Joshua Wilson Black*, Jennifer Hay, Lynn Clark and James Brand

The overlooked effect of amplitude on within-speaker vowel variation

https://doi.org/10.1515/lingvan-2022-0086
Received July 12, 2022; accepted April 5, 2023; published online November 24, 2023

Abstract: We analyse variation in vowel production within monologues produced by speakers in a quiet, well-controlled environment. Using principal component analysis (PCA) and generalized additive mixed models (GAMMs), applied to a large corpus of naturalistic recordings of New Zealand English speakers, we show that the first formant of monophthongs varies significantly with variation in a speaker's relative amplitude. We also find that amplitude variation is used, potentially agentively, to mark the beginning and ending of topical sections within single-speaker monologues. These results have significant methodological consequences for the study of vocalic variation in the context of research on speaker style and language variation and change. While laboratory research has shown a connection between variation in F1 and amplitude in loud environments or with distant interlocutors, this has not been seen in quiet environments with unscripted speech of the sort often used in sociolinguistcs. We argue that taking account of this variation is an important challenge for both within-speaker investigation of stylistic covariation and across-speaker investigation. In the latter case we recommend, as a minimal step, the inclusion of a measure of relative amplitude within regression models.

Keywords: amplitude; monophthongs; covariation; Lombard effect; language variation and change

1 Introduction

The predominant methodology in sociophonetics is the extraction and analysis of vowel formants, in order to compare vowel productions across different groups of speakers. This methodology often incorporates some form of vowel space normalization to enable direct comparison across speakers while controlling for the effect of differences in vocal tract length (Adank et al. 2004). Such methodologies usually analyse many tokens of the target vowel(s) from each speaker to estimate their overall production patterns (e.g. Labov 2001; Stuart-Smith et al. 2017; Watson et al. 1998).

Simultaneously, work on speaker style and stance emphasizes the need to move beyond the individual sociolinguistic variable, and study how speakers use the full range of resources available to them over the course of a conversation (Eckert 2012, 2016; Eckert and Labov 2017; Podesva 2008; Tamminga 2021). This work posits that variants co-vary with one another in "clusters" associated with different linguistic styles or stances. For example, Podesva (2008) shows, by manual coding and inspection, that extreme variants of *t*/d releases, intonational patterns, and falsetto voice qualities co-vary in stylistically meaningful ways within a single speaker.

The work reported in this paper started with exploration of this kind of within-speaker covariation, asking if it is possible to identify, bottom-up, how the production of vowels co-varies across the course of speakers' monologues and whether we can identify structure that is visible in the co-patterning of many variables, when viewed across many recordings.

***Corresponding author: Joshua Wilson Black**, New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand, E-mail: joshua.black@canterbury.ac.nz. https://orcid.org/0000-0002-8272-5763
Jennifer Hay and Lynn Clark, New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand; and Department of Linguistics, University of Canterbury, Christchurch, New Zealand. https://orcid.org/0000-0003-3282-6555 (L. Clark)
James Brand, Institute of Czech Language and Theory of Communication, Charles University, Praha, Czech Republic. https://orcid.org/0000-0002-2853-9169

Our investigation revealed an unexpectedly large source of covariation in our data driven by the relationship between vowel formants and fluctuations in speaker amplitude. We argue that this variation is directly relevant to the analysis of stylistic within-speaker covariation, especially as amplitude appears to be actively used by speakers to demarcate variation in topic. We also argue that this has significant consequences for the common methodology of extracting formants for cross-speaker comparison, and suggest that doing so without regard to amplitude variation is problematic.

1.1 Speaker style

Our initial research question concerned the covariation of vowels in the creation of speaker style. When a speaker uses a more innovative NURSE vowel (fronted and raised), for example, do we also see a more innovative GOOSE vowel (fronted and lowered) in the same stretch of speech?[1] We were interested in how vowels may work together across the course of a conversation or monologue. This relates to a literature which predicts that such covariation should exist (Eckert and Labov 2017; Podesva 2008), but which so far lacks a systematic large-scale statistical attempt to reveal such structure in a suitably large corpus of speech.

If within-speaker covariation is discoverable, its relation to patterns of covariation we have observed across speakers would become a pressing question. Brand et al. (2021) looked at patterns of covariation across speakers, finding, for example, that if a speaker has a more advanced KIT vowel with respect to the NZ short front vowel shift, they are also more likely to have a more raised DRESS. They identified three sets of vowels that vary systematically across speakers. We asked whether, if we found patterns of vowels that co-vary within the speech of individuals, these would be the same patterns found across individuals? This seemed possible, given the claim that variation within individual speakers can animate and reflect broader societal patterns of language variation and change (Eckert 2019).

We do not review the literature on speaker style in any depth here, because, although it was our starting place, it is not the focus of the current paper. Rather, we report a major source of covariation uncovered in our search for stylistic variation – the covariation of formants with amplitude. We report this background as part of a commitment to increased transparency in exploratory research and to emphasize the methodological importance of the covariation of formants with amplitude for future work on speaker style. We now turn to literature on amplitude and its relationship with F1.

1.2 Amplitude variation and formants

A well-known and well-studied effect – the Lombard effect – leads to an increase in speech intensity when talkers are speaking in a noisy environment (Lombard 1911). It is well documented that, when speaking in noise, speakers not only adjust amplitude, but a variety of other features, including increased word duration, F0, and effects on spectral tilt (Brumm and Zollinger 2011; Cooke and Lu 2010; Draegert 1951; Junqua 1993; Tartter et al. 1993; Van Summers et al. 1988; Zhao and Jurafsky 2009). This literature has also shown effects on formants, although these can be more variable. When formant effects are reported, they show increasing F1 in louder environments (Van Summers et al. 1988), and decreasing F2 (Pisoni et al. 1985).

There are related effects observed in tasks involving manipulation of vocal effort. For example, Liénard and Di Benedetto (1999) varied the distance of an interlocutor from 0.4 to 6 m, and found that F1 was highly correlated with vocal effort, but not F2. Koenig and Fuchs (2019) review a variety of literature on "loud speech", which suggests that louder or shouted speech may involve more "open articulatory postures" than typical speech, which is more extreme for low vowels than higher vowels (cf. Schulman 1989). The literature reviewed by Koenig and Fuchs suggests that, while speakers tend to show higher F1 values in loud speech, the magnitude of this varies

[1] We follow Wells (1982) in using lexical set words to refer to sets of vowels in the same phonemic category. Thus DRESS refers to all vowels sharing the same vowel phoneme as the word *dress*.

across speakers and studies. Their own study, manipulating interlocutor distance with German speakers, concludes that effects of loud speech on formants can vary across speakers and across vowels.

This link between F1 and loud speech is further clarified in work on articulatory changes in loud conditions. Scobbie et al. (2012), for example, point out that the increased intensity associated with Lombard speech is partly enabled by an increased opening of the mouth. This, in turn, affects vowel quality, with vowels having a lower jaw and tongue position, both of which contribute to a raised F1. They recorded a single speaker of Scottish English in quiet and in noise, who produced increased intensity in noise, and a higher F1 for all vowels but/i/. Their analysis of lip movement shows increased lip opening in the Lombard speech, suggestive of a lowered jaw, and ultrasound tongue imaging shows general evidence of tongue lowering, although this is not equal across all vowels. This supports a variety of other studies which also show increased jaw and lip opening in Lombard speech (Garnier et al. 2006; Kim et al. 2005; Šimko et al. 2016).

