting of GOOSE (and to a lesser extent, NURSE in the front vowel space). For these vowels, speakers are either innovative at the back, or the front, but not both. The degree of GOOSE fronting is also mediated by the frontness of DRESS. For speakers with a backer DRESS vowel, GOOSE is able to move further forward, taking up a position as one of the frontest vowels in the vowel space. Such a relationship, between a vigorously fronting vowel, and an already front vowel, is also seen in the strong pairwise correlation between NURSE and FLEECE. NURSE fronts and FLEECE backs – speakers with backer FLEECE have fronter NURSE and vice versa.

For ease of reference, the vowels linked together by the different PCs are illustrated in Figure 12. From this visualisation, it is easy to see that all vowels are included in at least one of the PCs, illustrating that there is no vowel whose production is independent of all the others. There is clear evidence, then, that the vowels are not independent of each other, and that studying change in one vowel, independent of all others, may well miss an important part of the overall picture.

We now return to the literature summarised in Section 1, to consider how our results from NZE may be able to offer answers to the outstanding questions relating to phonetic co-variation.

### *4.1. Known Sources of Vowel Co-variation*

First, it is useful to note that none of the three PCs look like they are related to different overall articulatory biases, or normalisation failures. There is no pattern of strong cross-vowel co-variation in a particular direction that is limited to a particular part of the vowel space, for example, and we see no evidence of co-variation reflecting underlyingly different vowel space sizes. This suggests that the normalisation procedure we used achieved its goal.

Second, as was shown in Figure 4, all sounds have changed over the time period we are studying (albeit some in a non-linear way). Exploration of the supplementary Shiny app will quickly reveal that the vowel spaces of earlier born speakers are different from those of later born speakers. By virtue of the fact that the vowels are changing over this time period, therefore, they are all inherently correlated across speakers. However, our model fitting procedure attempts to control for this type of year-of-birth-mediated co-variation, and this control appears to have been successful. In Figures 7, 9 and 11 we see that the PCs we analyse capture co-variation between vowels, but crucially, this co-variation is independent of year of birth (and indeed gender). We are therefore finding that there is correlational structure across speakers' vowel productions that is not simply due to the fact that the vowels are all undergoing change. This

