Multidimensional acoustic variation in vowels across English dialects

James Tanner[a] **Morgan Sonderegger**[a] **Jane Stuart-Smith**[b]
[a]McGill University [b]University of Glasgow
james.tanner@mail.mcgill.ca
morgan.sonderegger@mcgill.ca
Jane.Stuart-Smith@glasgow.ac.uk

Abstract

Vowels are typically characterized in terms of their static position in formant space, though vowels have also been long-known to undergo dynamic formant change over their timecourse. Recent studies have demonstrated that this change is highly informative for distinguishing vowels within a system, as well as providing additional resolution in characterizing differences between dialects. It remains unclear, however, how both static and dynamic representations capture the main dimensions of vowel variation across a large number of dialects. This study examines the role of static, dynamic, and duration information for 5 vowels across 21 British and North American English dialects, and observes that vowels exhibit highly structured variation across dialects, with dialects displaying similar patterns within a given vowel, broadly corresponding to a spectrum between traditional 'monophthong' and 'diphthong' characterizations. These findings highlight the importance of dynamic and duration information in capturing how vowels can systematically vary across a large number of dialects, and provide the first large-scale description of formant dynamics across many dialects of a single language.

1 Introduction

Both the classification and measurement of vowels have long been central, intersecting, issues for phonetic research. Vowels are dynamic in production, yet language-specific vowel descriptions typically use broad categories referring to more or less general 'movement' of a vowel, such as distinguishing between *monophthongal* and *diphthongal* vowel realizations. At the same time, it is still unclear in what low-dimensional space vowels themselves vary: which acoustic properties best capture differences between vowels, and how securely categories like 'monophthong' and 'diphthong' can be established empirically within and across languages. Do these discrete categories reflect the ways in which vowels vary, or are vowel distinctions better characterized by a spectrum, reflecting various degrees of 'movement'? This study addresses this question by examining vowel variation *within* a language – across dialects – to consider how both *static* and *dynamic* properties of vowels capture dialectal variation across English.

Static measurements of formants, taken at a single time-point within the vowel, have long provided useful approximations for cues to vowel properties such as height and backness (e.g. Peterson and Barney, 1952; Hillenbrand and Gayvert, 1993), and have been central to previous descriptions of how vowels vary across dialects (e.g., Hagiwara, 1997; Clopper et al., 2005; Labov et al., 2006). Beyond single-point measurements of vowels, however, the importance of time-dependent *dynamic* information – such as spectral change and duration – has also been recognized since the earliest phonetic studies of vowel production and perception (e.g. Peterson and Barney, 1952; House, 1961; Gay, 1968).

Research on English has shown that this dynamic information may be utilized for better distinguishing vowels within a language (Harrington and Cassidy, 1994; Watson and Harrington, 1999; Williams and Escudero, 2014; Docherty et al., 2015), can reflect detailed dialectal and sociolinguistic meaning (Risdal and Kohn, 2014; Farrington et al., 2018; Williams et al., 2019), play a role in the development of dialect-specific vowel shifts (Evans, 1935; Labov, 1991; Clopper et al., 2005; Labov et al., 2006; Fox and Jacewicz, 2017), and constitute a robust source of variation across speakers (e.g. MacDougall, 2006; Morrison, 2009). Studies on single dialects have demonstrated that vowels vary in their average duration (House and Fairbanks, 1953; Peterson and Lehiste, 1960; Crystal and House, 1982), though our understanding of how vowel durations systematically vary across dialects is relatively limited (e.g., Bailey, 1968; Wetzell, 2000; Fridland et al., 2014; Tauberer and Evanini, 2009).

