CAMBRIDGE
UNIVERSITY PRESS

How stable are patterns of covariation across time?

Gia Hurring[1] (iD), Joshua Wilson Black[2] (iD), Jen Hay[1] and Lynn Clark[1]

[1]Department of Linguistics, University of Canterbury, Christchurch, New Zealand and [2]Department of Linguistics, New Zealand Institute of Language, Brain, and Behaviour, University of Canterbury, Christchurch, New Zealand
Corresponding author: Gia Hurring. Email: gia.hurring@pg.canterbury.ac.nz

Abstract
Recent research into vowel covariation has suggested that speakers can be identified as leaders or laggers in multiple ongoing sound changes. What remains unclear is how stable a speaker's patterns of covariation are over time and whether these leaders and laggers of sound changes remain leaders and laggers over time. We employ corpus data from 51 New Zealand English (NZE) speakers who were recorded at two time-points (eight years apart) and explore covariation between 10 monophthongs using principal component analysis (PCA). The results indicate significant stability across the time-points in two unique vowel clusters, suggesting that speakers' covariation position within their community remains stable over time. The overall covariation patterns also replicate patterns previously observed in a different corpus of NZE, indicating that patterns of vowel covariation observed with PCA can be stable and replicable across multiple corpora.

Keywords: New Zealand English; principal component analysis; vowel covariation; covariation stability; lifespan change

Introduction

The most common way of exploring language variation and change is by studying individual sociolinguistic variables. But do sociolinguistic variables really operate in isolation of each other? There are many reasons to assume that the answer is "no." For example, social meaning is likely to be distributed across multiple different variables and meaning manifests in how speakers combine or cluster different variants (Brand, Hay, Clark, Watson, and Sóskuthy, 2021; Campbell-Kibler, 2021). Moreover, when linguistic variables are involved in variation or change, they may be integrally linked together, for example through chain-shifts. In considering the covariation of variables, there are two sets of important questions in the literature. One relates to which variables work together—Are there subsystems of vowels, for example, that seem to cluster

together across speakers? A second relates to which speakers cluster together—Are some speakers reliably "leading" in all ongoing sound changes in a dialect?

In Brand et al. (2021), a methodology was developed that allows for the study of patterns of covariation that can address both questions. Brand et al. (2021) explored covariation and change in the production of monophthongs in the Origins of New Zealand English (ONZE) corpus—a large longitudinal spoken corpus of New Zealand English (NZE) (Maclagan & Hay, 2007). They found clusters of vowels that are statistically and systematically linked to each other. This suggests that if we know how a speaker produces one vowel in a particular cluster, we can make predictions about how they are likely to produce other vowels in that cluster.

One important question that stems from this work is: *How robust are these patterns of vowel covariation?* One way to check for this is to consider whether these clusters still exist in a different corpus of NZE, recorded much later when the variety is more stable. In this paper, we apply Brand et al.'s (2021) analysis to the QuakeBox corpus, a collection of earthquake stories told by survivors of the earthquakes which struck the Canterbury region in 2010/2011 (Walsh, Hay, Bent, Grant, King, Millar, Papp, & Watson, 2013). This corpus was recorded in 2012 and includes speakers born later than the youngest speakers in ONZE. We show that almost identical patterns of vowel covariation exist in the QuakeBox corpus, even though many of the sound changes have considerably progressed. Our results suggest that even when historic ongoing change becomes more stable, speakers may continue to use the same vowel clusters to index social meanings.

We also examine within-speaker stability in real time by looking at recordings from QuakeBox 2; earthquake stories retold by 51 of the original QuakeBox speakers eight years later. We ask: *Do the innovative speakers in QuakeBox still produce innovative variants eight years later?* We examine this in our corpus of re-recorded stories and find a high degree of correlation between the two sets of recordings. Where a speaker sits within the landscape of covariation in the first recording can reliably predict where they will be in the second recording. This suggests that even if a speaker's vowel realizations change over time, their position (as innovative or conservative) within a particular cluster of sound changes remains stable when compared with others in the community. In other words, speakers who can be identified as leaders of sound change at one time-point may remain leaders at later points in their life.

For readers who are interested in referring to the *R Markdown* file used for this paper, we have provided this as supplementary materials (R Core Team, 2021)[1]. Viewing the supplementary materials is not necessary to follow this paper, but they provide substantial extra context and detail for interested readers.

Background

Covariation, style, and social meaning

There are multiple reasons why vowels might cluster meaningfully together. In the absence of very clear structural relationships (e.g., in a chain-shift), one leading explanation could be that a cluster of vowels works together in the construction of social meaning. The term *social meaning* is used to refer to any form of indexical association

between linguistic material and some aspect of the "speaker as a social actor in the speech situation" (Eckert & Labov, 2017:3). Social meaning may appear to index categories of speakers (e.g., men, older, working class) and indeed identifying linguistic differences between broad categories of speakers was the focus of early sociolinguistic work (Labov, 2001). It is now understood that these categories of speakers are not the root cause of observed variation but rather reflect communities of practice, and patterns of behavior linked to broader constraints on the social order (Eckert, 2019). Individuals also deploy stylistic resources in the construction of personae—"holistic, ideological social types that are recognizably linked with ways of being and speaking" (D'Onofrio, 2020:1). Personae are linked with recognizable social types and are used to convey social meanings through associated stylistic resources (Agha, 2003; D'Onofrio, 2020). Recent work on personae raises the question of how they might be linked with "holistic style packages" (D'Onofrio, 2020:11). To the degree that we observe coherent subsets of vowels varying together across speakers, one potential interpretation is that they are working as a "style package."

Approaches to the study of covariation

Studies of potential patterns of covarying variables have been largely supported by careful analyses of variation within and across small numbers of speakers and largely using pairwise comparisons of two variables (Guy, 2013; Tamminga, 2019). The overall results of such studies are mixed. Some studies show that the realizations of some variables appear to be correlated across speakers, but many other pairs of changes appear unrelated.

Brand et al. (2021)

Brand and colleagues (2021) developed a technique to investigate covariation in a large corpus. Their methodology shows how elements of sound systems covary across hundreds of speakers and demonstrates how constellations of vocalic variables operate together. Brand et al.'s (2021) dataset is composed of NZE monophthongs from 481 speakers born between 1864 and 1982. They first obtained statistical estimates of how advanced each speaker was with respect to changes in each of the 10 vowels studied (detailed below). They then used principal component analysis (PCA) on these estimates to investigate the underlying covariation that exists across the vocalic variables. PCA finds covariation in multivariate data by taking the original variables and replacing them with principal components (PCs), where each new PC is selected to capture the maximum possible variation in the data. Each PC consists of numerical "loadings" assigned to the original variables which explain the relationship between the PC and original variables (Wilson Black, Brand, Hay, & Clark, 2023). Every speaker is assigned a "score" for each PC, which indicates their position with respect to the PC[2].

   This technique provides a pathway for answering questions both about vowels—such as which vowels appear to be working together—but also about *speakers*—such as whether any speakers are consistently ahead in sets of sound changes and, if so, who

those speakers are. The ability to statistically identify patterns of covariation of this kind lays the groundwork for the field to move to more large-scale analyses investigating covariation across multiple variables and multiple speakers. This paper adopts the methodology reported by Brand et al. (2021) and the steps detailed by Wilson Black et al. (2023) to address the stability of this observed covariation across apparent time and in different corpora, and within speakers in real time.

Linguistic change over the lifespan

Historically, adults' vowel production was thought to be stable, and indeed, this is the basis of the apparent time hypothesis, on which much work in language variation and change relies (Bailey, Wikle, Tillery, & Sand, 1991). However, it is now known that speakers can continue to shift their production of vowels over their lifetime. For example, Harrington, Palethorpe, and Watson (2000), in their data on Queen Elizabeth II's speech, exemplified how speakers may undergo individual/unrelated acoustic shifts over a lifetime, in the direction of community change. These shifts may be indicative of age-graded variation (Wagner, 2012) or they may be the result of a speaker continuing to participate in community sound change ("lifespan change"), or they may even move in the opposite direction to the community ("retrograde change") (cf. Sankoff, 2019).

Sheard (2023) came closest to investigating the stability of clusters of covarying variables in real and apparent time. This work explored real-time variation and change in 11 phonetic variables in the speech of five Greek- and five Italian- background Australians who were recorded in Sydney in the 1970s as teenagers (see Horvath, 1985) and again in 2019 as adults. These real-time changes were contextualized against an apparent time study of 117 speakers from the Sydney Speaks corpus.[3] Sheard (2023) showed that individual speakers in the real-time study consistently change over time in the direction of the community and that they change more for the variables that they were behind in when first recorded in the 1970s. There is also some attempt to explore the speakers' "'cohesion' in relative positioning" of a cluster of sound changes in progress (i.e., whether relative position in one sound change will correlate with relative position in other [structurally or socially] related changes) (Sheard, 2023:291). However, this work does not use PCA to explore cohesion; it presupposes which variables are likely to covary and explores pairwise correlations between sets of variables. Sheard (2023) found little evidence of cohesion across structurally related variables in the 1970s, and no evidence of cohesion across structurally unrelated variables, which is perhaps why this work does not then attempt to explore cohesion further in the 2019 data. The inability to find evidence of cohesion among these variables is perhaps an artifact of the very small sample size of only 10 participants (5 representing each speech community).