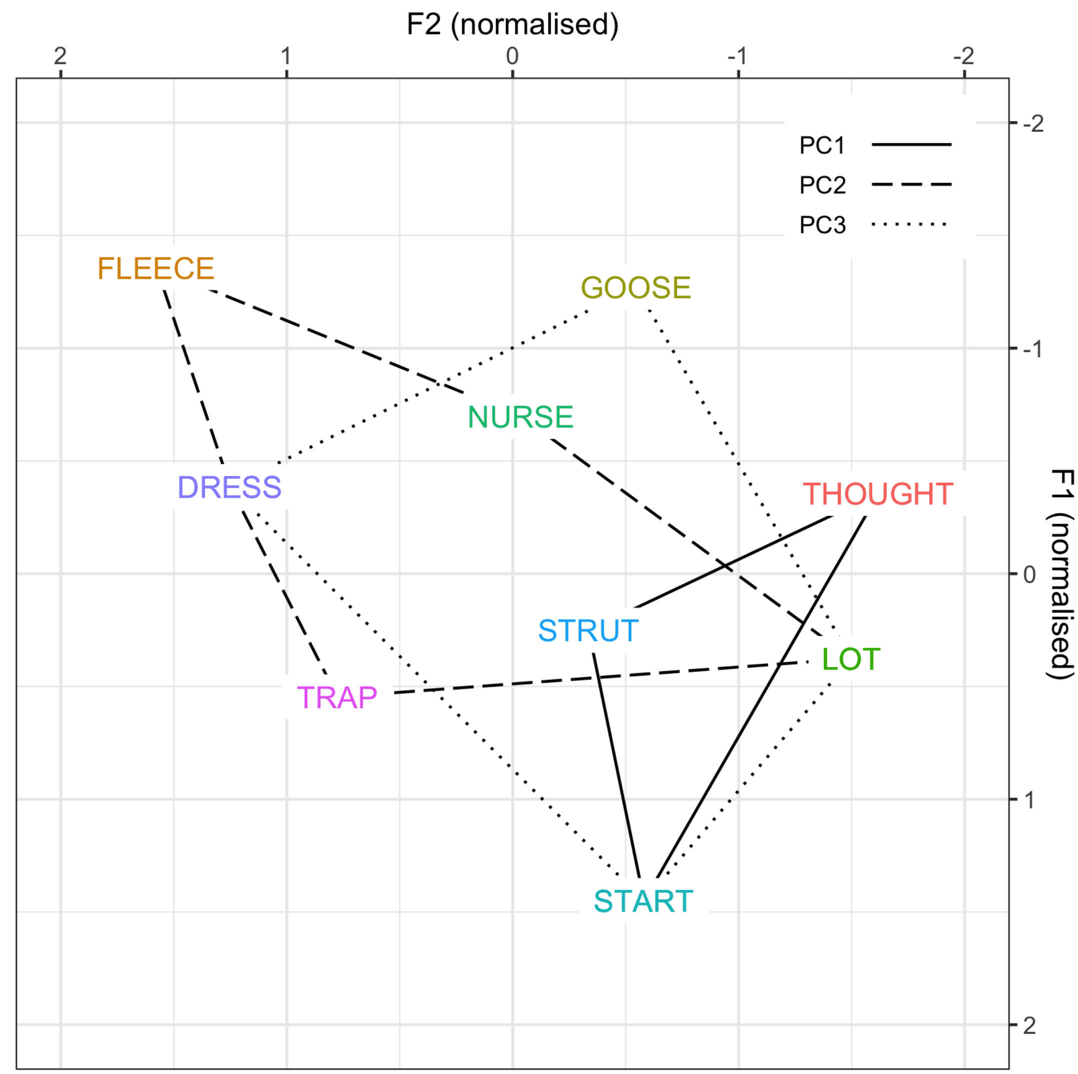


Figure 12: Simple schematic showing vowel space (from an older speaker), illustrating the vowels linked together by PC1, PC2 and PC3

methodology, then, considerably broadens the data-sets that can be used to study patterns of co-variation: we are no longer limited to homogeneous data-sets of speakers of the same age
830 and social background (Tamminga, 2019), as long as these factors can be controlled for by a mixed-effects modelling approach.

*4.2. Hypothesised Causes of Vowel Co-variation*

*Chain-shifting relationships*

The literature on chain-shifting posits that there can be tight pair-wise relationships between adjacent vowels, such that changes in one of the vowels can be causally responsible for changes in others (Labov, 1994; Gordon, 2002). In NZE, there is a reported chain-shift involving (at least) trap, dress and kit (Hay et al., 2015). These are all captured by PC2, which is aligned with these ongoing changes. This could potentially be taken as evidence that the chain-shift exists within individuals – if a speaker is a bit advanced in the raising of trap, for example, this pushes on their dress, and so on, leading them to be more advanced in the entire chain-shift. In this case, the chain-shifting mechanism could be construed to exist within individuals, and not just at a societal level. However, a number of vowels are loaded onto PC2, and there is no strong evidence that the chain-shifting vowels are more closely interlinked than the other vowels in this PC, lot and nurse. It is therefore not possible to tell whether chain-shifting has a causal effect on these vowels that is separate from the social motivations outlined in the next section.

PC1 involves the expansion of vowels at the back into a linear organisation, with strut and start lowering and (eventually) backing, and thought raising and (eventually) fronting. From PC1, we see that individuals were advanced either in the backing of strut or in the raising or thought, but not both. This is consistent with the interpretation that these vowels are repelled from one another in a pushing relationship. For some speakers, thought moves the quickest, removing the pressure on strut. For some other speakers the backing of strut is more vigorous, and thought is slower to move. However, the PC suggests that there are few speakers who do not maintain the distance between these two sets of vowels through moving one or the other, which supports a chain-shifting interpretation.