The Lombard and the "loud speech" literature suggest a clear relationship between amplitude/loudness and vowel formants, at least in cases where the amplitude variation is extreme. We are not aware of literature that looks at normal variation in amplitude over the course of speech in quiet environments, and the effects that such variation may have on observed F1. Sociolinguistic studies that investigate vowel variation will often include controls for factors such as preceding and following phonological environment, or word frequency (e.g. Hay et al. 2015). They may also control for speech rate (e.g. Podesva et al. 2012; Villarreal and Clark 2022), as faster articulation rates have been linked with more reduced vowel spaces (Fourakis 1991) and with undershoot of vowel targets (see Van Son and Pols 1992). However we are not aware of any studies that take the role of variation in amplitude into account.

1.3 Overview

In a principal component analysis (PCA) we observe substantial covariation of F1, which we link to amplitude variation (Section 3). We subsequently focus an analysis directly on exploring the substantial covariation observed between formants and amplitude (Section 4), and the degree to which amplitude may be used to indicate topic structure (Section 5). Substantial supplementary materials, along with anonymized data, are available.[2]

2 Data

We analyse monologues produced by New Zealand English (NZE) speakers in the QuakeBox (QB) corpus (Clark et al. 2016; Walsh et al. 2013). The corpus comprises a database of audio and video recordings, where members of the public were invited to record stories of their experiences of the earthquakes that hit Christchurch, New Zealand, in 2010–2011. The stories were recorded in a sound-protected environment. Participants wore a head-mounted microphone and were prompted with "tell us your earthquake story". The recordings took place in 2012 and were forced-aligned using the HTK aligner (Young et al. 2002), with the corpus stored on a LaBB-CAT instance (Fromont and Hay 2008).

The corpus was queried for all instances of the following monophthongs: DRESS, FLEECE, FOOT, GOOSE, KIT, LOT, NURSE, SCHWA, START, STRUT, THOUGHT, TRAP. We exclude speakers who did not grow up in New Zealand. Before filtering, the data has 288 speakers and 486,073 monophthong tokens. First and second formants at the vowel midpoints, pitch, and amplitude were extracted using LaBB-CAT's interface with Praat (Boersma and Weenink 2018). Amplitude is measured as the maximum amplitude in the word. Amplitude extraction was carried out at the word level because it requires spans of at least 0.064 s. This would cause unnecessary data loss if carried out at the vowel

[2] Code is shared in the form of four R Markdown documents, one each for the data filtering, data representation, PCA, and modelling stages of the paper. These will be referred to by "SMx, §y", where "x" is the number of the R Markdown document (visible in the filename) and "y" is the section number within the document. These markdowns, supplementary scripts, data, and fit models are available at https://osf.io/m8nkh/.

Table 1: Number of speakers by demographic category and length of recording after filtering.

Variable	Factor level	Number of speakers
Age category	18–25	41
	26–35	19
	36–45	41
	46–55	68
	56–65	55
	66–75	34
	76–85	11
	85+	3
	NA	8
Gender	Female	186
	Male	94
Ethnicity	NZ European	257
	NZ Māori	8
	NZ mixed ethnicity	9
	Other	6
Length	Short (<10:00)	162
	Medium (10:00–20:00)	84
	Long (>20:00)	34

Table 2: Vowel token counts after filtering.

Vowel	Tokens
DRESS	23,453
KIT	18,117
FLEECE	17,404
STRUT	16,241
LOT	12,681
TRAP	12,678
THOUGHT	10,101
GOOSE	7,704
START	6,948
NURSE	6,082
FOOT	3,998
Total	135,407

level. The variables are *z*-scored by speaker and by vowel to put all vowels and speakers on the same scale. Table 1 and Table 2 present the composition of the data set after filtering to remove stopwords, unstressed tokens (including SCHWA), and outliers (see SM1).

3 Principal components analysis

3.1 Methods

We use PCA to explore within-speaker covariation in the first and second formants of NZE monophthongs. The principal components (PCs) which PCA produces reveal structure in the original variables of the data set. For instance, Brand et al. (2021) used PCA to investigate the historical development of NZE, showing that one across-speaker ingredient of variation of NZE monophthongs is the increase of KIT F1 and TRAP F2 while FLEECE F1 decreases (and vice versa). To explore within-speaker variation we divided monologues into intervals of fixed length and applied PCA to the resulting data set.

3.1.1 Interval representation

Our approach divided each speaker's monologue into intervals of fixed duration, within which we could take the mean value for each variable. We then asked whether vowels co-vary randomly across the different intervals, or whether we could find structure that might represent systematic vowel covariation. PCA requires complete observations. That is, it cannot handle missing data. Consequently, each interval required an F1 and F2 value for each vowel (see Wilson Black et al. 2023).

Selecting the interval length required a trade off between stability and resolution. Formant measurements of vowels have greater variance than the shifts in vowel space location which we are attempting to investigate within the speech of a single speaker. Consequently, if an interval, by chance, consisted of two or three tokens with extreme values, our mean value for the interval would be similarly extreme while not representing meaningful variation in central tendency. Moreover, the vast difference in token counts for each vowel (Table 2) means that interval lengths appropriate for one vowel were insufficient for another. Both considerations pushed towards longer interval lengths. However, the longer the interval length, the harder it is to detect very local stylistic covariation, as this is likely to occur over short time scales.

We decided upon an analysis of intervals of 240 and 60 s. The 240 s intervals reliably captured sufficient data from each variable. The 60 s intervals are significantly shorter, but required imputation of values in intervals with missing data. Data for FOOT was too sparse to be included in the analysis. Once we generated intervals, we assigned values to them by taking the mean for each variable (see SM2).

For intervals that do not contain data for up to three vowel types, we topped up any variables with fewer than three tokens within an interval with additional tokens representing the overall speaker mean. This ensured complete observations for the remaining intervals and enabled us to preserve intervals with sparse data for some vowels without fear of including extreme values which do not represent real changes in the overall vowel space. After imputation, we were left with 2,835,60 s intervals and 673,240 s intervals (see SM2, §4).

Imputation affects around 1.6 % of the data at an interval length of 240 s and 33.5 % of the data at the 60 s interval length. The two interval lengths thus offer a way to check whether imputation affected our results. Moreover, given that the imputation process tends to reduce the number of intervals with extreme values, by bringing them closer to the mean value for the vowel, imputation is more likely to reduce covariation than it is to introduce spurious covariation.

Figure 1 displays an example speaker, both with their original formant and amplitude data (as points) and their interval values for both formant and amplitude data at both interval lengths. The fill colour denotes the interval value, with higher values shaded more yellow and lower values more blue. Intuitively, our PCA analysis asked whether, and in what way, the colour patterns in this plot are related to one another.

3.1.2 PCA methodology

PCA reveals underlying structure in the correlation matrix of a data set. By default, the notion of correlation used by PCA is Pearson's. However, we allowed for non-linear relationships by applying a rank transformation to our data. That is, we replaced Pearson correlation with Spearman correlation (cf. Aluja-Banet et al. 2018: §3.4).

We carried out PCA using the R function stats::prcomp (R Core Team 2023). We applied the function separately to the rank transformed amplitude and formant measures from the 60 and 240 s intervals (see SM3).