In the context of our study of covariation, we can ask a more specific question regarding speaker stability. Imagine that vowel realizations do shift considerably within speakers over time so that if we record the same cohort of speakers some years later, the vowels have moved. Will individuals' *positions* within the distributions for the sets of clustering vowels remain stable? Are the leaders at time-point one in a cluster of sound changes, still the leaders at time-point two?

Research questions

We have two research questions. These relate to the stability of patterns of vowel covariation, and the stability of a speaker's position within the wider distribution of vowel covariation.

1) *Stability of patterns of vowel covariation.* Are patterns of vowel covariation observed with PCA stable and replicable across multiple corpora?

We address this by replicating the analysis presented in Brand et al. (2021) on an entirely different corpus, recorded with younger NZE speakers.

2) *Stability of speaker position within axes of vowel covariation.* Are individual speakers stable in their position within covarying vowel clusters?

We address this by conducting a covariation analysis of two different sets of recordings by the same speakers—recorded eight years apart—and investigating how stable speakers are in terms of their position within the identified PCs. Both of our primary research questions were preregistered (AsPredicted.org, #86574).

To answer the research questions, PCA is conducted on two sets of recordings with the same speakers produced eight years apart and compares the results to each other, as well as to Brand et al. (2021). We conclude that despite considerable sound change between ONZE and QuakeBox (and minor sound changes between the two timepoints in our dataset), both the patterns of vowel covariation across corpora, and the positions of individual speakers within this covariation, are remarkably stable.

Methodology

Corpora

The analysis is conducted on the QuakeBox corpus, a collection of earthquake stories from the 2010/2011 Canterbury earthquakes (Walsh et al., 2013). In 2019–2020, many of the original participants returned to retell their stories, and recount what has happened since. We focus on speakers who are in both sets of recordings—referred to as QuakeBox 1 (QB1) and QuakeBox 2 (QB2). Analysis of the discourse structure of the "earthquake" story components has revealed that they have extremely similar structure across the two recordings (Blain & Millar, 2023). Recordings in QB2 tend to be longer (median QB2 = 19 minutes, median of same speakers in QB1 = 7.5 minutes). This is predominantly because they include not only the retold earthquake story but also a reflection of what has happened since. The recording context, content, and purpose across the two recordings are relatively equivalent. We thus have two recordings of each speaker, with approximately an eight-year span between them. The recordings have been transcribed and uploaded to a *LaBB-CAT* database which is a browser-based corpus management tool (Fromont & Hay, 2008). Using HTK, transcripts have been automatically aligned at the phoneme level. Vowel F1 and F2 values (at the vowel midpoint), along with other relevant phonetic and social information (see supplementary

Table 1. The number and percentage of vowel tokens per monophthong in all QB1

Vowel[5]	Tokens (QB1)	% (QB1)
DRESS	18,838	18.0
FLEECE	13,922	13.0
GOOSE	5982	5.7
KIT	14,734	14.0
LOT	10,192	9.8
NURSE	4672	4.5
START	5426	5.2
STRUT	12,720	12.0
THOUGHT	7949	7.6
TRAP	10,033	9.6
Total	**104,468**	**100**

materials S1), were extracted using *LaBB-CAT* and its integration with *Praat* (Boersma & Weenink, 2022).

### Visualizing overall patterns of change within QB1

First, we explore the direction of change for all 239 speakers in QB1 who remain after filtering ($F = 168$). This allows us to interpret our comparisons between real-time QB1 and QB2. The filtering process closely follows Brand et al. (2021) and detailed steps are provided in the supplementary materials (S2). Briefly, the filtering process removed measurement errors, missing speaker and formant data, stop words (due to their unstressed vowel realisations), tokens that preceded /l/ and /r/, and applies a 2.5SD outlier removal to each vowel token. After filtering, we normalized the data using the *Lobanov 2.0* method detailed in Brand et al. (2021). The final token counts are presented in Table 1.[4]

We fit generalized additive mixed models (GAMMs) on each formant. These models predict the normalized formant values with a parametric term for *gender* and a four-knot smooth for *age* at both levels of *gender*. We also fit *articulation rate*[6] as a control variable and random effect intercepts for *speaker* and *word*[7]. We note that the focus is on exploring (and ultimately controlling) known social factors associated with vowel variation, and there are likely to be unmodeled nonsocial influences, including phonological environment and lexical frequency. However, we have left these unmodeled in the interests of fitting simple parallel models for each vowel and formant, and to closely follow the structure in Brand et al. (2021). The random effect intercepts are likely to absorb any major sources of variation associated with the word.

Figure 1 shows the age effects from the GAMMs. This figure represents the apparent change over time, separately for men and women, with the beginning of the arrow representing the oldest age group (76–85) and the arrowhead representing the direction of change (i.e., the youngest age group, 18–25). GOOSE, THOUGHT, and LOT are not significant for men while STRUT and LOT are not significant for women. This is

A
QB1 Change Over Apparent Time by Gender
From 76–85 to 18-25
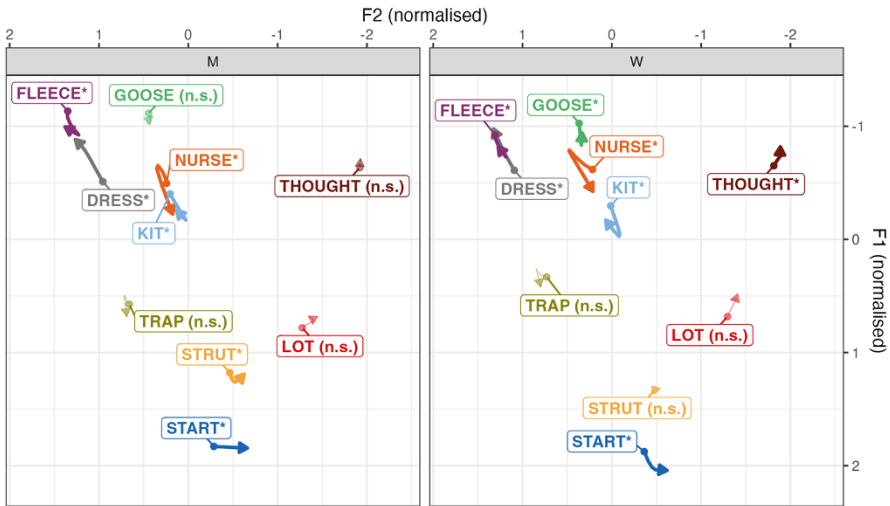
B
ONZE Change Over Real Time by Gender
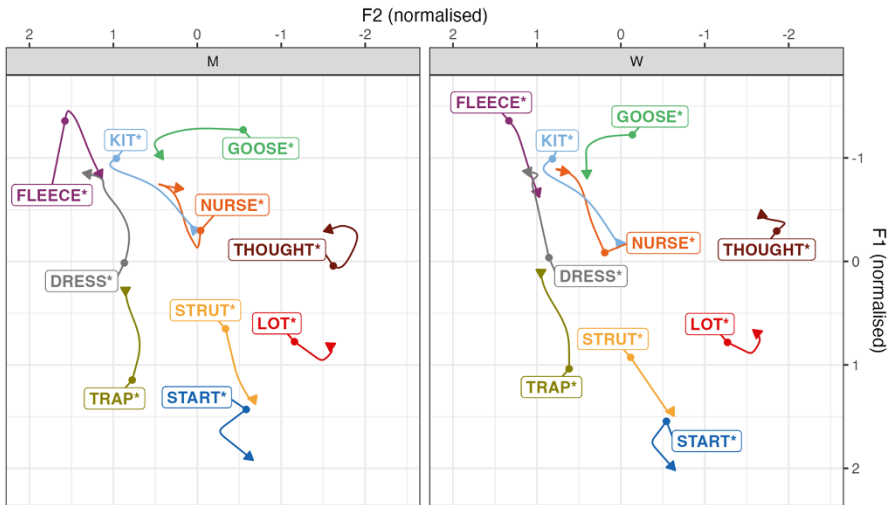Year of birth 1864 to 1982

Figure 1. Vowel changes in apparent time for QB1 (panel A) compared to the vowel changes in real time for ONZE (panel B). Men are on the left and women are on the right. Data for panel B comes from Brand et al. (2021) (CC BY-NC-ND 4.0). Vowels with statistically significant change over time in either F1 or F2 are displayed with a star, nonsignificant changes are marked with "(n.s.)."

compared, in the bottom panel, with the real-time sound changes observed within the ONZE corpus (Brand et al., 2021), which spans an earlier and longer period, and has significant effects for all vowels.

When comparing apparent time change within QB1 to the older ONZE data, we see some consistencies and some differences. With respect to the well-established short front vowel shift (Maclagan & Hay, 2007), we can observe the continued raising of DRESS, whereas TRAP is now relatively stable. KIT has been lowering and retracting for many years in NZE, but it appears to start reversing for our youngest speakers. The fronting of GOOSE has stabilized, although there is some movement in the F1 dimension. START continues to lower (women) and back (men), whereas STRUT looks to be raising for younger men. THOUGHT appears to be very high, and relatively stable, moving slightly up and back for the women. LOT shows some nonsignificant raising for women. NURSE is now lowering, as opposed to the raising trajectory seen in ONZE (Brand et al., 2021). We see a continued lowering of FLEECE[8], which swaps position with DRESS for the women, and then a reversal of this trajectory in the youngest speakers.