While this is a plausible account, it cannot be fully separated from an account in which these adjacent vowels simply carry opposing social meaning and/or are being led by different groups. PC3 certainly contains some vowels with opposite loadings that are more distant from each other, indicating that the existence of multiple social forces is a distinct possibility. We discuss these in the next section.

*Socially-driven parallel changes*

We are now in a position to return to Guy & Hinskens (2016)'s question about whether innovative variants tend to be used together by leaders or if different sets of changes are led by different speakers.

The loadings on PC2 provide evidence that there are identifiable leaders/laggers for sub-sets of vowels. This, however, does not imply that *all* vowels need be involved. Over 50% of the co-variation captured by PC2 involves movements in just 6 co-varying vowels. While there are more that appear likely implicated, below the 50% threshold, there are also others that are not loaded onto PC2 at all.

Indeed, the full set of results provides clear evidence that if you are a leader in one set of vowel changes, this does not imply you are a leader in all others. In both PC1 and PC3 we see evidence that there are sets of vowels, such that if speakers are conservative for one they tend to be innovative for another (e.g. PC3 – lowering of START but fronting for GOOSE). This may mean that speakers concentrate their innovation in certain parts of the vowel space, or it may mean that there are different types of 'leaders', in that different sound changes also carry different social meanings. The proposal that several variables can cluster together because their speakers share social motivations or personas is consistent with the conclusions of smaller correlational studies examining this question (Becker, 2016; Tamminga, 2019).

Our results demonstrate that there can indeed be leaders who use innovative variants for a substantial sub-set of phonetic variables (as shown by PC2). However, they also provide insights into how a speaker may exhibit opposing patterns, whereby the use of an innovative variant for one (or a number of) vowel(s) implies the use of a conservative variant for another (as shown by PC1 and PC3). Previous literature investigating pairwise correlations between limited sets of phonological variables has reported positive correlations, but not, to our knowledge, negative ones. This again highlights the broad utility of our methodology to uncover such patterns of co-variation, whether they function in an aligned or indeed, an opposing manner.

*Shared Stylistic Meaning*

The question of the stylistic meaning of variants is interwined with the question of leaders and laggers in the previous section. In our data, all of the vowels are undergoing some change, so it is not possible to disentangle co-varying changing vowels from stylistic co-variation that is unrelated to change. Certainly the data generally supports the prediction from Eckert's work that there should be broad patterns of correlations across sets of variables, and that variables do not work in isolation of one another. To definitively link this to different styles and speaker personas, we would need close and detailed stylistic analysis of the co-varying vowels in context.

*Phonologically-driven parallel changes*

Is there evidence for the shifting implementation of entire phonological classes – sets of vowels that can be referred to by phonological features?

There is absolutely no evidence within this data on co-variation that can be linked to the phonetic implementation of a single feature (such as ‘high’ or ‘back’). The sound change data does not show any such coordinated movement over time, and the PCs also do not reveal any patterns of this kind. It is always possible that this co-variation may have existed, and could have been removed through the normalisation process, which scales the edges of the vowel spaces to cover the same area. But the socially driven parallel changes described above are certainly not explainable as artifacts of phonologically driven parallel changes.

We do, however, see at least one pair of adjacent vowels which appear to move quite closely together. For earlier born speakers, START fronts while STRUT lowers. Around 1920, STRUT is quite close to START and it visually looks as if they become linked, and undergo backing together (see Figure 6). The F2 of both vowels is linked together on PC1, and there is a direct pairwise correlation of $r = .48$ – the strongest positive pairwise correlation between vowels. This is consistent with the interpretation that the primary difference between the two vowels in contemporary NZE is a quantity difference (and associated peripherality effects; see Warren 2006; Bauer & Warren 2008; Warren 2018).