19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, pages 72 - 82
July 14, 2022 ©2022 Association for Computational Linguistics

Looking across *many* English dialects, however, it still remains unclear how best to characterize, on one hand, vowel variability across multiple acoustic dimensions (including how robustly monophthong/diphthong categories hold up across dialects), and on the other hand, the extent to which dynamic representations compare with static measures for characterizing differences between dialects on the basis of vowel realization. This study takes a computational and exploratory approach to addressing these issues, by considering the following research question: *to what extent do dynamic representations of vowels (formant trajectories, duration) capture additional information (over static F1/F2 position) in describing vowel variation across English dialects?* Concretely, answers to this question are addressed in two ways: 1. through an exploratory analysis of English vowel variability (Section 3.1), which enables inspection of the 'same' vowel across different dialects, including the evidence for monophthong/diphthong classification; 2. through a dialect classification experiment, where different combinations of formant position, trajectory shape, and duration are compared in their ability to correctly classify the dialect of a given vowel (Section 3.2). The exploratory analysis is motivated by the phonetic literature discussed above, which uses formant dynamics to characterize the vowel space of a given dialect, while the classification experiment is inspired by by the computational literature on dialect classification, where different kinds of acoustic information have been found to independently help differentiate dialects (e.g. Woehrling et al., 2009; Hanani et al., 2013; Chittaragi and Koolagudi, 2019).

The study takes a 'large-scale' approach, through the consistent extraction of the same measures for a large amount of data collected from speech corpora of 21 English dialects. Scaling up the analysis across multiple dialects is made possible by tools for automatic annotation (e.g. Schiel, 1999; Fromont and Hay, 2012; McAuliffe et al., 2017a), acoustic analysis (Rosenfelder et al., 2014; Mielke et al., 2019), and integrating information across idiosyncratic data formats (McAuliffe et al., 2017b, 2019). To our knowledge, this is the largest cross-dialect study to date of formant dynamics.

Vowels for the study were selected to provide a spectrum of qualities which are described in the English dialectological literature as ranging from largely monophthongal through to usually diphthongal, varying dialectally by the presence of a glide (Ladefoged and Maddieson, 1993), reflected in the degree of formant change over their timecourse. Specifically, the vowels were the following, as represented in terms of lexical sets (a characteristic word of a particular vowel) (Wells, 1982): FLEECE, FACE, PRICE, MOUTH, and CHOICE. FLEECE is expected to be monophthongal across dialects; MOUTH, PRICE, and CHOICE are expected to be diphthongs, which vary across dialects in both the degree of dynamic change and overall position (e.g. 'monophthongization' of PRICE in Southern US varieties, 'Canadian raising' of MOUTH in some Canadian/US varieties: Thomas, 2001; Labov et al., 2006; Boberg, 2010). FACE is expected to be intermediate between monophthongs and diphthongs, dependent on the specific dialect (e.g. Trudgill, 1999; Labov et al., 2006; Haddican et al., 2013).

2 Data

This study examines variation in stressed vowels from 21 British and North American English dialects, using corpus data collated as part of the SPeech Across Dialects of English (SPADE) project (Sonderegger et al., 2022, https://spade.glasgow.ac.uk/), including multi-dialect corpora from the United Kingdom (Coleman et al., 2012; Grabe, 2004; Anderson et al., 2007) and North America (Godfrey et al., 1992; Greenbaum and Nelson, 1996), as well as multiple individual English dialect corpora (Pitt et al., 2007; Dodsworth and Kohn, 2012; Stuart-Smith et al., 2017; Rosen and Skriver, 2015; Fabricius, 2000; Holmes-Elliott, 2015). Here, North American dialects refers to dialects in Canada and the United States as outlined in *The Atlas of North American English* (Labov et al., 2006). Due to the relative sparsity of Canadian data compared with United States and British dialects, Canadian dialects were distinguished along rural and urban dimensions instead of geographical location (Greenbaum and Nelson, 1996; Rosen and Skriver, 2015). Dialectal distinctions for British English used Trudgill's (1999) modern dialectal groupings, based on both phonological and lexical distinctions. Speakers for Scottish dialects were grouped based on information from *The Scottish National Dictionary* (Skretkowicz and Rennie, 2005).