This is the landscape in which our investigation of covariation takes place. We may expect some differences from Brand et al. (2021) given the later and more compressed time-period, and some differences in terms of the directionality of vowel change. Crucially, whatever patterns of covariation we find in QB1, we expect some stability in the observed covariation to carry through to QB2, given that it contains the same set of speakers.

Visualizing real-time speaker change between QB1 and QB2

In this section, we visualize the stability of the vowels across the 51 speakers who are in both QB1 and QB2 ($F = 37$). QB2 contains nearly double the number of tokens as QB1 (because participants talked for longer in QB2), but the distribution of vowels is roughly equivalent, as shown in Table 2. A similar set of GAMMs to those detailed in the aforementioned section were run on this subset of data with *corpus* as a predictor but dividing the data by gender to reduce model complexity, giving 40 distinct models (10 vowels × 2 formants × 2 genders).

Table 2. The number and percentage of vowel tokens per monophthong in QB1 and QB2 after outlier removal

Vowel	Tokens (QB1)	% (QB1)	Tokens (QB2)	% (QB2)
DRESS	4018	18.0	7807	18.0
FLEECE	3005	13.0	5542	13.0
GOOSE	1311	5.8	2594	6.0
KIT	3235	14.0	7164	17.0
LOT	2249	10.0	4008	9.2
NURSE	1014	4.5	2259	5.2
START	1218	5.4	2036	4.7
STRUT	2630	12.0	5302	12.0
THOUGHT	1670	7.4	2709	6.2
TRAP	2179	9.7	3982	9.2
Total	**22,529**	**100**	**43,403**	**100**

The models fit a smooth for QB1 and a "difference smooth" which represents the difference between QB1 and QB2, divided into differences in the shape of the smooth (the "smooth term") and the overall height of the smooth (the "parametric term").[9]

Figure 2 visualizes the vowel spaces across the two corpora, plotting the predicted values for QB1 and QB2 for men and women. QB1 and QB2 are plotted on the same graph, with the QB1 vowels in orange and QB2 vowels in green. QB2 vowels are only shown if their parametric terms are significantly different from those in QB1. That is, Figure 2 represents change in real time while controlling for apparent time change, articulation rate, and speaker and word effects.

To the extent that we observe change across the corpora, this could relate to several possible sources. One is that there may be lifespan or retrograde change, in which case the QB2 vowels would be either in the direction pointed to by the corresponding arrow in Figure 1 (the potential "direction of change") or in the reverse direction, respectively. The second is that age differences seen in QB1 could reflect age-graded differences. If so, we might expect the QB2 vowels to be shifted in the opposite direction from the arrow, reflecting the fact that speakers have grown older, and shifted toward the more conservative variants used by older speakers in QB1. Finally, there may be stylistic variation between the corpora. The recording context differs, the urgency of the stories differs, and QB2 more often has an interlocutor whereas the QB1 stories are mostly monologues. Wholesale shifts related to such differences would be removed by the normalization process. But stylistic variation encoded in different vowels or clusters of vowels may be visible across the corpora, in a way that could be independent of the observed age effects.

For all speakers, FLEECE significantly lowers between the two sets of recordings. The women show a substantial raising of NURSE and DRESS and a slight fronting of STRUT. The men show a significant raising of KIT, and a small but significant fronting of GOOSE and backing of LOT. THOUGHT slightly raises (women) and backs (men). Some of these shifts are aligned with apparent time changes (Figure 1, panel A) and some are not. The youngest speakers in the apparent time analysis appear to be now raising FLEECE, following a period of lowering (Figure 1, panel A). In this real-time data, both speaker groups lower their FLEECE. THOUGHT shows a significant raising effect for women in apparent time (Figure 1, panel A), and the women in Figure 2 show significant raising of THOUGHT between the recordings. This is consistent with an interpretation in which speakers are changing in the direction of community norms (or "lifespan change"). The apparent time data (Figure 1, panel A) shows a recent raising by the youngest speakers of KIT—significant for women and nonsignificant for men, and the real-time data show men producing higher KIT. This, too, may be a case of speakers shifting in the direction of emerging change. NURSE, on the other hand, shows an overall lowering in the apparent time analysis (Figure 1, panel A), but considerable raising in the real-time analysis for women (Figure 2), indicating that there are also within-speaker changes that are not consistent with the direction of change and may instead indicate some age-grading or stylistic variation. We do not have sufficient data to robustly interrogate the degree to which these changes are equally carried across speakers of all ages. Work on life-span change indicates that different types of change may be more likely at different ages

Statistically Significant Changes in Parametric Term, QB1 v. QB2
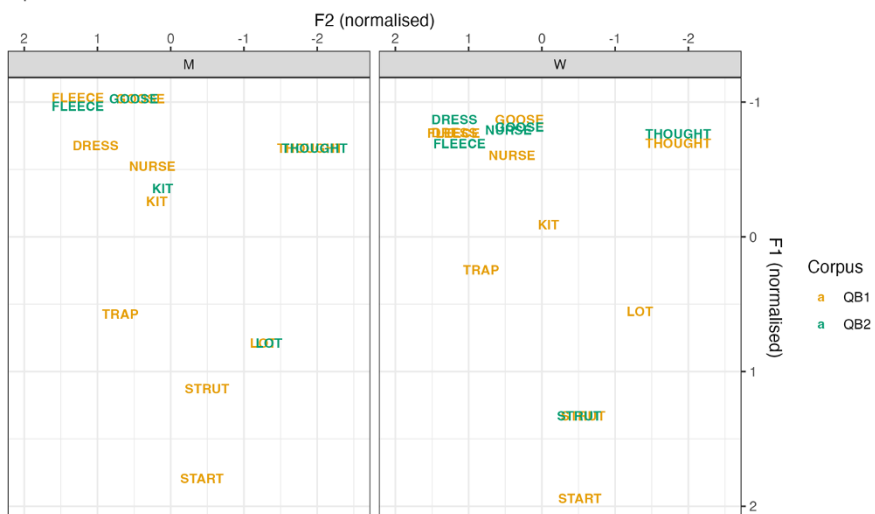Alpha = 0.025

Figure 2. Vowel spaces from QB1 (orange) and QB2 (green), plotted only in cases of significant difference with men on the left and women on the right. The significance test uses a Bonferroni correction (α = .025).

(Sankoff & Wagner, 2006), but our data are too sparsely distributed to look for robust age differences. The significant effects in the models simply tell us that there is a change that is carried by a nontrivial number of speakers, but not who those speakers are.[10]

Overall, the results show some stability between the corpora, and some change. The change is not always in a uniform direction and is not carried equally across both genders. There is unlikely to be a single factor explaining these shifts. Indeed, explaining the individual effects is not our focus. Rather, we ask, even though we have some observed change between the two time-points, are speakers' positions within patterns of covariation nonetheless stable?

Modeling covariation patterns in QB1 and QB2

This section reports results of the PCA performed on the 51 speakers in QB1 and QB2. This analysis follows the procedure outlined by Brand et al. (2021) but with additional techniques as detailed in Wilson Black et al. (2023). Both studies suggest that PCA is a promising method to explore holistic vowel covariation if the data that undergoes PCA has already been controlled for known factors of sound change. To do this, we fit GAMMs following the same structure reported above to the speakers who are in both QB1 and QB2. We then extract the random *speaker* intercepts from these models, which represent the differences between speakers in formant values which remain after gender, age, and articulation rate have been controlled for (Wilson Black et al., 2023)[11]. The resulting dataset gives distinct vowel intercept values per speaker, per corpus, which are used as input data to a PCA.

Overview of PCA

PCA returns similar results when applied to QB1 and QB2. Both PCAs were robust when compared with random permutations of the data, suggesting that PCA is reflecting real covariation in both datasets. Following Wilson Black et al. (2023), we apply another permutation test to explore 95% confidence intervals of the variance explained in each PC. This is done by bootstrapping the data to estimate the stability of the variance explained by each PC. These random permutations and the bootstrapped samples are depicted in Figure 3, with QB1 in the top panel and QB2 in the bottom panel. PC1 accounts for 33.1% of the original variation in QB1 and 25.9% of the original variation in QB2. PC2 accounts for 15.9% of original variation in QB1 and 14.1% of the original variation in QB2. PC1 and PC2 explain more variance than expected from randomly permuted data in QB1. Consequently, we focus our analysis on PC1 and PC2 (unlike Brand et al., 2021, who explored three PCs). In QB2, PC1–PC3 explain more than expected from randomly permuted data. However, the overlap between sampling distributions (red) for PC2 and PC3 in both QB1 and QB2 indicates instability in the loadings of these PCs (in a way to be explained).