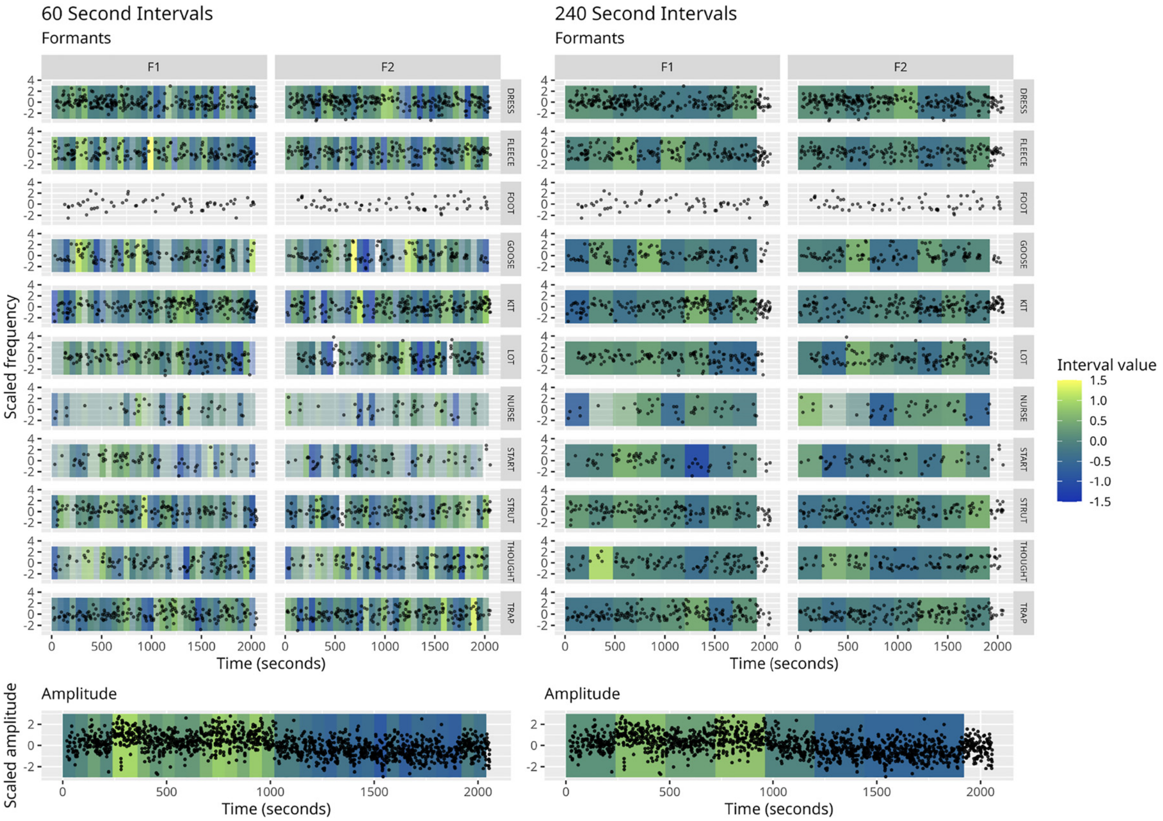

Figure 1: Data from example speaker (QB_NZ_F_369) divided into 60 and 240 s intervals. For the lower panels, fill indicates the mean of by-speaker *z*-scored amplitude. For the upper panels, fill indicates the mean of by-speaker *z*-scored formant frequency after imputation. Due to data filtering, no interval values are given for FOOT.

3.2 Results

We report a PCA analysis of the 60 and 240 s intervals including both formant and amplitude data. Our initial analysis, carried out when looking for vocalic covariation driven by style, omitted amplitude and showed substantial covariation in F1 (SM3, §3). We subsequently added amplitude and repeated the analysis, which revealed that amplitude was associated with the F1 covariation (SM3, §5).[3]

Figure 2 consists of two variable plots. The axes represent the first and second PCs. The arrows correspond to the original variables. Variable plots show how the original variables correspond to the PCs. The colour gradient and length of each arrow represents the strength of association between an original variable and the relevant PCs.[4]

The first PC, at both interval lengths, contains a relationship between distinct vowels. All first formants pattern together, along with amplitude, on PC1.[5]

[3] We also repeated the analysis with a corpus balanced so that each speaker contributes the same amount of data to the PCA in order to demonstrate that speakers with long monologues are not skewing our results (SM3, §4).

[4] As an illustration, the second PC of the 60 s interval PCA is easiest to interpret. This PC is represented by the *y*-axis of the upper plot. We see two of the original variables are strongly associated with PC2 – the first and second formants of GOOSE. Holding all other PCs constant, increases in PC2 correspond to an increased value of GOOSE F2 and a decreased value of GOOSE F1. Since this PC is not a relationship *between* vowels, we will not discuss it any further.

[5] We subsequently used permuted versions of the data to provide a simulated estimate of the results of PCA if all temporal structure was removed from the data. This analysis supports our view that our PCA is detecting non-random covariation (SM3, §5.4).

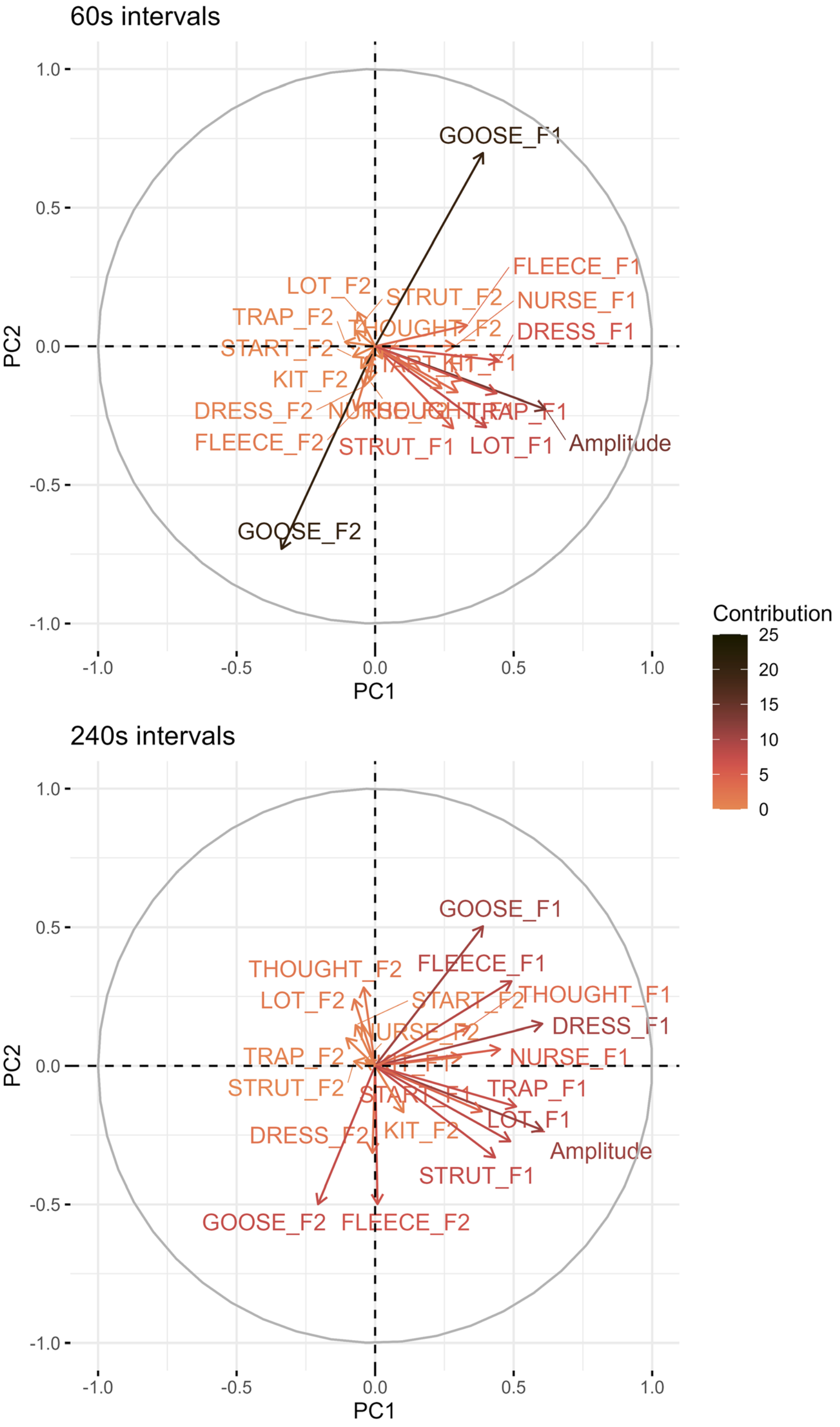

Figure 2: Variable plots for PCA analysis of 60 and 240 s intervals with both formant and amplitude data. Position on the *x*-axis indicates the contribution of a variable to PC1, while position on the *y*-axis indicates contribution to PC2. Colour indicates the magnitude of contribution.