Some other adjacent pairs with reasonably high pairwise positive correlations are LOT and THOUGHT F1 ($r = .34$) and TRAP and DRESS F1 ($r = .30$). Unlike START and STRUT F2, these could also be explained by a chain-shifting relationship (as one is moving in the direction of the other) and we can also not eliminate the explanation that all three pairs are simply linked through social factors. Thus, while there is no evidence that phonetic implementation of wholesale phonological features is responsible for our observed patterns, we do see some co-variation that is consistent with (but not conclusive of), parallel forces acting on some pairs of adjacent vowels.

*Evolutionary pressures on the vowel space*

Finally, it is not out of the question that some of the co-varying patterns result from overall evolutionary pressure. Indeed chain-shifting itself arises from a pressure that maximises dispersion and contrast (Labov, 1994; Gordon, 2002; Hay et al., 2015). The interlinked vowels at the back, which move to a linear organisation, are likely influenced by such forces, and are related to each other both through PC1 and PC3.

*4.3. Overall Summary*

Our methodology presents a way of controlling for known overarching influences on production patterns, and identifying subsystems of vowel productions – vowels whose productions are not independent of each other. Using these methods, we find that all vowels work together as part of a complex system. There is no vowel which is not strongly loaded onto at least one of the three PCs we discuss. That there are three separate components shows that there is some structure to the co-variation – each vowel is more tightly linked to some subset of vowels than to others.

The most compelling evidence from our results addresses the question of whether there are overall 'leaders' and 'laggers' in sound change. Our analysis provides evidence that sound changes that are not independent of each other are also not necessarily linked structurally. For some vowels, if we know that a speaker is advanced in that vowel change, it follows that they are also likely to lead in a constellation of other, possibly unrelated, vowel changes. It is not the case, however, that being a leader in one vowel change implies that they are a leader in all vowel changes.

Some local patterns of co-variation may be related to chain-shifting, or repulsive relationships, and some may link two vowels by featural similarity, but it is difficult to state conclusively that such co-variation relationships are not purely the result of shared social meanings. A socio-stylistic explanation is needed for at least some of the relationships, and thus could explain all of them. To confirm whether all the observed patterns of co-variation are related to underlying social meaning, we would need close stylistic analyses that reveal the social meanings of these variables when observed in usage.

Although we cannot definitively confirm this with our current data, it seems unlikely that linked social meanings between vowels arise entirely accidentally. While not all the observed co-variation can be explained by them, there are many hints of structural relationships in our data. If independent reasons exist for vowels to be linked – whether through chain-shifting (TRAP/DRESS/KIT), a mutual pushing relationship (STRUT/THOUGHT), closely parallel productions (STRUT/START), or jostling for the edge of the vowel space (GOOSE/DRESS) – and this leads them to co-vary within the same speakers, this likely provides the seeds for similar social meanings to become associated with correlated variants of the vowels. Most or all of the proposed mechanisms underlying co-variation patterns may exist, and conspire to create strong patterns of co-variation: sub-systems of co-varying vowels.

Finally, studying the NZE monophthongs as a system has revealed several interesting relationships and sound changes within the time-course of our data set that have not previously received much attention. First, while much previous research has focused on the short front vowels, we present here new insights into how there has also been a considerable rearrangement of the back vowels. The rearrangement of this system has perhaps been less apparent in research which analysed vowels independently of each other, before arriving to their contemporary arrangement, we can see a number of non-linear changes (most notably in START and THOUGHT F2). The PCA results clearly reveal that these movements are interlinked. Second, while the fronting of GOOSE and NURSE has been widely commented upon, the consequences of them potentially colliding with the front vowels (FLEECE, DRESS, TRAP) have not. The sound change trajectories (cf. Figure 4), show that toward the end of the time period, GOOSE and NURSE start to back again, while DRESS fronts. However, PC3 (and the pairwise correlation between FLEECE and NURSE) reveals that for some speakers, the fronting vowels are able to 'overtake' the previous frontest vowels, and become the frontest in the vowel space (see speakers with low PC3 in the Shiny app, and the examples in Figure 11). The degree to which this is possible interacts with how front the speaker's FLEECE, DRESS and TRAP vowels are. There is thus some jostling for the frontest part of the vowel space, which would not be apparent if each vowel's trajectories were analysed independently of all the others.