In short, we are focusing on the two PCs that are significant in our data. The first one is heavily weighted for the vowel measurements involved in ongoing changes, on which speakers are ranging along a leader–lagger continuum. The second is loaded for back vowel measurements. The ONZE results found by Brand et al. (2021) found two comparable PCs with similar loadings but weighted in the opposite order. See below for further discussion on this PC switch.

PC1: leaders and laggers of change

PC1 loadings for both corpora are given in Figure 4, with QB1 in the top panel and QB2 in the bottom panel. Vowels are listed from left to right with vowels on the right having greater contribution to the PC. Vowels that are positively loaded are indicated with a black cross. Vowels that are negatively loaded are indicated with a minus sign. The direction of the loadings is arbitrary in that the plus and minus signs could be swapped without affecting the interpretation. We have manually flipped the loadings for PC1 to match the orientation of the corresponding PC in Brand et al. (2021). Vowels with loadings and confidence intervals outside the null distribution are the most stable in the PC. In QB1, F1 FLEECE and F1 KIT are both negatively loaded. The fact that they are loaded in the same direction means that when a speaker has a low F1 for FLEECE, they will also have a low F1 for KIT. In other words, vowel formants sharing a sign are positively correlated with each other, and vowel formants with opposite signs are negatively correlated with each other.

The PC1 loadings for the different corpora are compared in Table 3, with loadings from a matching PC (PC2) in Brand et al. (2021). Included in the table are all vowel formants that fall above the 50% threshold in PC1—meaning these vowels collectively account for at least 50% of variation in that PCs' total loadings. In other words, these vowel formants are the "top contributors" to the PC loading. Although some vowels are more strongly loaded in some of the analyses than others, we can see that the same pattern of covariation is found in all three corpora. In all cases, speakers with high
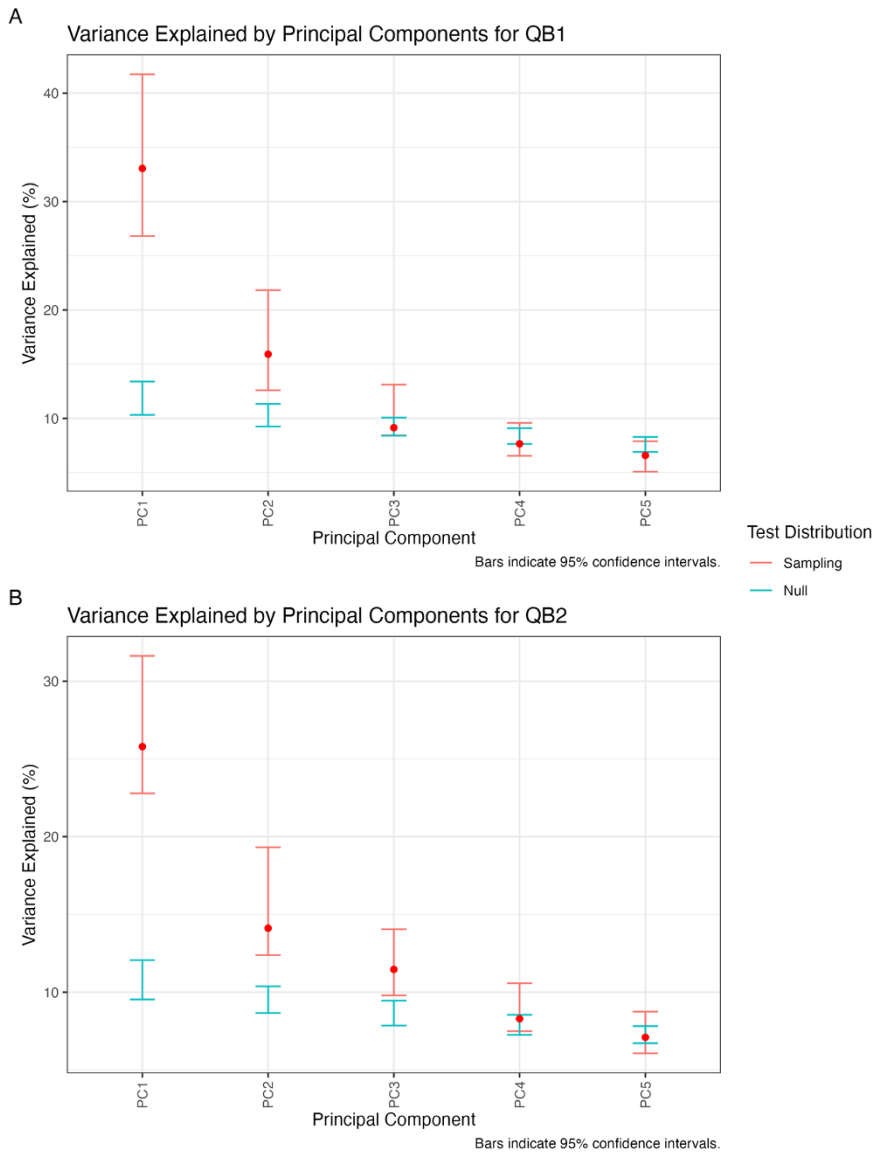
A

Variance Explained by Principal Components for QB1

Bars indicate 95% confidence intervals.

Test Distribution
— Sampling
— Null

B

Variance Explained by Principal Components for QB2

Bars indicate 95% confidence intervals.

Figure 3. Variance explained by top five PC loadings for QB1 (panel A) and QB2 (panel B). The red "sampling" distribution represents the range of variance explained attained across the bootstrapped analyses, and the blue "null" distribution indicates the variance explained by random permutations of the data. Bars indicate 95% confidence intervals.

loadings on this PC are producing innovative variants of these vowels and speakers with low loadings are behaving conservatively.

This PC includes New Zealand's short front vowel shift (Maclagan & Hay, 2007) as centered on raising TRAP and DRESS vowels, and a lowering and centralizing KIT
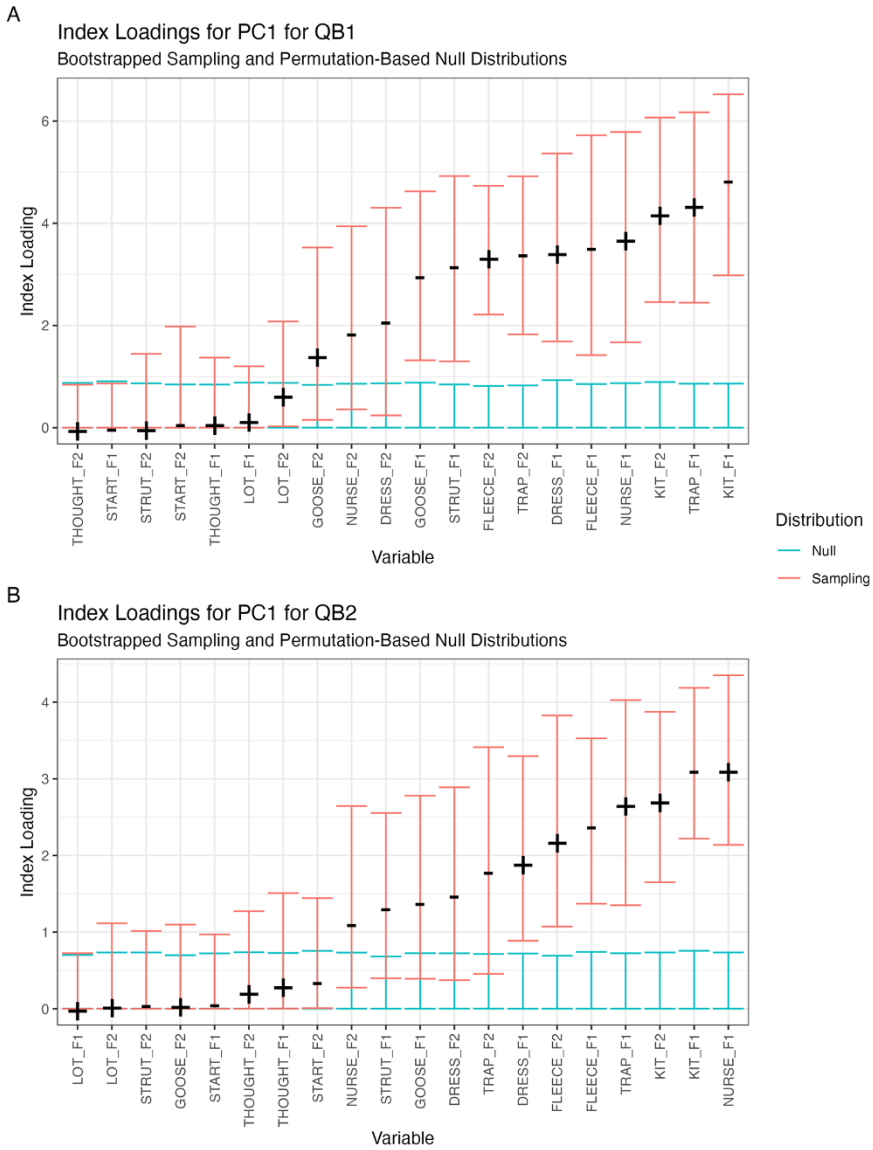
Figure 4. PC1 loadings for QB1 (panel A) and QB2 (panel B).

vowel. These vowels account for the four strongest loadings of PC1 in both QB1 and QB2—with positive loadings for DRESS and TRAP F1 and KIT F2 (conservative speakers having lower DRESS, TRAP, and fronter KIT), and a negative loading for F1 KIT (with conservative speakers having higher KIT).