Tokens with a duration shorter than 50 milliseconds were not extracted, in line with previous studies of vowel formants (Dodsworth, 2013; Frue-

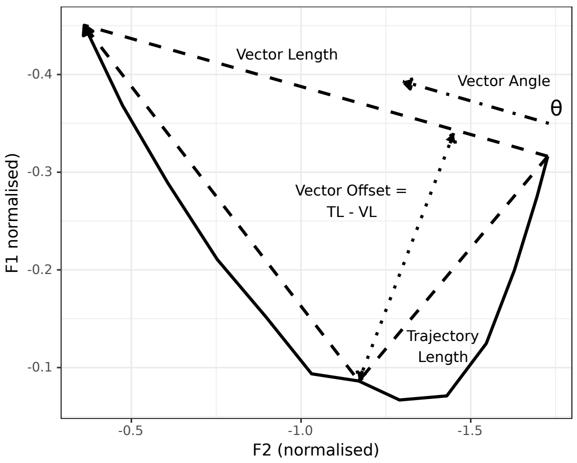

Figure 1: Schematic of all dynamic measures (dashed lines) used in the study mapped to a hypothetical CHOICE vowel trajectory (solid line).

hwald, 2013). Vowels with a duration longer than 500 milliseconds were also excluded. Formants were extracted in Hertz at 21 equally-spaced points, and were automatically measured with PolyglotDB (McAuliffe et al., 2017b) using the measurement scheme described in Mielke et al. (2019). The first and last 20% of the vowel was excluded to minimise the influence of surrounding segments (Fox and Jacewicz, 2009; Williams and Escudero, 2014; Williams et al., 2019). The remaining middle 60% of the vowel (13 points) was then z-score normalized against all vowels produced by the speaker ('Lobanov normalization', Lobanov, 1971).

In order to inspect spectral change across dialects more easily, and to allow comparison of our exploratory formant-based analyses with existing cross-dialect research (Section 3.1), we calculated a set of measures which are based on calculations of 'vowel section length' (VSL): the Euclidean distance between two formant points (n, m):

$$VSL_{n,m} = \sqrt{(F1_n - F1_m)^2 + (F2_n - F2_m)^2}$$

A measure of the overall spectral change (called 'Vector Length') is derived from calculating the VSL of the vowel onset and offset, whilst more complex representations of the trajectory can be derived from the summation of VSLs calculated from subsets of the points, such as onset to midpoint + midpoint to offset (Fox and Jacewicz, 2009). Figure 1 illustrates these measures on a hypothetical formant trajectory. A wide range of measures have been utilized within the vowel dynamic literature for capturing the dynamic properties of a formant trajectory, such as polynomial functions (MacDougall and Nolan, 2007; Van der Harst et al., 2014; Themistocleous, 2017), discrete cosine transforms (Watson and Harrington, 1999; Williams and Escudero, 2014), target-locus scaling (Broad and Clermont, 2017), and additive models (Kirkham et al., 2019; Renwick and Stanley, 2020) – the choice to use vector-based measurements of formant trajectories was motivated by their use in numerous studies of dialectal variation in English (Fox and Jacewicz, 2009; Cardoso, 2015; Farrington et al., 2018) and other languages (Mayr and Davies, 2011; Schoorman et al., 2015). Whilst these methods have not been explicitly compared, the decision to make use of the vector-based measurements in this study is based around the relative comparability with the previous cross-dialectal work using this measure, as well as its relative interpretability as a representation of spectral change. More information about the data, vowel formant extraction, and measurement calculation methods used in this study can be found in Tanner (2020). In total, 323,060 tokens (6259 types), corresponding to 1245 speakers from 21 dialects of North American and British English, were analyzed (Table 1).

3 Results

Figure 2 shows the vowel plot for each dialect included in the study, with arrows reflecting the vowel trajectories for each of the five vowels. Even from the empirical data, two findings are immediately clear: dialects are variable in their phonetic implementation of a given vowel, but there are also consistent patterns for the same vowel across dialects, including the anticipated monophthong-diphthong spectrum: from least movement for FLEECE to visible trajectories for CHOICE, PRICE, and MOUTH, with FACE showing dialect-specific variation consistent with monophthongal realization in Scottish dialects (Central Scotland, Edinburgh, Glasgow, N. Scotland & I) to diphthongs in other regions (East England, Midwest US). Again, the Scottish dialects show a distinct fronting pattern for MOUTH (shown as a reduction in normalized F2) compared with other dialects where MOUTH typically shows a backing pattern as it raises.