4 Generalized additive mixed model of F1

To empirically support our interpretation that systematic covariation in F1 is driven by variation in amplitude, we fit a generalized additive mixed model (GAMM; see Sóskuthy 2017) using the mgcv package

Table 3: GAMM independent variables and smooths.

Variable name	Description
participant_gender	Gender of participant (factor, 2 levels).
Vowel	NZE monophthong (factor, 11 levels).
s(speaker_scaled_time, by=vowel)	Thin plate regression splines on time through speaker monologue scaled to (0, 1) for each level of vowel.
s(speaker_scaled_amp_max, by=vowel)	Thin plate regression splines on *z*-scored speaker amplitude for each level of vowel.
s(speaker_scaled_art_rate, by=vowel)	Thin plate regression splines on *z*-scored speaker utterance articulation rate for each level of vowel.
s(speaker_scaled_pitch, by=vowel)	Centred thin plate regression splines on *z*-scored speaker pitch for each level of vowel.
s(speaker, bs="re")	Random intercepts for each speaker (factor, 171 levels).
s(speaker, vowel, bs="re")	Random slopes on vowel for each level of speaker.

Table 4: *p* values estimated by model comparison.

Variable	*p* value
speaker_scaled_time	2.215e-07
speaker_scaled_art_rate	<2e-16
speaker_scaled_amp_max	<2e-16
speaker_scaled_pitch	<2e-16

(Wood 2017). The model has first formant value, in hertz, as the dependent variable. The independent variables and associated smooth terms, along with the random effects, are presented in Table 3. This model, and a series of other models exploring the same question with different modelling strategies, are covered in supplementary material (SM4, §2).

We fit this model on the filtered data, before the introduction of intervals and imputation. The effect of amplitude assessed here is thus more local to the vowel than the effect as captured by PCA applied to intervals. This also means that foot is included in the model. Table 4 presents p values for the key terms.

Figure 3 depicts the smooths on F1 for each vowel type. We see strong evidence for increases in F1 accompanying increases in amplitude. This effect is stronger for some vowels than for others. The estimates for LOT and TRAP cross one another as amplitude increases. We also see clusters of vowels which are distinguishable in F1 at some amplitudes but not others. Concerningly, this would lead an analyst to reach different conclusions about the configuration of the vowel space at higher versus lower amplitudes.[6]

4.1 The effect of amplitude on vowel spaces

Articulation rate is widely understood to affect vowel space area, and to be worth controlling for when modelling formant values. Amplitude variation is usually ignored. In order to consider their respective effects on the vowel space of speakers in the QuakeBox corpus, we fit a GAMM of F2 using the same model structure used in the previous section. We then generated F1 and F2 predictions across the range of scaled amplitude and articulation rate values and plotted them within the vowel space (Figure 4).

[6] For plots of the control variables, see SM2, §3.2.

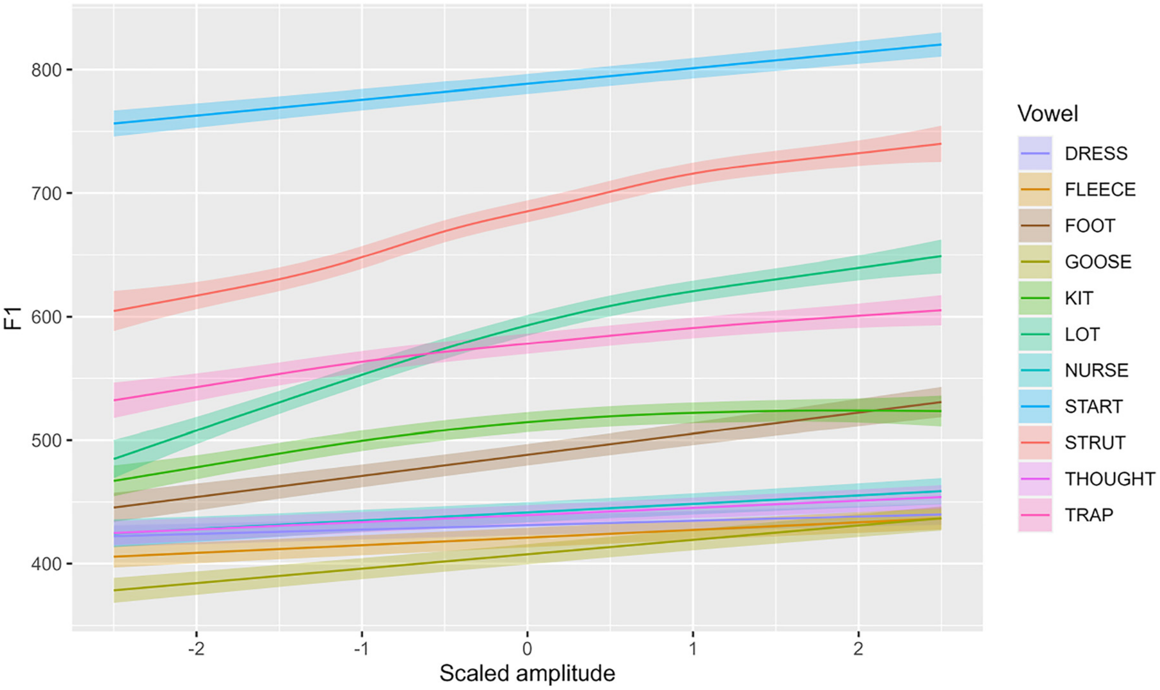

Figure 3: Model predictions of F1 for each vowel given scaled amplitude assuming female speaker, halfway through monologue, with mean articulation rate and pitch. Model predictions were extracted using the itsadug package (van Rij et al. 2022).

Figure 4 shows that amplitude variation results in larger shifts to the vowel space than are produced by articulation rate. Furthermore, the vertical shifts of various vowels are quite different in magnitude and there is also strong lateral movement, especially in GOOSE and THOUGHT. These lead to differences in the character of the vowel space which cannot be corrected by simple vertical translation.

Figure 5 is an example plot from an R Shiny interactive which can be used to explore the effect of amplitude on judgements of vowel space similarity. The interactive enables users to select intervals which are high amplitude and those which are not, view their vowel spaces, and view the speaker vowel spaces which are most similar to the high amplitude interval, in the top row, and the low amplitude interval, in the bottom.

The top row of Figure 5 shows the three most similar overall vowel spaces to a high amplitude interval vowel space from a speaker, and the bottom row shows the most similar overall vowel spaces to a low amplitude interval vowel space from the same speaker. In both cases Lobanov 2.0 normalization has been applied (Brand et al. 2021). The high amplitude interval shows significant lowering in the vowel space of THOUGHT, LOT, and STRUT compared to the low amplitude interval. Similarity is measured by mean Euclidean distance (see SM3, §6). The takeaway is that different speakers are picked out in both rows.

Consideration of speakers in the Shiny reveals that the specific speakers they appear most similar to will depend on whether their high amplitude or their low amplitude vowel space is being considered. This is true whether or not the vowel spaces are normalized.[7]

Taken together, Figures 4 and 5 indicate a concerning effect of amplitude on speaker vowel spaces. The former shows that this effect is larger than that of articulation rate, which is already taken seriously by sociolinguists. The latter suggests the effect is enough to undermine rankings of speakers by vowel space similarity.