This PC is not limited to the short front vowels—other vowels are also loaded in the direction of well-documented sound change in NZE, particularly the raising of NURSE,

Table 3. Comparison of matching principal components from QB1, QB2, and ONZE (Brand et al., 2021) (CC BY-NC-ND 4.0); + and − represent loadings within the top 50% of the contribution to the PC, and parentheses indicate lower contributions (below 50%)

	QB1 (PC1)	QB2 (PC1)	ONZE (PC2)
F1 KIT	−	−	(−)
F1 TRAP	+	+	+
F2 KIT	+	+	+
F1 NURSE	+	+	+
F1 FLEECE	−	−	−
F1 DRESS	+	(+)	+
F2 LOT	(+)	(+)	+

and the lowering and backing of FLEECE (McKenzie, 2005). Despite the remarkable consistency across these analyses, there are some differences in the strength of involvement of different vowels. LOT F2 was strongly loaded in Brand et al. (2021) but is absent in QB1 and QB2. Conversely, the QB analyses show KIT F1 as one of the most strongly loaded formants, whereas this fell below the threshold for interpretation in Brand et al. (2021).

Our first research question asked: *Are patterns of vowel covariation observed with PCA stable and replicable across multiple corpora?* We have shown that the vowel covariation first discovered in NZE in Brand et al. (2021) is replicated in both of our more recent QuakeBox corpora. The degree of replication is particularly notable given that the patterns of sound change (Figure 1, panel A) for the QuakeBox speakers is not the same as seen in the earlier ONZE corpus (Figure 1, panel B). For example, a low KIT F1 may be conservative when considering the QuakeBox speakers as a population, but we are starting to see a reversal of this trend with KIT raising again in the vowel space in our youngest speakers. Despite this, in the overall covariation pattern, a low F1 (i.e., raised KIT) patterns with the conservative variants of other vowels with which it has been shown to covary for some time. Similarly, TRAP F1 remains robustly represented, despite the complete lack of an age effect in the GAMMs. It is interesting to observe well-documented patterns of "innovative" vowels clustering together even in a corpus that does not contain evidence within itself of ongoing change. This may indicate that the social meaning associated with "innovative" and "conservative" speech can permeate users' speech for some time after there is an observable age effect within the community.

On the other hand, LOT is largely backing in the ONZE corpus, whereas in QuakeBox, LOT is no longer in the set of vowels characteristic of the "leaders," perhaps suggesting that some previous social meaning associated with LOT F2 is no longer present. Similarly, NURSE is raising in ONZE, and in the beginning of QuakeBox, but then begins to lower. In QB1 and QB2, we see high NURSE patterning with the innovative vowels. These relatively minor fluctuations over time seem to suggest that the innovative/conservative axis of variation can remain relevant for specific vowels for some time after community-level age-effects are apparent.

We turn now to question 2, which asked: *Are individual speakers stable in their position within covarying vowel clusters?* To answer this, we can extract the speaker scores

from the PCA, as a measure of where individuals sit with respect to this cluster. A Spearman correlation test between these two sets of loadings (i.e., PC1 in QB1 and QB2) shows the scores have a strong, significant relationship, displayed in Figure 5. Thus, even though several vowels change in their production between QB1 and QB2 (Figure 2), individual speakers' position within this covariation remains remarkably stable. In terms of research question 2, then, at least with respect to PC1, the answer
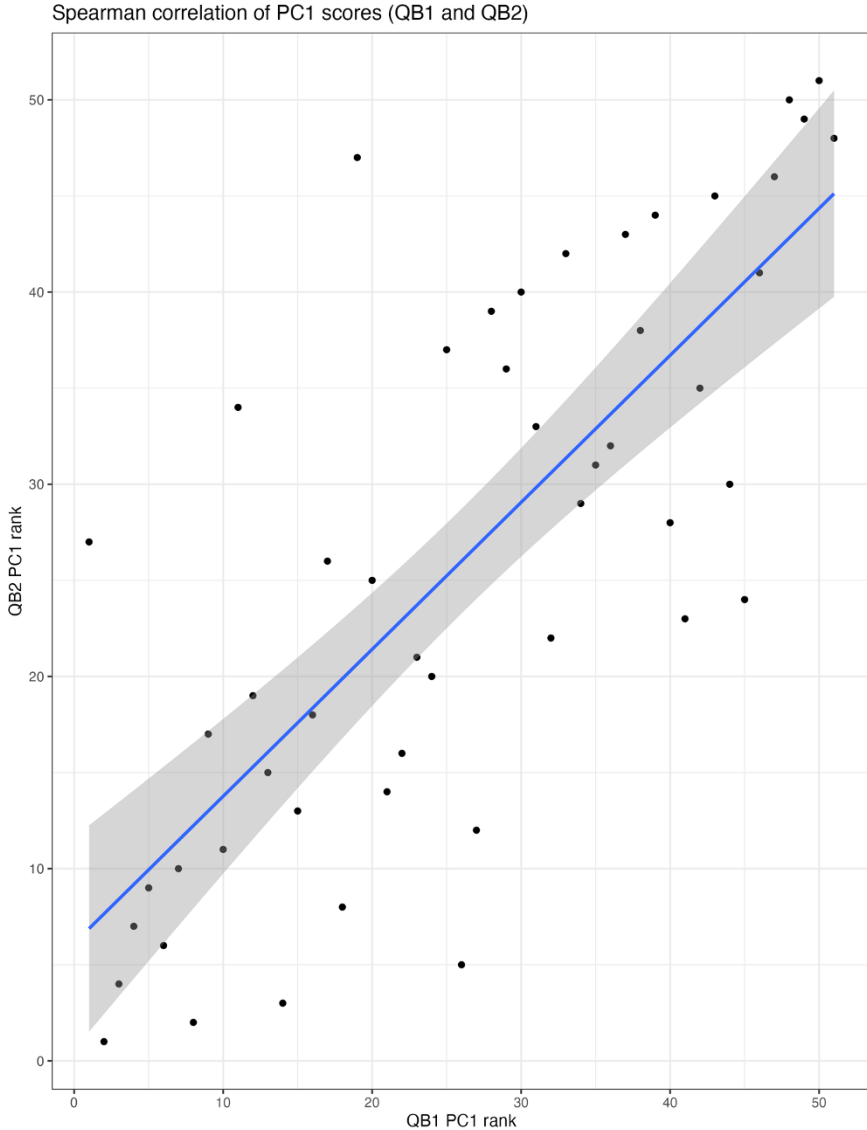

Figure 5. PC1 speaker loadings correlated across QB1 and QB2 (rho = .77, *p* < .001).

is yes—speakers are very stable in their positioning within the covariation over time. Speakers who are innovative in QB1 continue to be innovative in QB2, eight years later.

PC2: the back vowels

PC2 loadings for both corpora are shown in Figure 6, with QB1 in the top panel and QB2 in the bottom panel. PC2 is less stable than PC1, with many of the confidence intervals overlapping the null distribution. We only consider vowels where the loadings signs (±) are above the null distribution (indicating that they explain more variation than expected by chance) and vowels whose sampling confidence intervals do not touch zero (indicating that their appearance in a PC with the other identified variables is stable) (Wilson Black et al., 2023). This includes (reading right to left on the *x*-axis) STRUT F2 through to THOUGHT F2 in QB1 and LOT F2 through to THOUGHT F1 in QB2. In this PC, speakers with fronter START vowels also have a fronter STRUT, and a fronter and higher THOUGHT. Additionally, speakers in QB1 have a fronter and higher LOT covarying with the previous vowels, whereas LOT F1 drops from this covariation in QB2, leaving only LOT F2 to covary with START, STRUT, and THOUGHT.

Table 4 shows the loadings on PC2 (from this analysis) and from Brand et al.'s (2021) PC1, which is the corresponding PC from their analysis. It also indicates the general direction of sound change in each corpus based on Figure 1. For Brand et al. (2021), we have simplified the description of sound change to only include the trajectory at the end of the time-period studied. In ONZE, all vowels in the cluster were aligned in the direction of change except for F2 THOUGHT. It was proposed that speakers participated in a linear reorganization of the back vowel space in different ways—speakers who were innovative in their F2 for a set of back vowels were conservative for the raising of THOUGHT, and vice versa.

In QuakeBox, we see THOUGHT much higher in the vowel space, and generally less movement amongst all the vowels (Figure 1, panel A). When they are moving, they are moving in the same direction as they were in Brand et al. (2021) except for THOUGHT F2, which is now is backing (see the final row of Table 4). The polarity of the loading of THOUGHT F2 on the PC has also flipped.

Brand and colleagues (2021) claimed that this vowel cluster encodes innovation in a set of vowels (including THOUGHT F2), together with conservatism in THOUGHT F1. In QB2, this interpretation can hold, but THOUGHT is now backing, and so the specific realization in terms of what it means to be innovative has changed. When THOUGHT was fronting, speakers who were innovative in START, STRUT, and LOT were innovative in the fronting of THOUGHT. Now that it is backing, it is still the same speakers that are leading that change.