3.1 Exploratory analysis

To capture the formant position, the speaker-normalized F1 and F2 values were taken from the 20% and 80% points, corresponding to the vowel **Onset** and **Offset** respectively. Figure 3 (top) illus-

Continent	Dialect	Corpus	Speakers	Tokens
North America	Canada (rural)	Canadian-Prairies	44	20042
	Canada (rural)	ICE-Canada	8	2764
	Canada (urban)	Canadian-Prairies	67	38021
	Canada (urban)	ICE-Canada	8	877
	Midwest US	Buckeye	40	17669
	New England	Switchboard	18	2868
	North Midland US	Switchboard	44	7126
	Northern US	Switchboard	53	7494
	NYC	Switchboard	19	3183
	Raleigh US	Raleigh	100	64659
	South Midland US	Switchboard	106	20327
	Southern US	Switchboard	37	5595
	Western US	Switchboard	45	6376
United Kingdom	Central Scotland	SCOTS	23	5237
	East Central England	Audio BNC	30	3877
	East England	Audio BNC	100	13429
	East England	Hastings	49	25477
	East England	IViE	12	972
	East England	IViE	11	992
	East England	ModernRP	48	2811
	Edinburgh	SCOTS	18	2361
	Glasgow	SCOTS	26	4432
	Glasgow	SOTC	155	45487
	Lower North England	Audio BNC	41	5445
	Lower North England	IViE	11	891
	Lower North England	IViE	10	760
	North East England	Audio BNC	10	917
	North East England	IViE	12	1018
	Northern Scotland & Islands	SCOTS	31	3998
	South West England	Audio BNC	37	3458
	West Central England	Audio BNC	32	4497
Total	**21**	**11**	**1245**	**323060**

Table 1: Speaker and token count for each dialect used in this study, separated by the corpus from which the data was originally sourced.

trates the position of the onset and offset of each dialect, for each of the five vowels. This figure again captures overall consistency in the broad realization of a given vowel across dialects, but also the substantial differences between dialects in occupying the formant space for each vowel. The degree of this difference, however, varies by vowel. For example, dialects are somewhat diffused for CHOICE (outer left) FACE, (inner left), and PRICE (outer right), whilst maintaining some similarity in the difference between the onset and offset (reflected in the direction of the arrow) across dialects.

Three measures were calculated to capture properties of a vowel's formant trajectory independent of its position in formant space. The first, **Vector Length** (calculated from VSL, Equation 2), was calculated between the onset and offset value, reflecting the overall degree of linear spectral change over the vowel's timecourse. One measurement commonly used in studies of trajectory shape, trajectory length (Fox and Jacewicz, 2009; Mayr and Davies, 2011; Farrington et al., 2018) is calculated as the summation of two VSLs: one measuring the distance from the vowel onset to midpoint, and another measuring the distance from the midpoint to the vowel offset. As trajectory length is highly correlated with Vector Length ($r = 0.99$, $p < 0.001$ for this data), we derived our second measure, **Vec-**