[7] Kendall's Tau was used to determine whether the lists of most similar speakers for each speaker's highest and lowest amplitude intervals is correlated. The vast majority of correlations, with normalization applied, are between 0 and 0.1, with none above 0.2 (SM3, §6.3.2.1).

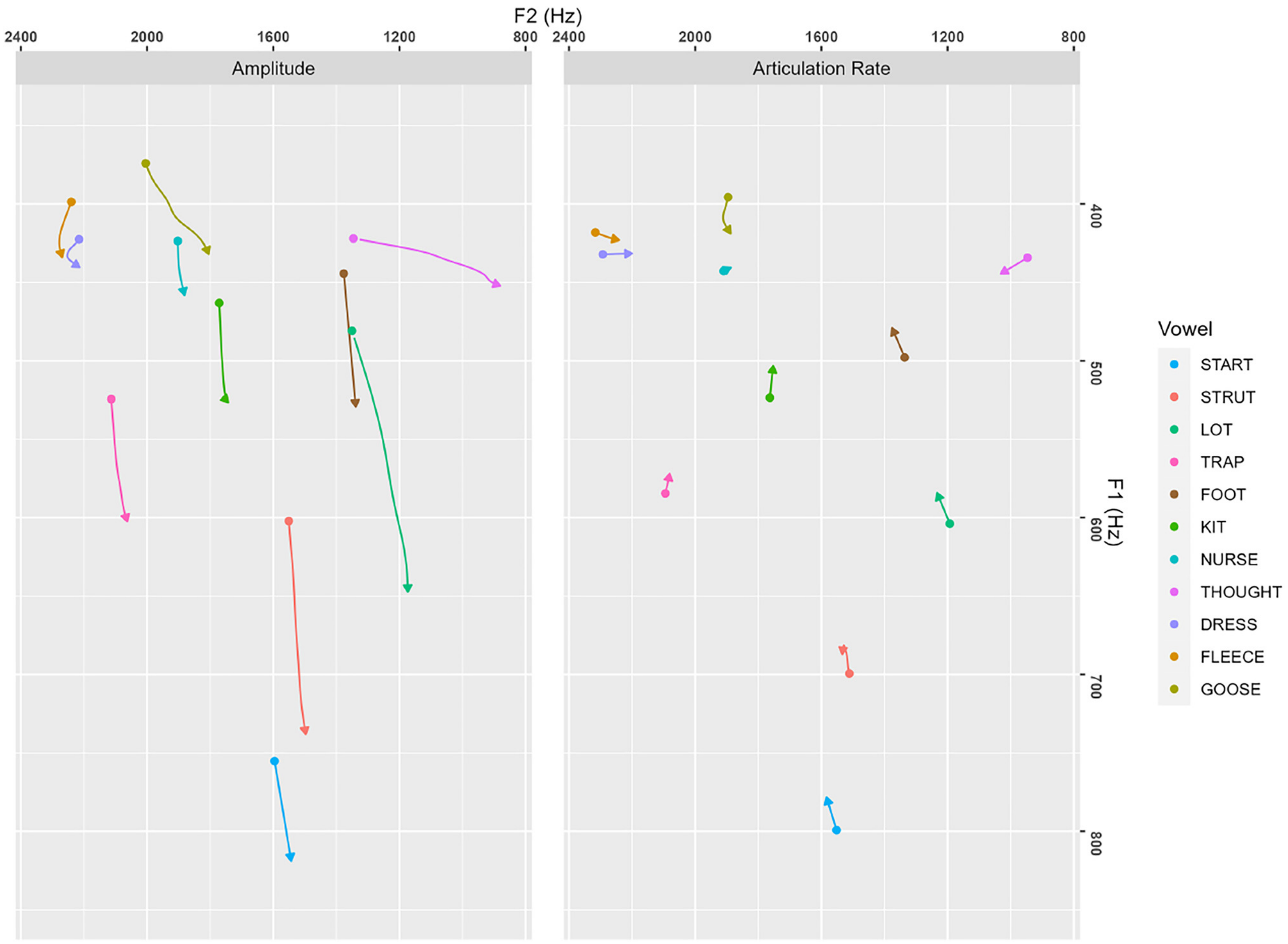

Figure 4: Model predictions of the vowel space given scaled amplitude and articulation rate, assuming female speaker halfway through monologue, speaking at mean pitch. Points indicate the *z*-scored F1 and F2 predicted at a value of −2.5 for amplitude (*left*) and articulation rate (*right*). For animated and interactive versions of this plot, see SM4, §2.3, Figures 2.36 and 2.37 (https://nzilbb.github.io/amp_f1_public/supplementary_material/SM4_models.html#fig:animated-plot).

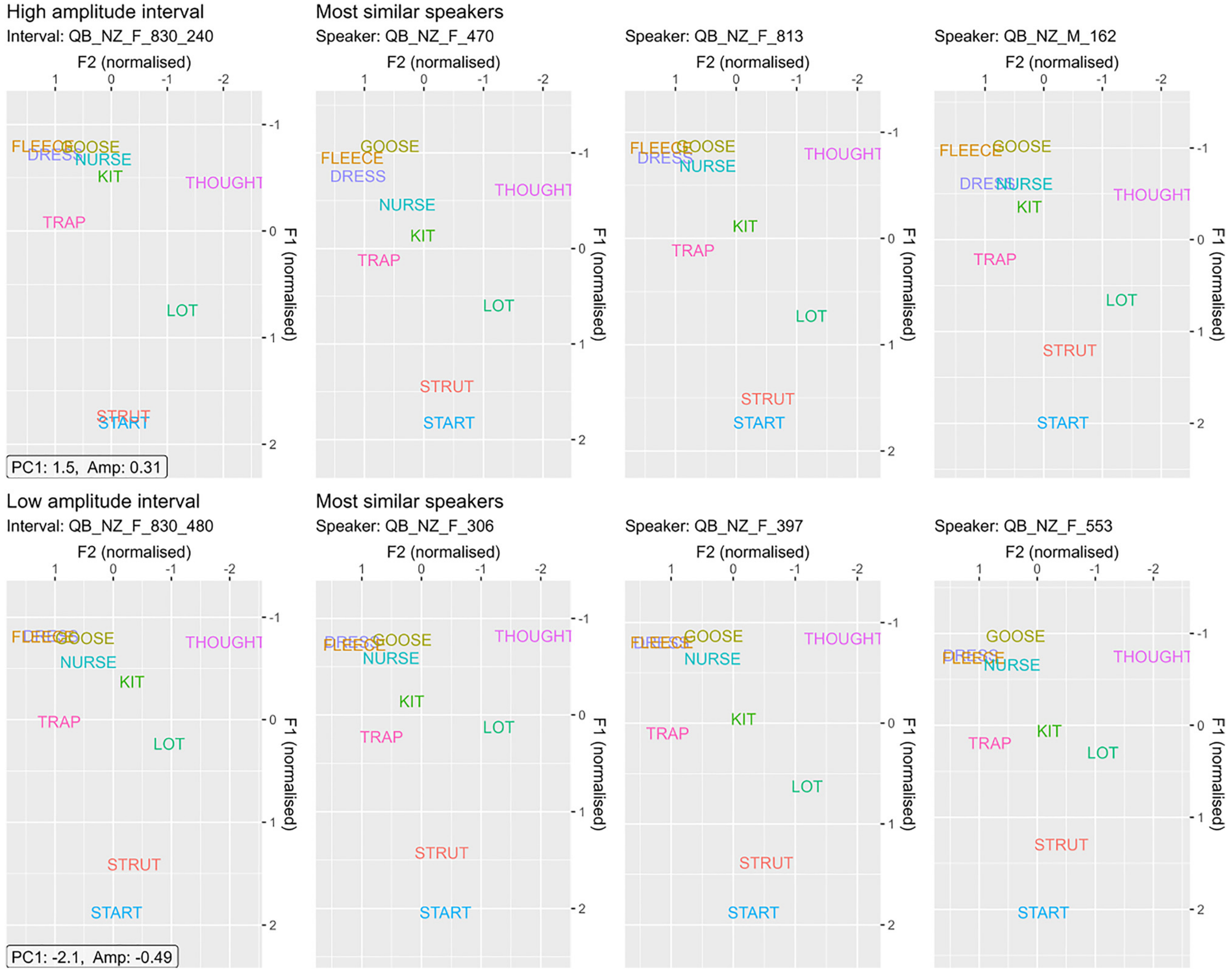

Figure 5: An example plot from the Shiny interactive for viewing high and low amplitude intervals from a given speaker (in this case QB_NZ_F_830) and the speakers with the most similar vowel spaces in the corpus. View the interactive at https://nzilbbshiny.canterbury.ac.nz/app/vs_similarity_shiny/.