It is interesting to note that the F1 of THOUGHT is still loaded in the opposite direction, with respect to the historic sound change; speakers who are innovative in F2 in the back vowels are "conservative" in THOUGHT F1. Though THOUGHT F1 is not significantly changing within the QuakeBox corpus, the previously observed structural relationship between it and the other vowels remains, despite its different overall position and lack of current momentum. This provides further support for

the interpretation that covariation in these vowels can be stable and potentially socially meaningful in a sense that is not directly tied to any currently ongoing change.
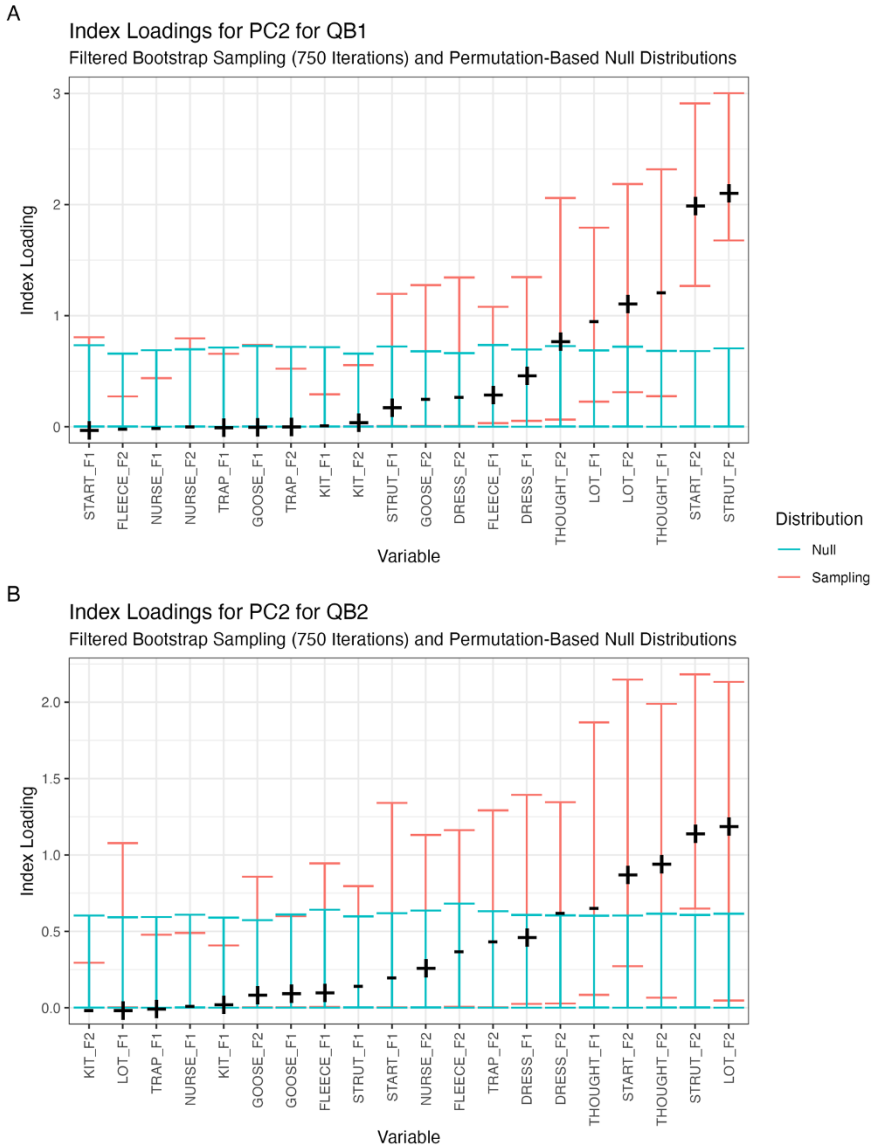

Figure 6. PC2 loadings for QB1 (panel A) and QB2 (panel B). Only the 750 bootstrapped iterations with the highest values for the largest magnitude loading are included. Sampling intervals indicate a conditional 95% bound. For instance, the sampling distribution for START F2 in QB1 indicates a 95% interval in cases where STRUT F2 is in the top three quarters of its distribution.

Table 4. Comparison of matching principal components from QB1, QB2, and ONZE (Brand et al., 2021) (CC BY-NC-ND 4.0); + and − represent loadings within the top 50% of the contribution to the PC, and parentheses indicate lower contributions

	QB1 (PC2)	QB2 (PC2)	ONZE (PC1)
F2 STRUT	+	+	+
F2 START	+	+	+
F1 THOUGHT	-	(-)	-
F2 LOT	(+)	+	(+)
F2 THOUGHT	(+)	+	-

Despite these differences in the direction of sound change between ONZE and QuakeBox, what we see from Table 4 and Figure 6 is relative consistency across the corpora in terms of the axes of covariation. For PC2, the answer to research question 1 is that this vowel cluster is stable and observable across multiple NZE corpora. In this PC, however, we see a change in the alignment of one of the vowels that may be linked to interim changes in the overall direction of sound change.

Again, we can test whether individual speakers are stable in their position within this covariation between QB1 and QB2, shown in Figure 7. The Spearman correlation is significant ($p = .02 < .05$), although much weaker than for PC1[12]. If a speaker shows a particularly vertical configuration of their back vowels in QB1 they are likely to retain that pattern into QB2.

The changed relevance of PC1/PC2 across corpora

As outlined above, the relative ordering of the two PCs has switched between Brand et al. (2021) and our analysis. In Brand et al. (2021), their PC1 (back vowels) accounted for 17.2% of the variation, and their PC2 (leaders and laggers) accounted for 15.8% of the variation. The two components were not dramatically different from each other in how much variation they explained. By contrast, in QB1, our PC1 (leaders and laggers) accounts for 33.1%, and PC2 (back vowels) accounts for 15.9%. QB2 is similar. It is important to note that the absolute numbers cannot be directly compared across analyses. There are many reasons why the absolute value of a PC could be different in QB1 than ONZE, including the fact that the recordings are newer and likely to be less noisy, and that only two PCs surface as reliable in that analysis. We cannot say that "leaders and laggers" has become more important, because it now explains 33.1% of the variance instead of 15.8%, nor can we say that "back vowels" have become less important. What we can conclude is that a gulf has opened between these patterns of covariation. The "leaders and laggers" cluster is now much more explanatory—compared to the back vowels—than it used to be. If we consider the patterns of changes over the history of NZE, we see that there is a history of change in the back vowels (Figure 1, panel B), but that they are relatively stable now (Figure 1, panel A). We know from our analysis that they still covary in a meaningful way. But it is possible that the weight of social meaning carried by these now relatively stable vowels is waning, and/or that the vowels represented by the leader/lagger continuum (most of which are still changing) are more meaningful at this moment in time. Our analysis, and its comparison to Brand et al.

(2021), shows that patterns of covariation in a dialect can endure across corpora, across different recordings of the same speaker, and across evolving sound change patterns. But this does not imply that the relevance of such patterns is completely static. The fluctuation in relative importance of these clusters over time suggests that patterns of covariation can themselves be subject to change.
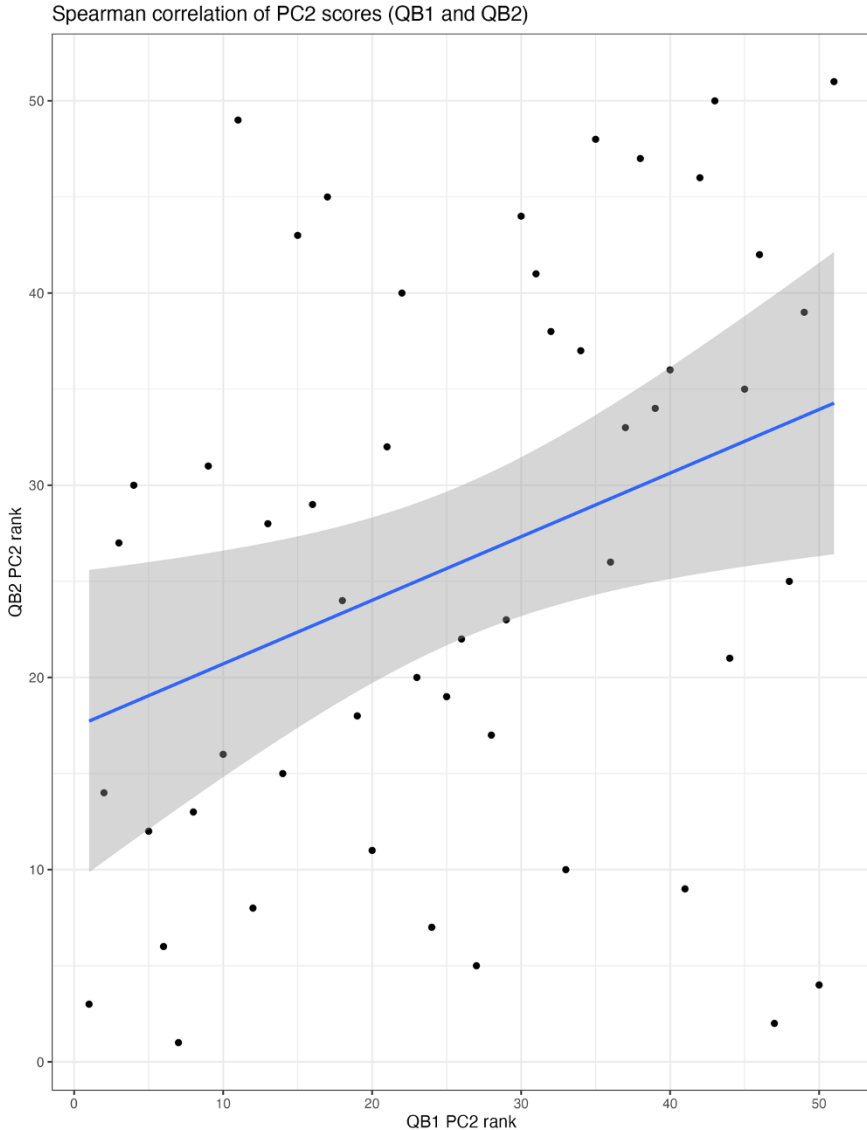

Figure 7. PC2 speaker correlations across QB1 and QB2 for the original models (rho $= .33, p < .05$).