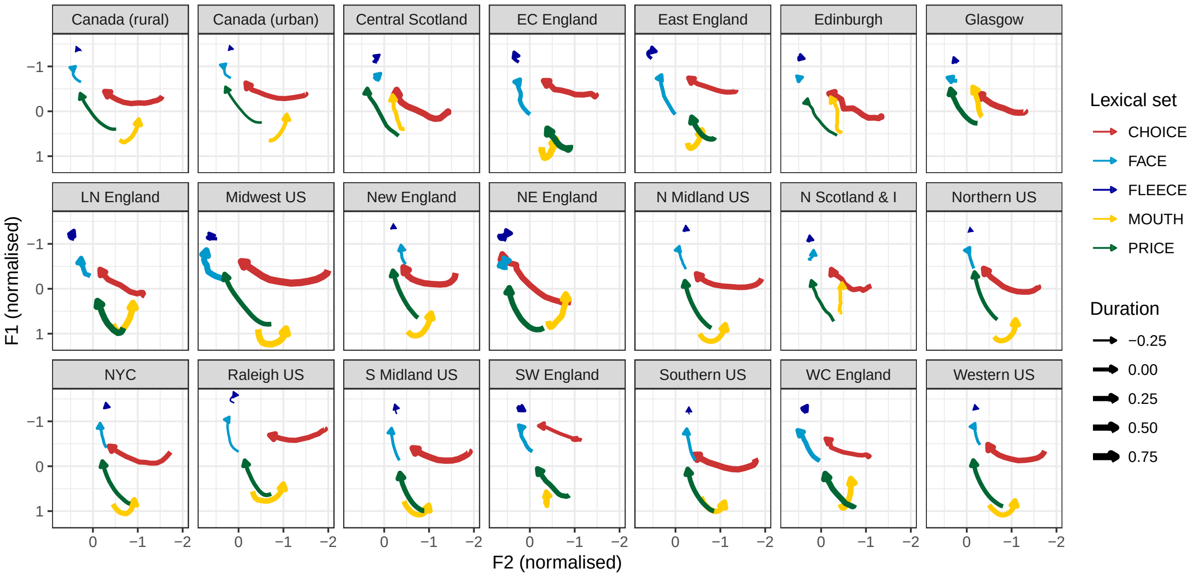

Figure 2: normalized by-dialect vowel trajectories for the central 60% of the five vowels analyzed, averaged over all tokens for that dialect. Duration corresponds to the within-speaker z-score normalization.

tor Offset, as trajectory length subtracted from Vector Length, reflecting the residual difference between the two measures. Finally, **Vector Angle**, the measure of a vowel's *direction* of change, was derived from the onset and offset position, on a $180/-180°$ scale (e.g., $\uparrow = 0$ °, $\leftarrow = 90$ °). Figure 3 (bottom) illustrates the dialectal variation in both Vector Length (a dialect's distance from the centre of the compass) and Vector Angle (the orientation around the compass). This figure demonstrates that, as with formant position (Fig. 3 top), the degree of dialectal variation for these dimensions differs between vowels, while showing some consistency within-vowel. FACE and PRICE show little dialectal variation in Vector Angle; instead, dialects differ in Vector Length. CHOICE and MOUTH show dialectal variation in both Vector Angle and Vector Length, within a clear range. (For example, CHOICE always points between $-90°$ and $0°$.) FLEECE shows very little overall spectral change, reflected in all dialects clustered around the centre of the compass.

Vowel **Duration** was calculated by z-score normalizing the vowel's force-aligned duration against all of the speaker's vowels (including vowels not analyzed in this study). As with previous measures, duration exhibits a wide range of variability across dialects, but this variability is somewhat structured within-vowel, roughly along the anticipated monophthong–diphthong axis: FLEECE shows the lowest average duration across dialects, with the least variability, followed by FACE (higher average, more variability), followed by PRICE/MOUTH/CHOICE.

Overall, the exploratory analysis shows that dialects tend to vary in how they produce the 'same' vowel, in fairly constrained ways, across both formant position and dynamics, consistent with the intuitive axis of degree of 'movement': FLEECE < FACE < PRICE, MOUTH, CHOICE in terms of how much dialectal variation there is in both spectral change and duration.