5 Amplitude and topical subsections

Amplitude in the monologues affects a speaker's F1. But is amplitude itself a simple effect of vocal fatigue, or is it manipulated in a potentially agentive fashion? To examine this, we exploited the fact that the monologues are organized into subtopics. The QuakeBox corpus contains topical annotations for 193 speakers after filtering.

We tracked speaker amplitude at different parts of topics by simply dividing each topic into thirds: beginning, middle, and end. We then filtered our topics to ensure that there are at least five tokens in each part of each topic to enable a meaningful mean value to be taken for each.

We defined a linear mixed model, allowing the model to fit the beginning, middle, and end of the topic independently (SM4, §3.2). The model has scaled amplitude as its dependent variable, with topic part, time through monologue on a scale from −0.5 to 0.5, and scaled pitch as fixed effects. Random slopes were fit for each topical section in the corpus and for each speaker. Setting the time through monologue variable to run between −0.5 and 0.5 means that we did not need to fit a random intercept for each speaker.

In an exploratory spirit, we interpret *t* values as "expressions of surprise" (cf. Baayen et al. 2017: 227). As the magnitude of *t* values increases from around 2, which in simple models corresponds to something like a 0.05 *p* value, we are increasingly "surprised". Table 5 shows a clear decrease due to time through monologue. We also see a surprising decrease in amplitude at the end of topical subsections and an increase at the beginning.[8]

We thus have evidence that amplitude differences are being used to mark position within topical subsections of monologues and that this is distinct from variation in pitch. This is a starting point for further investigation of the sociolinguistic use of amplitude variation.[9]

6 Discussion

6.1 Natural amplitude variation in quiet environments correlates with F1

As outlined above, there are clear results in the experimental literature on speech production that link large changes in amplitude, either due to a loud environment or a distant interlocutor, to variation in F1. This is well understood as linking a more open articulation setting to louder speech, which affects resonant frequencies of the vocal tract, and thus F1, with more extreme effects observed for low vowels (cf. Koenig and Fuchs 2019). Our results do show different magnitudes across different vowels, but these are not straightforwardly mapped to

Table 5: Estimates and *t* values for topic part model. Estimates are in terms of by-speaker *z*-scored amplitude. The topic_part estimates are estimated mean values, while the speaker_scaled_time estimate is a gradient term, indicating the reduction in by-speaker *z*-scored amplitude over the course of a monologue.

Variable	Estimate	*t* value
topic_part (start)	0.059	5.206
topic_part (middle)	−0.010	−0.789
topic_part (end)	−0.072	−4.676
speaker_scaled_time	−0.203	−4.533
speaker_scaled_pitch	0.542	152.190

[8] We further test whether these values are surprising by generating fake topics from the data, rerunning the analysis 1000 times, and collecting the distributions of the model coefficients. The *t* values we obtained for `speaker_scaled_time` and `topic_part (start)` are in the 5 % tails of the distribution from fake topics, while the value for `topic_part (end)` is completely outside the distribution. For more details see SM4, §3.4.4.

[9] For a GAMM approach to the same question, see SM4, §3.3.

vowel height. Our lowest vowel, START, for example, does not show a particularly strong effect, whereas the shorter low vowels STRUT and LOT show the biggest effect. Amongst the highest vowels, FLEECE and GOOSE appear identical in height at higher intensities but are well separated in F1 at lower intensities.

Our speakers show a range of amplitude variation across the recordings. When examined as a mean over 60 s intervals, as in our PCA, the mean speaker range between intervals is 4 dB–5 dB (i.e. this is the range visualized in the Shiny app). When measured at the word level (as in our GAMMs), the mean observed range is in the order of 24 dB. This is a substantial degree of variation, likely comprising variation related to the overall structure, as well as more local variation to do with the role that the word is playing within the local phrasal structure.

By comparison, speakers in Van Summers et al. (1988) show less than a 10 dB difference in their wordlist reading between quiet and speaking in 100 dB noise. Liénard and Di Benedetto (1999) also show a comparable difference in isolated vowel production in a quiet environment between a close interlocutor and a distant one. Both studies show some correlation with F1.

This suggests that, when we move away from isolated or controlled speech materials towards unscripted communicative monologue, the degree of amplitude variation across the monologue can actually be as big as, or bigger than, that seen in experimental work, and with substantial consequences. This is not just a phenomenon observed in environments requiring prototypically “loud” speech.

The magnitude of the Lombard effect itself has been seen to vary across different types of task and communicative conditions (Villegas et al. 2021). An interesting question for future work, then, is whether the degree of amplitude variation, and subsequent F1 covariation observed here, is also observed in other types of speech contexts, such as reading or conversation.

6.2 Amplitude variation can be related to topic structure

What causes the amplitude variation within the monologues? The speakers were recorded with a head-mounted microphone, so there will not be substantial within-speaker variation due to changing microphone proximity. And this is not simply very local variation, relevant to the vowel in question. The PCA shows that there is more general amplitude variation that shows up when averaged across intervals of 60 or 240 s and that this relates to F1.

Our analysis of the patterning of this variation shows that amplitude declines over the course of a monologue. This would make sense as a reflection of general fatigue. However our analysis of topical structure shows that this is more agentive, with amplitude tending to be higher at the beginning of identifiable subtopics, and lower towards the end of each subtopic. Amplitude, and associated F1 (and likely other) variation, then, may be being used in a somewhat agentive fashion in the aid of signalling topic structure.

This links to a wider literature which explores how phonetics can be used to mark discourse structure (Pickering 2004; Local and Walker 2005; Zellers and Post 2012) and “manage the flow of talk” (Local 2007). Local and Walker (2005), for example, discuss the use of amplitude in turn taking, and the use of stand-alone *so*. When used to hold the floor, *so* has a higher amplitude than trailing *so*, which can occur turn-finally. There is much here to be explored about the ways in which amplitude may be being used to mark topic structure in these monologues.

6.3 Challenges for the study of within-speaker stylistic covariation of vowels

Our initial goal was to identify stylistically driven covariation in vowel production. Our analysis reveals that a very big source of covariation is F1 variation, linked to amplitude. This covariation would need to be well controlled in order to identify anything more locally stylistic. We note that the 60 s intervals in the PCA (but not the 240 s intervals) show a small amount of above-random structure in PC3 and PC4. It may be that once amplitude and F1 covariation is controlled for in a high PC, evidence of local stylistic covariation may still emerge in these lower PCs. This is a topic for future work.

This finding may also be relevant to other reports of phonetic covariation in the literature, particularly the literature on the degree to which articulations of vowels can prime each other, either within or across vowels. Villarreal and Clark (2022), for example, argue that there is a cross-vowel priming effect between DRESS, TRAP, and KIT, such that if a word is produced with a higher F1, a nearby word containing one of these vowels will also be produced with a higher F1. Our results suggest that this effect could be partly attributable to local variation in amplitude.