Additional analyses

We conducted further analyses on the QuakeBox data to investigate whether our stability is an artifact of other mechanisms such as uncontrolled social factors or repetition of content. Adding an *ethnicity* control into the GAMMs did not significantly change the vowel clusters. This suggests that our stability is not an artifact of the unmodeled social factor ethnicity that is a known source of variation in NZE (Maclagan et al., 2009).

Additionally, parts of the QB2 corpora contain retold phrases and common lexical items due to speakers retelling their earthquake story. However, when we divided QB2 into "new" and "repeated" speech, we found no evidence that the correlation is stronger for the retold material.

We also explored whether these clusters are explainable in more universal terms. We conducted PCA on 30 speakers from an entirely different corpus, the Origins of Liverpool English corpus (Clark & Watson, 2016), which resulted in unique PC clusters unlike those that have been found in NZE. Thus, the patterns of covariation reported in this paper and in Brand et al. (2021) really are patterns that relate to NZE[13].

Discussion

We have analyzed patterns of vowel covariation in a set of 51 speakers, each recorded twice (eight years apart). Our data show some differences from Brand et al.'s (2021) analysis in terms of the progression and, in some cases, emerging reversals of sound changes. The two time-points in our data also show some differences, with speakers significantly shifting several of the vowels between the two time-points. Our research questions relate not to these overall changes, but rather to the stability of relationships between clusters of vowels in the context of these changes. First, we asked *whether patterns of vowel covariation observed with PCA are stable and replicable across multiple corpora.* Using the QuakeBox corpus, we largely replicated the two major patterns of covariation reported in Brand et al. (2021).

Second, we asked *whether speakers are stable in their position within covarying vowel clusters.* Spearman correlation tests reveal that individual speakers' PC1 and PC2 scores in QB1 are statistically predictable by their scores in QB2 (Figures 5 and 7). This effect is much stronger for PC1 than PC2. Although many of the vowels in question change significantly between the two corpora (see Figure 2), we find reasonable stability in individuals' positions within the distribution of covariation. The stability of PC1 suggests that what counts as "innovative" might look slightly different in 2019 than 2012, but whether a speaker is "innovative" is relatively stable across the two corpora. The implication is that speakers are finely tuned in to directions of change in the community even as they age, staying in the leadership position even as the community moves in a new direction of sound change. We note that, in the context of a lifespan, the time between our two corpora (eight years) is reasonably short. Whether such high levels of stability would be observed at longer intervals is a question which would warrant further study.

Relationships between sound change and "social meaning"

PC1 links a set of vowels that have significantly changed over the history of NZE. There are few changes between the patterns of covariation picked up by Brand et al. (2021)

and the patterns of covariation documented here, even though many of the sound changes that "leaders" were leading seem to have since been completed or reversed. This appears, then, to reflect social meaning that may have been linked with innovativeness but has remained stable beyond the momentum of the sound changes themselves. As outlined by Guy, Oushiro, and Mendes (2022:55), we might expect that speakers are likely to use variants together that "jointly or separately index some particular social meaning." Once a set of covarying sound changes carries social meaning, ongoing change is not necessarily required for that social meaning to persist. It has been shown that social variation which appears to be tied to demographic features can take on more locally relevant and nuanced meanings. For example, Eckert (2008:462) showed how variables that historically distinguish geographic dialects "can take on interactional meanings based in local ideology." In the case of our sound changes, some of the social stratifications may no longer persist, but the covariation continues to exist and so it is likely that the ideological orientation associated with the covariation does too.

Eckert and Labov (2017) discussed a variety of sound changes through Silverstein's (2003:4) notion of orders of indexicality, "through which a variable can progress from indexing a social position (e.g., with a region, a class position, an ethnic group) to indexing some quality apparently associated with the people occupying that social position." If the patterns of covariation picked up by our PCs relate to social constructs such as persona, then our results suggest that, while these may not be entirely independent of patterns of sound change, they can persist beyond the change itself.

We note that Eckert and Labov (2017) explicitly excluded patterns of sound change such as vowel shifts from the type of structure likely to attract social meaning because complex sequences of sound change are not likely to be objects of social perception. Our PC1 includes not only a vowel shift but also a range of other forms that underwent change at the same time. We have demonstrated that the covariation across speakers persists over a long period of time, which we believe indicates that there is some persona-based social meaning associated with this covariation. However, to properly understand whether these patterns of covariation are picking up on stylistic persona-based variation, we need to carry out careful perceptual work.

Covariation in real and apparent time

When exploring changes in vowels over real time (Figure 2), there are some differences in the production of certain vowels in the eight-year gap between QB1 and QB2. Let us suppose that these are due to vowel changes that take place over the lifespan of the speakers (rather than stylistic differences between the two corpora[14]). The fact that these changes in real time are present is not unexpected. Gillian Sankoff and Blondeau (2007) showed that individuals can continue to change their linguistic production post-adolescence in the direction of the community in a process known as "lifespan change." G. Sankoff and Wagner (2006) also described "retrograde change," where speakers move away from the direction of community change over time and adopt more conservative variants with age. We may have evidence of both lifespan and retrograde change in our data, although we note that to fully untangle what is going

on with these individual vowels, a more detailed study would be needed, including investigation of the degree to which speakers of particular age-groups may be disproportionally responsible for these changes. In our apparent time data (Figure 1, panel A), FLEECE is lowering and KIT is beginning to raise after a period of lowering. In our real-time data (Figure 2), both men and women now have lower FLEECE realizations than they did eight years ago, and men are raising their KIT vowels, following the most recent community innovations. These are two examples of potential lifespan change. We can also see in our apparent time data that GOOSE is beginning to raise for the youngest women, but in our real-time data, women now have a lower GOOSE realization than they did eight years ago. This is a potential example of retrograde change.

Work on change across the lifespan has mainly focused on exploring variation and change in single linguistic variables. Our work shows that while there may be differences in the rate and direction of change of individual vowels when we compare apparent time with real-time data, the speakers' axes of covariation remain relatively stable over time, even as the vowels which make up the cluster of covariation changes over the same time-period. A speaker who was a "leader" (or produced more innovative articulations) in the cluster of sound changes represented by PC1 in 2012 is still a "leader" in 2020, eight years later.

Beaman and Sering (2022) explored sociolectal coherence over real and apparent time in two German communities (Stuttgart and Schwäbisch Gmünd). The real-time data came from 20 speakers who were interviewed in 1982 and again in 2017, and an apparent time set of "social twins" (i.e., speakers who are matched for age, sex, education, and location with the real-time speakers). A PCA was conducted on data from 20 linguistic variables and variation was documented between pairs of PCs (e.g., participants PC1 score from 1982 is plotted against their PC2 score from 1982, etc.). These were then used to calculate measures of dialect coherence. The real-time findings suggest that the 1982 Schwäbisch Gmünd speakers showed the most coherence with little variation and no overlap in the PCs between Schwäbisch Gmünd and Stuttgart. By 2017, however, both dialects showed less cohesion (i.e., greater variability) and there was overlap in the PCs, which Beaman and Sering attributed to dialect contact with speakers now "juggling" and "mixing" these two different varieties. The apparent time analysis tells a largely congruent story. This work confirms the importance of using PCA as a tool to understand clusters of variation in both real and apparent time, but because participants' PC scores from 1982 were not correlated against their PC scores from 2017, we are unable to tell whether the results from Beaman and Sering (2022) match our own with respect to the *stability* of a participant's position along the axis of variation for a particular cluster or PC.

In summary, "there is clearly much we still do not know about co-variation in real and apparent time, and very little we know about co-variation over the course of an individual's lifetime" (Sheard, 2023:63). We are left with substantial questions. For instance, if the position of an individual vowel changes in real time, but both the clusters and the individual's position on the axes of variation remains stable, what are the implications for social meaning? If patterns of covarying variables carry social meaning, do listeners track covariation and are they aware of patterns of relationships between phonetic variants?