3.2 Dialect classification experiment

We now turn to quantitative characterization of the extent to which dynamics (trajectory shape, duration) provide additional information about dialectal variability on top of static measures (F1/F2 position). In this experiment, different combinations of measures are used to train a supervised learning model to predict the dialect label associated with data from a single vowel/speaker pair. Support vector machines (SVMs) were trained on each vowel using the `e1071` package (Meyer et al., 2019) in R (R Core Team, 2019). SVMs are a class of supervised learning model, which can be trained to assign ('classify') a label (such as dialect, e.g., Southern US, Glasgow) to a datapoint based on predictor values such as formant, trajectory, and duration measurements. The radial basis function kernel was used for SVMs in this study, which allows for fitting non-linear decision boundaries, since we do not a priori expect boundaries between dialects to

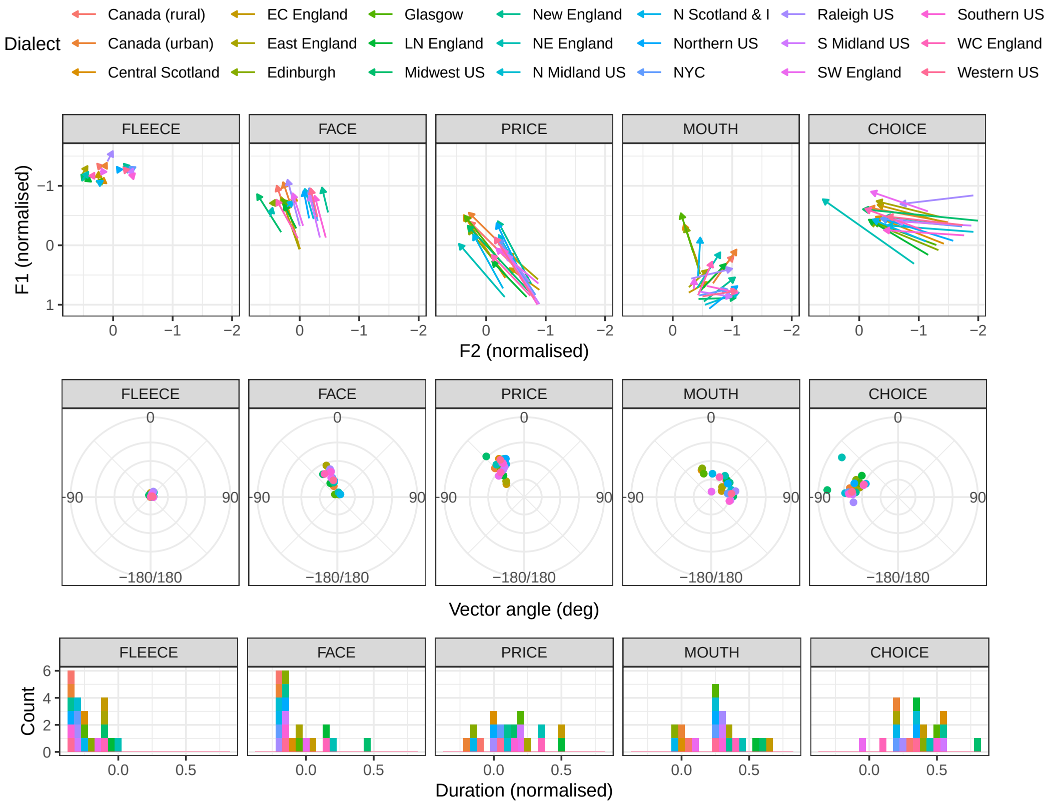

Figure 3: Top: Mean dialect F1 and F2 values for the 5 vowels (CHOICE, FACE, FLEECE, MOUTH, PRICE). One point per dialect. Onset value represented by the start point of the arrow; offset represented by position of arrowhead. Middle: Mean dialect values for Vector Angle (direction on compass) and Vector Length (distance from centre), for each of the five vowels in the study. Bottom: Mean z-normalized duration values per dialect.

be linear. We use a multiclass version of SVMs, to predict one of N-many possible dialect labels given prototypical formant position, trajectory shape, and duration values.