6.4 Challenges for the study of across-speaker covariation in sociolinguistics

The sociophonetic literature on language variation and change seeks to document structured differences across speakers in terms of their vowel space. This literature seeks to control known factors that might affect vowel space measurements, such as articulation rate, but does not attend to amplitude. Section 4 clearly showed that the effect of amplitude can be much bigger than the effect of articulation rate, and that an analyst could draw quite different conclusions about a speaker's vowel space if they measure their vowel space at different amplitudes. This is particularly problematic, because recording environments will seldom be comparable. There is thus no straightforward way to assess the actual amplitude at which the speaker is talking.

As pointed out by a reviewer, this point is also relevant for new data collection and phonetic fieldwork with production studies on languages in which little or nothing is known about the vowel space. Ideally, a workflow for the collection of new data would allow for assessment of relative amplitude differences between speakers, and for developing some understanding where in their normal amplitude range a speaker is operating during fieldwork recordings.

We note that our data is recorded with head-mounted microphones, and reinforce the importance of this for any data set in which an analyst hopes to control for the observed covariation between amplitude and F1. If, as we assume, F1 is co-varying with actual variation in produced loudness (as opposed to distance from the microphone), then compensating for this effect in recordings that use, for example, a table microphone, will be further complicated by the difficulty of separating variation in loudness from variation in proximity to the microphone.

For analysis of existing recordings, if we are fitting regression models to formant data, a measure of the relative amplitude of the section of speech from which measurements are extracted, together with speaker intercepts in our models, may go some way towards removing noise in our data. However, if we have sampled different speakers speaking at different levels within their normal amplitude range, and want to directly compare them, the problem becomes non-trivial.

The common act of normalizing the data to make it comparable across speakers also becomes complex in this context. This scales the data to span comparable F1 ranges, but we now know that if we scale relative to a whole monologue produced by a speaker, some parts of that monologue will not be scaled around zero for that speaker, and the extent of this offset will vary across vowels.

6.5 Implications for listeners

A final question for future work is how this covariation is dealt with by listeners. Literature shows that many known sources of variation are routinely compensated for by listeners. For example, listeners compensate for local coarticulatory effects (Beddor et al. 2002), speech rate (Miller and Liberman 1979; Port 1976), talker differences (Johnson and Sjerps 2021; Ladefoged and Broadbent 1957), and predicted social characteristics of speakers (Johnson et al. 1999).

In addition to talker characteristics and talking rate, Sommers et al. (1994) also manipulated amplitude as a control variable in their experiments under the assumption that "varying overall amplitude does not alter properties of the speech signal that are important for distinguishing phonetic categories" (1994: 1315). They found that mixing amplitudes does not provide a drop in accuracy in perception in their task, concluding that “certain types of variability, such as overall amplitude, are either ignored or are more easily accommodated by the speech-perception system, perhaps at very early stages of perceptual analysis” (1994: 1320).

In terms of a potential influence of amplitude on vowel categorization, it is perhaps pertinent that New Zealand listeners have been shown to adjust their vowel boundaries in contexts where they might expect Lombard speech (Hay et al. 2017), and also when they believe a token is being sung rather than spoken (Gibson 2019). Gibson (2019) shows that New Zealand singers use a higher F1 in their singing than their speaking for the vowels DRESS and TRAP. In a listening task, listeners are more likely to hear an ambiguous vowel as TRAP if they believe it is being sung, consistent with them expecting a higher F1 boundary between the vowels in singing. As pointed out by Gibson, this effect may relate to stylistic differences in speaking and singing, but there may also be some contribution from the more open jaw typically used in singing, with F1 consequences both in production and perception.

Overall, while literature is scarce on this topic, the degree of compensation observed in other domains suggests that it is possible that listeners readily accommodate the kind of amplitude variation that we have reported, together with the concurrent F1 variation. It is also possible that lexical and contextual cues are in most cases sufficient to lead to good identification rates, in spite of the observed F1 variation. The exact consequences of this variation for listeners are an interesting question for future work.

7 Conclusions

The amplitude at which a monophthong is measured can result in large changes within a speaker’s vowel space and thus impact sociolinguistic inferences. Moreover, we find evidence that amplitude itself does not vary randomly, nor in a manner that is completely linked to vocal fatigue, but rather it can be used by speakers to mark topic structure within monologues. Readers are encouraged to explore the associated Shiny interactive, which provides an intuitive sense of the extent to which a single speaker’s vowel space can alter when measured across different amplitudes. Variation in amplitude is a significantly underappreciated source of variation, which needs to be taken into consideration in studies comparing vowel acoustics both within and across speakers.

Acknowledgements: We thank the anonymous reviewers for their careful attention to this article and supplementary materials and for their many helpful comments. The QuakeBox project was a collaboration between UC CEISMIC (now the UC Arts Digital Lab) and NZILBB at the University of Canterbury, both of whom provided considerable support for the project. We thank all of the staff, students, research assistants, and community groups who supported the project. We are particularly grateful to Robert Fromont for his work programming LaBB-CAT – the corpus analysis tool used for this project. Our gratitude is also due to audiences at SocioPhonAus 2022 and the New Zealand Linguistics Society Conference 2022. Funding for this work was provided by the Marsden Fund (Grant: 20-UOC-064).

References

Adank, Patti, Roel Smits & Roeland Van Hout. 2004. A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116(5). 3099–3107.

Aluja-Banet, Tomas, Alain Morineau & Gaston Sanchez. 2018. Principal component analysis for data science. Available at: https://pca4ds.github.io.

Baayen, Harald, Shravan Vasishth, Reinhold Kliegl & Bates Douglas. 2017. The cave of shadows: Addressing the human factor with generalized additive mixed models. *Journal of Memory and Language* 94. 206–234.

Beddor, Patrice Speeter, James D. Harnsberger & Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30(4). 591–627.

Boersma, Paul & David Weenink. 2018. Praat: Doing phonetics by computer, version 6.0.37 [Computer program]. Available at: https://www.praat.org.

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

Brumm, Henrik & Sue Anne Zollinger. 2011. The evolution of the Lombard effect: 100 years of psychoacoustic research. *Behaviour* 148(11–13). 1173–1198.

Clark, Lynn, Helen MacGougan, Jennifer Hay & Liam Walsh. 2016. 'Kia ora. This is my earthquake story". Multiple applications of a sociolinguistic corpus. *Ampersand* 3. 13–20.

Cooke, Martin & Youyi Lu. 2010. Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *Journal of the Acoustical Society of America* 128(4). 2059–2069.

Draegert, G. L. 1951. Relationships between voice variables and speech intelligibility in high level noise. *Communication Monographs* 18(4). 272–278.

Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41. 87–100.

Eckert, Penelope. 2016. Variation, meaning and social change. In Nikolas Coupland (ed.), *Sociolinguistics: Theoretical debates*, 68–85. Cambridge: Cambridge University Press.

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

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

Fourakis, Marios. 1991. Tempo, stress, and vowel reduction in American English. *Journal of the Acoustical Society of America* 90(4). 1816–1827.

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

Garnier, Maëva, Lucie Bailly, Marion Dohen, Pauline Welby & Lœvenbruck Hélène. 2006. An acoustic and articulatory study of Lombard speech: Global effects on the utterance. In *Ninth international conference on spoken language processing*. INTERSPEECH.

Gibson, Andy. 2019. *Sociophonetics of popular music: Insights from corpus analysis and speech perception experiments*. Christchurch: University of Canterbury PhD thesis.

Hay, Jennifer, Janet B. Pierrehumbert, Abby J. Walker & Patrick LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139. 83–91.