Conclusion

The findings of this paper contribute to our understanding of lifespan variation and real-time change in a covariation context. The nature of the relationships between the vowels appears to be more stable over time than the specific realizations of the individual vowels themselves. Furthermore, individual speakers' positions within the axes of covariation are also relatively stable over time. It seems very likely that these patterns of observed covariation are socially meaningful, consistent with arguments put forward by Eckert, that "sociolinguistic variables do not occur independently, but as components of styles" (Eckert, 2019:752). A major question yet unanswered is the degree to which this social meaning is accessible to the listener. This is an important question for future work.

Notes

1. The supplementary materials can be viewed as a web page at https://nzilbb.github.io/qb_stability_public/markdown/QB-analysis.html. An editable *R Markdown* file and anonymized data are available at https://github.com/nzilbb/qb_stability_public.

2. While some parts of the observed patterns of covariation could be attributed to structural relationships such as chain-shifting, there are others which do not appear to be of this type. Brand et al. (2021) concluded that, in general, their results support the prediction emerging from the sociolinguistic literature that there should be broad patterns of correlations across variables, and that social variables might work together rather than in isolation.

3. https://legacy.dynamicsoflanguage.edu.au/.

4. Following Brand et al. (2021), we collapse FORCE and NORTH together with THOUGHT, and we collapse START and BATH. It is widely assumed that NZE speakers do not make phonological distinctions between these categories. While we analyze the set of vowels commonly analyzed as the set of "monophthongs" in NZE, we acknowledge that some of these vowels are likely have some diphthongal realizations for some speakers. By concentrating on the midpoint, we hope that we largely avoid the influence of any on-glides or off-glides in these vowels.

5. Brand et al. (2021) excluded FOOT from analysis due to low token numbers. We exclude FOOT as our aim is to replicate Brand et al. (2021) on a different NZE corpus and, in the process, also found low tokens of FOOT in our data (3184 tokens in QB1 and 1222 tokens in QB2).

6. Articulation rate refers to the number of syllables produced per second, after removal of silences/pauses.

7. See S3.1 of the supplementary material for full details.

8. Typical phonological analysis of monophthongs for NZE includes the FLEECE vowel, although the degree to which this is a monophthong is debatable. We assume that any effect of an on-glide or off-glide is minimized by taking a midpoint measure.

9. See S3.2 of the supplementary material for full details.

10. For readers interested in potential age effects in QB1 and QB2, detailed plots of the age effects for each vowel and each corpus are available in the supplementary materials (S4.1.3).

11. See S4.1 in the supplementary materials for full details.

12. The lower correlation value for PC2 suggests that there is some change in the back vowels for some speakers between QB1 and QB2 (Sheard & Wilson Black, 2024).

13. For more detail of these analyses, please refer to the supplementary materials (S5 and S6).

14. This is not to diminish the very real problem of disentangling lifespan change from style shifting (Rickford & Price, 2013), topic effects (Gregersen & Barner-Rasmussen, 2011), or differences in interviewer familiarity between corpora (Wagner & Tagliamonte, 2017). We simply make this assumption because while there are differences between these two corpora, there are many more similarities (including not only the same general topic of conversation but the same narrative).

References

Agha, Asif. (2003). The social life of cultural value. *Language & Communication*, *23*(3–4), 231–273. https://doi.org/10.1016/S0271-5309(03)00012-0.

Bailey, Guy, Wikle, Tom, Tillery, Jan, & Sand, Lori. (1991). The apparent time construct. *Language Variation and Change*, *3*(3), 241–264.

Beaman, Karen V., & Sering, Konstantin. (2022). Measuring change in lectal coherence across real- and apparent-time. In V. B. Karen, & R. G. Gregory (eds.), *The Coherence of Linguistic Communities* (pp. 87–105). Routledge.

Blain, Hayden, & Millar, Paul. (2023). Shifting discourses of togetherness and heroism in retold earthquake stories. *Narrative Inquiry*.

Boersma, Paul, & Weenink, David. (2022). *Praat: doing phonetics by computer* [Computer program]. Version 6.2.10, retrieved 17 March 2022 from http://www.praat.org/

Brand, James, Hay, Jen, Clark, Lynn, Watson, Kevin, & Sóskuthy, Márton. (2021). Systematic co-variation of monophthongs across speakers of New Zealand English. *Journal of Phonetics*, *88*, 101096.

Campbell-Kibler, Kathryn. (2021). The Cognitive Structure Behind Indexicality: Correlations in Tasks Linking /s/ Variation and Masculinity. In L. Hall-Lew, E. Moore, & R. Podesva (eds.), *Social Meaning and Linguistic Variation: Theorizing the Third Wave* (pp. 127–150). Cambridge University Press.

Clark, Lynn, & Watson, Kevin. (2016). Phonological leveling, diffusion, and divergence:/t/lenition in Liverpool and its hinterland. *Language Variation and Change*, *28*(1), 31–62.

D'Onofrio, Annette. (2020). Personae in sociolinguistic variation. *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(6), e1543. https://doi.org/10.1002/wcs.1543.

Eckert, Penelope. (2019). The limits of meaning: Social indexicality, variation, and the cline of interiority. *Language*, *95*(4), 751–776.

Eckert, Penelope. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, *12*(4), 453–476.

Eckert, Penelope, & Labov, William. (2017). Phonetics, phonology and social meaning. *Journal of Sociolinguistics*, *21*(4), 467–496.

Fromont, Robert, & Hay, Jen. (2008). ONZE miner: The development of a browser-based research tool. *Corpora*, *3*(2), 173–193.

Gregersen, Frans, & Barner-Rasmussen, Michael. (2011). The logic of comparability: On genres and phonetic variation in a project on language change in real time. *Corpus Linguistics and Linguistic Theory* 7–1:7–36.

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

Guy, Gregory R., Oushiro, Livia, & Mendes, Ronald B. (2022). Indexicality and coherence. In V. B. Karen & R. G. Gregory (eds.), *The Coherence of Linguistic Communities* (pp. 53–68). Routledge.

Harrington, Jonathan, Palethorpe, Sallyane, & Watson, Catherine. (2000). Does the Queen speak the Queen's English? *Nature*, *408*(6815), 927–928.

Horvath, Barbara M. (1985). Variation in Australian English. The Sociolects of Sydney. *Cambridge Studies in Linguistics London, 45*, 1–200.

Labov, William. (2001). *Principles of Linguistic Change: Social Factors, Volume 2*. Blackwell Publishers.

Maclagan, Margaret, & Hay, Jen. (2007). Getting fed up with our feet: Contrast maintenance and the New Zealand English "short" front vowel shift. *Language Variation and Change*, *19*(1), 1–25.

Maclagan, Margaret, Watson, Catherine I., Harlow, Ray, King, Jeanette, & Keegan, Peter. (2009). /u/fronting and/t/aspiration in Māori and New Zealand English. *Language Variation and Change*, *21*(2), 175–192.

McKenzie, Jayne. (2005). 'But He's Not Supposed to See Me in My Weeding Dress!'-the Relationship between DRESS and FLEECE in Modern New Zealand English. *New Zealand English Journal*, *19*, 13–25.

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rickford, John, & Price, Mackenzie. (2013). Girlz II women: Age-grading, language change and stylistic variation. *Journal of Sociolinguistics*, *17*(2), 143–179.

Sankoff, Gillian. (2019). Language change across the lifespan: Three trajectory types. *Language*, 95(2), 197–229.

Sankoff, Gillian, & Blondeau, Helene. (2007). Language change across the lifespan:/r/in Montreal French. *Language*, 83(8), 560–588.

Sankoff, Gillian. & Wagner, Suzanne E. (2006). Age-grading in retrograde movement: The inflected future in Montréal French. *University of Pennsylvania Working Papers in Linguistics*, 12(2), 1–14.

Sheard, Elena. (2023). *Explaining language change over the lifespan: A panel and trend analysis of Australian English*. Unpublished doctoral thesis. Australian National University

Sheard, Elena, & Wilson Black, Josh. (2024). *Rotating Principal Components to explore change over the lifespan in New Zealand English monophthongs*. Paper presented at Eighteenth International Conference on Methods in Dialectology. Melbourne, Australia. July 1-5.

Silverstein, Michael. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3–4):193–229.

Tamminga, Meredith. (2019). Interspeaker covariation in Philadelphia vowel changes. *Language Variation and Change*, 31(2), 119–133.

Wagner, S. E. & Tagliamonte, S. A. (2017). What makes a panel study work? Researcher and participant in real time. In E. W. Susan & I. Buchstaller (eds.), *Panel Studies of Variation and Change* (pp. 213–232). Routledge.

Wagner, Suzanne E. (2012). Age grading in sociolinguistic theory. *Language and Linguistics Compass*, 6(6), 371–382.

Walsh, Liam, Hay, Jen, Bent, Derek, Grant, Liz, King, Jeanette, Millar, Paul, Papp, Viktoria, & Watson, Kevin. (2013). The UC QuakeBox Project: Creation of a Community-Focused Research Archive. *New Zealand English Journal*, 27, 20–32.

Wilson Black, Josh, Brand, James, Hay, Jen, & Clark, Lynn. (2023). Using principal component analysis to explore co-variation of vowels. *Language and Linguistics Compass*, 17(1), e12479.