The data was prepared for SVM training by averaging formant, trajectory shape, and duration values for each speaker across each vowel, and separate SVMs were trained for each of the 5 vowels analyzed in this study. The choice to use one observation per speaker (compared to one value for each observation in the dataset) was motivated by the desire to abstract away from variability due to phonological environment, and instead achieve an 'average' value for a vowel for a speaker by averaging over all observations of that vowel by that speaker. To examine how different combinations of measures best contribute to accurately predicting the dialect, 7 SVMs were trained for each vowel on a different set of measurements (for a total of 35 SVMs):

1. Formant values (F1/F2 onset + offset)
2. Trajectory shape (Vec. Length, Offset, Angle)
3. Duration
4. Formants + duration
5. Trajectory + duration
6. Formants + trajectory
7. Formants + trajectory + duration

Each SVM was trained on a 80% subset of the data, and tuned to derive the best parameters (margin parameter C, kernel parameter γ) via 10-fold cross validation. A 'dummy classifier' model which returns the most common dialect label from the test set was also included as a baseline model. The performance on the 20% test set is evaluated using a metric that appropriately accounts for class imbalance. This measure, balanced accuracy, is the average of a model's sensitivity and specificity, and accounts for class imbalance by normalizing the true positive and negative rates by the relative number of samples (Kelleher et al., 2015).

(Note that balanced accuracy is 0.5 for the baseline.) Balanced accuracy was calculated using the `yardstick` package (Kuhn and Vaughan, 2020). To directly compare how different combinations of metrics aid in the classification of dialects, the differences in balanced accuracy for each vowel was calculated, and significance of the difference was evaluated through a one-sided permutation test, comparing the likelihood of whether the difference was greater than the average difference observed for 1000 permutations (Table 3), and were subject to within-vowel Benjamini-Hochberg False Discovery Rate (FDR) adjustment for multiple comparisons.

Table 2 shows the classification performance for each SVM, which demonstrates that using all SVMs improve over the naive baseline model (row 1), and the best-performing SVM includes dynamic information (trajectory or duration), for every vowel. Table 3 shows the performance differences between SVMs trained with different combinations of measurements, specifically comparing how dynamic measurements aid in distinguishing dialects relative to formant-only models (rows 1-2), as well as how utilizing *all* measurements compare with removing either trajectory measurements (row 3) or duration (row 4).

Comparing how dynamic (trajectory and duration) information provides additional resolution for distingiushing between dialects, the use of duration as a cue alongside formant information provides a large and significant increase in accuracy across all vowels (Table 3 row 1); alongside the observation that duration in isolation largely returns the lowest accuracy of all model sets (Table 2 row 4), this suggests English dialects do not sufficiently vary in duration for duration to uniquely distinguish dialects, but instead is a meaningful cue alongside a vowel's formant position. The additional effect of duration is mitigated when all measurements are included (Table 3 row 4), though including duration still results in significantly better classification accuracy for FLEECE, MOUTH, and PRICE. The additional role of trajectory information relative to formant position, in contrast, is much more variable across vowels (Table 3 row 2). Trajectory information plays the largest role for distinguishing MOUTH vowels across dialects, reflecting the fact that both Vector Length and Vector Angle vary substantially across dialects (Figure 3), with MOUTH in Scottish dialects fronting over its timecourse.

4 Discussion

This study has examined variability in English vowel realization across 21 dialects, to address the broad question of how to characterize variability in the 'same' vowel, across multiple acoustic dimensions, considering both static formant position and time-dependent dynamic information (trajectory shape, duration). What low-dimensional space does vowel variability lie in, does it line up with traditional notions of 'monophthong' vs. 'diphthong', and what role do static versus dynamic information play?

Our exploratory analysis (Section 3.1) found that while dialects vary in the static and dynamic realization of vowels, this cross-dialectal variation is clearly structured: the 'same' vowel patterns similarly with respect to dynamic realization, across dialects. As a first approximation, the patterns of dynamic variation within vowels seems to broadly correspond to the general monophthong/diphthong characterization, related to varying degrees of formant 'movement' during the vowel timecourse: FLEECE exhibits the least change, followed by FACE, with PRICE, MOUTH, CHOICE showing the most change; duration patterns similarly. Future work should incorporate more vowels into the analysis, to fully map out the structure of variability within and between dialects, and assess its possible sources.