## 5. Conclusions

The degree to which sociolinguistic variants, or vowel productions operate independently of each other is a topic of current debate, but is methodologically challenging. We present a novel approach that controls for known predictors of co-variation, and finds clusters of variables that are linked systematically. Applying this to a data-set of NZE which has considerable time-depth, we find strong evidence of patterns of co-variation, which leads us to the conclusion that no vowels in our data operate independently of each other.

Some observed relationships are predicted from literature that has already proposed structural relationships (the NZ short front vowel shift). And some seem to suggest previously overlooked structural relationships, including a linear reorganisation of the back vowels, driven by a repulsive relationship between THOUGHT and STRUT/START; and jostling amongst the high front vowels for the frontest part of the vowel space.

Not all observed relationships can be explained through such structural relationships though, and indeed we see strong evidence that vowels can be linked together socially – for example in PC2, where we have innovative versions of many vowels clustering within speakers ('leaders' of these changes). Such social co-variation between vowels does not always go in the same direction, as in both PC1 and PC3, where innovation in one set of vowels is linked with conservatism in the other. Once we accept that a likely explanation for some co-variation is social, we are unable to rule out the possibility that all the co-variation we observe is of this kind. However, it seems very likely that structural relationships play some role, perhaps seeding potential social interpretations, which then strengthen the emergent patterns of co-variation.

To fully disentangle the interpretations, some close sociolinguistic analysis of the speakers and variants would be helpful. Our analysis is also overly simplistic in that it assumes a constant degree of co-variation across the time period studied. This may not be sufficient to account for vowels like STRUT/START which appear visually linked only in the later half of the data, and LOT/THOUGHT, which both change direction during the time period studied. If nascent structural co-variation becomes amplified by emergent social co-variation, this predicts increasing degrees of co-variation over a long time period. Further, if co-variation reflects leaders of particular sound changes, then such co-variation should be strongest during periods where the sound changes are most vigorous. Future analysis revealing how patterns of co-variation evolve and change over time will lead us still further toward understanding the mechanisms underlying patterns of co-evolution in the system as a whole. Such nuances present clear methodological challenges beyond those tackled in this paper, and remain for future work.

The study of how vowels change over time has traditionally examined individual vowels. When examining change in any particular vowel, research tends to find socially meaningful distributions of speakers, with speakers on the margins of the distribution representing 'leaders' or 'laggers' in that particular change. But it has proven challenging to move beyond individual variables, towards an understanding of sound systems. We have provided a significant step forward in this endeavour, by overcoming long-standing methodological challenges to the understanding of co-variation. Our analysis provides novel insights into the structure of sound systems, demonstrating the existence of structured patterns in the realisations of specific vocalic variables across a large group of speakers.

**Acknowledgements**

This research was made possible by a Royal Society of New Zealand Marsden Research Grant (17-UOC-049). The authors would like to thank members of the New Zealand Institute of Language, Brain and Behaviour for their feedback and comments as this work progressed. An earlier version of this work was presented at the International Congress of Phonetic Sciences, Melbourne, Australia 2019 and benefited from feedback.