Hay, Jennifer, Ryan Podlubny, Katie Drager & Megan McAuliffe. 2017. Car-talk: Location-specific speech production and perception. *Journal of Phonetics* 65. 94–109.

Johnson, Keith & Matthias J. Sjerps. 2021. Speaker normalization in speech perception. In Jennifer S. Pardo, Lynne C. Nygaard, Robert E. Remez & David B. Pisoni (eds.), *The handbook of speech perception*, 145–176. Hoboken, NJ: Wiley-Blackwell.

Johnson, Keith, Elizabeth A. Strand & Mariapaola D'Imperio. 1999. Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics* 27(4). 359–384.

Junqua, J. C. 1993. The Lombard reflex and its role on human listeners and automatic speech recognisers. *Journal of the Acoustical Society of America* 93(1). 510–524.

Kim, Jeesun, Chris Davis, Guillaume Vignali & Harold Hill. 2005. A visual concomitant of the Lombard reflex. In *Proceedings of the Auditory Visual Speech Processing conference*, 17–22. https://www.isca-speech.org/archive/avsp_2005/kim05_avsp.html (accessed 23 October 2023).

Koenig, Laura L. & Susanne Fuchs. 2019. Vowel formants in normal and loud speech. *Journal of Speech, Language, and Hearing Research* 62(5). 1278–1295.

Labov, William. 2001. *Principles of linguistic change, vol. 2: Social factors*. Malden, MA: Blackwell Publishers.

Ladefoged, Peter & Donald Eric Broadbent. 1957. Information conveyed by vowels. *Journal of the Acoustical Society of America* 29(1). 98–104.

Liénard, Jean-Sylvain & Maria-Gabriella Di Benedetto. 1999. Effect of vocal effort on spectral properties of vowels. *Journal of the Acoustical Society of America* 106(1). 411–422.

Local, John. 2007. Phonetic detail and the organisation of talk-in-interaction. In *Proceedings of the 16th icphs saarbrücken, germany*. https://pure.york.ac.uk/portal/en/publications/phonetic-detail-and-the-organisation-of-talk-in-interaction (accessed 23 October 2023).

Local, John & Gareth Walker. 2005. Methodological imperatives for investigating the phonetic organization and phonological structures of spontaneous speech. *Phonetica* 62(2–4). 120–130.

Lombard, Étienne. 1911. Le signe de l'élévation de la voix. *Annales des Maladies de L'Oreille, du Larynx, du Nez et du Pharynx* 37. 101–119.

Miller, Joanne L. & Alvin M. Liberman. 1979. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics* 25(6). 457–465.

Pickering, Lucy. 2004. The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes* 23(1). 19–43.

Pisoni, D., R. Bernacki, H. Nusbaum & M. Yuchtman. 1985. Some acoustic-phonetic correlates of speech produced in noise. In *IEEE international conference on acoustics, speech, and signal processing*: ICASSP'85, vol. 10, 1581–1584.

Podesva, Robert J. 2008. Three sources of stylistic meaning. In *Texas Linguistic Forum (Proceedings of the symposium about language and society – Austin 15)*, vol. 51, 134–143.

Podesva, Robert J., Lauren Hall-Lew, Jason Brenier, Stacy Lewis & Rebecca Starr. 2012. Condoleezza Rice and the sociophonetic construction of identity. In Juan Manuel Hernandez-Campoy & Juan Antonio Cutillas-Espinosa (eds), *Style-shifting in public: New perspectives on stylistic variation*, 65–80. Amsterdam: John Benjamins.

Port, Robert. 1976. Influence of tempo on the closure interval cue to the voicing and place of intervocalic stops. *Journal of the Acoustical Society of America* 59(S1). S41–S42.

R Core Team. 2023. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org.

Schulman, Richard. 1989. Articulatory dynamics of loud and normal speech. *Journal of the Acoustical Society of America* 85(1). 295–312.

Scobbie, James M., Joan K. Y. Ma & Joanna D. White. 2012. The tongue and lips in Lombard speech: A pilot study of vowel-space expansion. In *Casl working papers wp-21*. https://eresearch.qmu.ac.uk/handle/20.500.12289/3601 (accessed 23 October 2023).

Šimko, Juraj, Štefan Beňuš & Martti Vainio. 2016. Hyperarticulation in Lombard speech: Global coordination of the jaw, lips and the tongue. *Journal of the Acoustical Society of America* 139(1). 151–162.

Sommers, Mitchell S., Lynne C. Nygaard & B. David Pisoni. 1994. Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America* 96(3). 1314–1324.

Sóskuthy, Márton. 2017. Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction. *arXiv preprint*. Available at: https://arxiv.org/abs/1703.05339.

Stuart-Smith, Jane, Brian José, Tamara V. Rathcke, Rachel Macdonald & Eleanor Lawson. 2017. Changing sounds in a changing city: An acoustic phonetic investigation of real-time change over a century of Glaswegian. In Chris Montgomery & Emma Moore (eds.), *Language and a sense of place: Studies in language and region*, 38–64. Cambridge: Cambridge University Press.

Tamminga, Meredith. 2021. Social meaning and the temporal dynamics of sound changes. In Lauren Hall-Lew, Emma Moore & Robert J. Podesva (eds.), *Social meaning and linguistic variation: Theorizing the third wave*, 338–362. Cambridge: Cambridge University Press.

Tartter, Vivien, Hilary Gomes & Elissa Litwin. 1993. Some acoustic effects of listening to noise on speech production. *Journal of the Acoustical Society of America* 94. 2437–2440.

van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Hedderik van Rijn. 2022. Itsadug: Interpreting time series and autocorrelated data using GAMMs, version 2.4. Available at: https://CRAN.R-project.org/package=itsadug.

Van Son, Rob J. J. H. & Louis C. W. Pols. 1992. Formant movements of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America* 92(1). 121–127.

Van Summers, W., David B. Pisoni, Robert H. Bernacki, Robert I. Pedlow & Michael A. Stokes. 1988. Effects of noise on speech production: Acoustic and perceptual analyses. *Journal of the Acoustical Society of America* 84(3). 917–928.

Villarreal, Dan & Lynn Clark. 2022. Intraspeaker priming across the New Zealand English short front vowel shift. *Language and Speech* 65(3). 713–739.

Villegas, Julián, Jeremy Perkins & Ian Wilson. 2021. Effects of task and language nativeness on the Lombard effect and on its onset and offset timing. *Journal of the Acoustical Society of America* 149(3). 1855–1865.

Walsh, Liam G., Jennifer Hay, Derek Bent, Jeanette King, Paul Millar, Viktoria Papp & Kevin Watson. 2013. The UC QuakeBox project: Creation of a community-focused research archive. *New Zealand English Journal* 27. 20–32.

Watson, Catherine I., Jonathan Harrington & Zoe Evans. 1998. An acoustic comparison between New Zealand and Australian English vowels. *Australian Journal of Linguistics* 18(2). 185–207.

Wells, John C. 1982. *Accents of English*, vol. 1. Cambridge: Cambridge University Press.

Wilson Black, Joshua, James Brand, Jen Hay & Lynn Clark. 2023. Using principal component analysis to explore co-variation of vowels. *Language and Linguistics Compass* 17. https://doi.org/10.1111/lnc3.12479.

Wood, Simon N. 2017. *Generalized additive models: An introduction with R*. Boca Raton, FL: CRC Press.

Young, Steve, Gunnar Evermann, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev & Phil Woodland. 2002. *The HTK book*. Cambridge: Cambridge University Engineering Department.

Zellers, Margaret & Brechtje Post. 2012. Combining formal and functional approaches to topic structure. *Language and Speech* 55(1). 119–139.

Zhao, Yuan & Dan Jurafsky. 2009. The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics* 37(2). 231–247.