**Supplementary Materials**

All files referred to in this manuscript as being in the Supplementary Materials can be accessed via:

- The online GitHub repository at:
  `https://github.com/nzilbb/Covariation_monophthongs_NZE`
- The Open Science Framework at:
  `https://osf.io/q4j29/`
- The online Shiny app (including the animated vowel spaces) can be accessed at:
  `https://nzilbb.shinyapps.io/Covariation_shiny/`

Both the GitHub and OSF repositories contain the same files (including the data, data filtering and analysis scripts and the Shiny application). The analysis file also contains additional analyses that were not considered crucial for inclusion in the paper, but may be of interest to readers.

## References


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*, 3099–3107. doi:10.1121/1.1795335.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2). *Distributed by the Linguistic Data Consortium, University of Pennsylvania*, .

Bauer, L., & Warren, P. (2008). New Zealand English: phonology. *Varieties of English*, *3*, 39–63.

Bauer, L., Warren, P., Bardsley, D., Kennedy, M., & Major, G. (2007). New zealand english. *Journal of the International Phonetic Association*, *37*, 97–102.

Becker, K. (2014). Linguistic repertoire and ethnic identity in New York City. *Language & Communication*, *35*, 43–54.

Becker, K. (2016). Linking community coherence, individual coherence, and bricolage: The co-occurrence of (r), raised BOUGHT and raised BAD in New York City English. *Lingua*, *172*, 87–99.

Boberg, C. (2019). A closer look at the short front vowel shift in Canada. *Journal of English Linguistics*, *47*, 91–119.

Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Ph.D. thesis University of Amsterdam.

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer 6.0.37 [computer program]. *https://www.praat.org*, .

Chodroff, E., & Wilson, C. (2017). Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *Journal of Phonetics*, *61*, 30–47.

Chuang, Y.-Y., Fon, J., & Baayen, R. H. (2020). Analyzing phonetic data with generalized additive mixed models. *PsyArXiv*, . URL: https://doi.org/10.31234/osf.io/bd3r4.

D'Arcy, A. (2017). *Discourse-pragmatic variation in context: Eight hundred years of LIKE* volume 187. John Benjamins Publishing Company.

Dickson, V., & Hall-Lew, L. (2017). Class, gender, and rhoticity: The social stratification of non-prevocalic /r/ in Edinburgh speech. *Journal of English Linguistics*, *45*, 229–259.

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

Eckert, P. (2000). *Language variation as social practice: The linguistic construction of identity in Belten High*. Wiley-Blackwell.

Eckert, P. (2016). Variation, meaning and social change. *Sociolinguistics: theoretical debates*, (pp. 68–85).

Eckert, P. (2018). *Meaning and Linguistic Variation: The Third Wave in Sociolinguistics*. Cambridge University Press.

Eckert, P. (2019). The limits of meaning: Social indexicality, variation, and the cline of interiority. *Language*, *95*, 751–776.

Fant, G. (1970). *Acoustic theory of speech production*. 2. Walter de Gruyter.

Foulkes, P., & Hay, J. B. (2015). The emergence of sociophonetic structure. *The handbook of language emergence*, *87*, 292.

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

Fruehwald, J. (2013). *The phonological influence on phonetic change*. Ph.D. thesis University of Pennsylvania.

Fruehwald, J. (2017). The role of phonology in phonetic change. *Annual Review of Linguistics*, *3*, 25–42.

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

Gordon, E., Maclagan, M., & Hay, J. (2007). The ONZE corpus. In *Creating and digitizing language corpora. Volume 2: Diachronic Corpora* (pp. 82–104). Palgrave.

Gordon, M. J. (2002). Investigating chain shifts and mergers. *The handbook of language variation and change*, *11*, 244–266.

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

Guy, G. R. (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics*, *52*, 63–71.

Hay, J., & Sudbury, A. (2005). How rhoticity became /r/-sandhi. *Language*, (pp. 799–823).

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.

Hebdige, D. (1984). *Subculture: The meaning of Style*. New York: Methuen.

Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. B. Fry, P. MacCarthy, N. C. Scott, & J. L. M. Trimm (Eds.), *In honour of Daniel Jones: Papers Contributed on the Occasion of His Eightieth Birthday* (pp. 73–84). London: Longmans.

Johnson, K. (2018). Speech production patterns in producing linguistic contrasts are partly determined by individual differences in anatomy. *UCBerkeley PhonLab Annual Report*, *14*, 256–282.

Keating, P. (2003). Phonetic and other influences on voicing contrasts. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 375–378).

Labov, W. (1994). Principles of linguistic change volume 1: Internal factors. *LANGUAGE IN SOCIETY-OXFORD-*, *29*.

Labov, W. (2001). Principles of linguistic change volume 2: Social factors. *LANGUAGE IN SOCIETY-OXFORD-*, *29*.

Lévi-Strauss, C. et al. (1962). *La pensée sauvage* volume 289. Plon Paris.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, (pp. 839–862).

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

Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, *49*, 606–608.

Love, J., & Walker, A. (2013). Football versus football: Effect of topic on /r/ realization in American and English sports fans. *Language and speech*, *56*, 443–460.

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*, 1–25.

Maclagan, M., Watson, C. I., Harlow, R., King, J., & Keegan, P. (2009). /u/ fronting and /t/ aspiration in Māori and New Zealand English. *Language Variation and Change*, *21*, 175–192.

Maclagan, M., Watson, C. I., Harlow, R., King, J., & Keegan, P. (2017). Investigating the sound change in the New Zealand English Nurse vowel /ɜː/. *Australian Journal of Linguistics*, *37*, 465–485. doi:10.1080/07268602.2017.1364126.

Mendoza-Denton, N. (2014). *Homegirls: Language and cultural practice among Latina youth gangs*. John Wiley & Sons.

Nevalainen, T., & Raumolin-Brunberg, H. (2011). The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change*, *23*, 1–43. doi:10.1017/S0954394510000207.

Oushiro, L. (2016). Social and structural constraints in lectal cohesion. *Lingua*, *172*, 116–130.

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

Pratt, T., & D'Onofrio, A. (2017). Jaw setting and the california vowel shift in parodic performance. *Language in Society*, *46*, 283.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: https://www.R-project.org.

Roettger, T. B., Winter, B., & Baayen, H. (2019). Emergent data analysis in phonetic sciences: Towards pluralism and reproducibility. *Journal of Phonetics*, *73*, 1–7.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). Major trends in vowel system inventories. *Journal of phonetics*, *25*, 233–253.

Sóskuthy, M. (2015). Understanding change through stability: a computational study of sound change actuation. *Lingua*, *163*, 40–60.

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

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

Sóskuthy, M., & Stuart-Smith, J. (to appear). Voice quality and coda /r/ in Glasgow English in the early 20th century. *Language Variation and Change*, . URL: `preprintavailableathttps://psyarxiv.com/td4px/`.

Stevens, K. N. (2000). *Acoustic phonetics* volume 30. MIT press.

Tamminga, M. (2019). Interspeaker covariation in Philadelphia vowel changes. *Language Variation and Change*, *31*, 119–133.

Thomas, B., & Hay, J. (2005). A pleasant malady: The ellen/allan merger in New Zealand English. *Te Reo*, *48*, 69.

Trudgill, P. (1974). *The social stratification of English in Norwich*. Cambridge University Press.

Warren, P. (2006). Oops. i've done a futt: quality and quantity in a New Zealand vowel contrast. *Te Reo*, *49*.

Warren, P. (2018). Quality and quantity in New Zealand English vowel contrasts. *Journal of the International Phonetic Association*, *48*, 305–330.

Waters, C., & Tagliamonte, S. A. (2017). Is one innovation enough? leaders, covariation, and language change. *American speech*, *92*, 23–40.

Watt, D. J. (2000). Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change*, *12*, 69–101.

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.

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*, 7–18.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

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.