The dialect classification experiment (Section 3.2) showed that whilst both formant position and trajectory shape can separately inform the prediction of a given dialect, accuracy is improved with both types of measures are used together. While previous work has shown that trajectory information is informative *within* a given dialect, these results demonstrate that characterizations of the formant trajectory also provide additional resolution as to the ways vowels can systematically differ across individual dialects. This study utilized one particular representation of trajectory shape: Vector Length/Offset/Angle. Testing other representations of trajectory shape, such as DCTs (Watson and Harrington, 1999; Williams and Escudero, 2014; Williams et al., 2019), would be a useful avenue for future research, especially if these improve on dialect classification accuracy, which is fairly low when using Vector Length/Offset/Angle.

Our understanding of cross-dialectal variation in vowel duration has been largely limited to studies of North American dialects, especially in the

Measures	**FLEECE**	**FACE**	**PRICE**	**MOUTH**	**CHOICE**
Baseline (most common dialect label)	50	50	50	50	50
Formants (F1, F2 onset + offset)	58	61.3	62.2	61.4	56.7
Trajectory (Vector Length, Offset, Angle)	54.5	62.1	56	63.6	56
Duration	55	52.9	57.4	52.7	51.9
{Formants, duration}	62.5	**65.3**	66.2	66.4	**60.3**
{Trajectory, duration}	56.7	65.1	60.6	65.4	55.9
{Formants, trajectory}	60.8	62.7	65	67.4	57.6
{Formants, trajectory, duration}	**63.4**	**64.2**	**69.4**	**70**	**59.2**

Table 2: Balanced accuracy (%) for each SVM, trained with different configurations of formant position, trajectory shape, and duration measures.

	FLEECE		**FACE**		**PRICE**		**MOUTH**		**CHOICE**	
Comparisons	**ΔBa.**	p	**ΔBa.**	p	**ΔBa.**	p	**ΔBa.**	p	**ΔBa.**	p
{F, D} vs F	4.5	**0.004**	4	**0.004**	4	**0.006**	5	**0**	3.6	**0.046**
{F, T} vs F	2.8	**0.034**	1.4	0.122	2.8	**0.018**	6	**0**	0.9	0.231
{F, T, D} vs {F, D}	0.9	0.148	−1.1	0.39	3.2	**0.009**	3.6	**0.008**	−1.1	0.364
{F, T, D} vs {F, T}	2.6	**0.034**	1.5	0.122	4.4	**0.002**	2.6	**0.021**	1.6	0.181

Table 3: Differences in balanced accuracy (ΔBa., %) between different combinations of measurements, with within-vowel FDR-adjusted p-values calculated using a one-sided permutation test with 1000 permutations (bold indicates $p < 0.05$). *F* = Formants, *T* = Trajectory, *D* = Duration.

US South (e.g. Jacewicz et al., 2007; Tauberer and Evanini, 2009; Fridland et al., 2014), leaving open the question of how duration varies across English dialects more generally. Results of the dialect classification experiment suggest that duration does contribute unique information over formant position and trajectory shape, but it is the least informative feature. However, this study only included vowels which are 'tense' in most dialects, which tend to be longer (than 'lax' vowels). Future work incorporating more vowels into the analysis would allow for better assessment of the role of duration, and would provide additional information about about dialectal differences in duration across English vowels in general.

To our knowledge, this is the largest study to date of formant dynamics (in terms of number of dialects, and tokens), for any language. Analyzing data at this scale was made possible due to access to a large number of corpora and tools for automated acoustic measurement. Previous large cross-dialectal analyses (e.g. Wells, 1982; Thomas, 2001; Labov et al., 2006) were multi-year enterprises requiring substantial time and labor-intensive manual annotation, making only simple characterizations of vowel dynamics (e.g. onset + offset) possible. Access to force-aligned speech corpora and the automatic measurement of formants allows the analysis to be 'scaled-up' easily relative to many other dialectal studies of vowel quality, but also requires recognition of a number of limitations for studies of this kind. Whilst this method has been shown to generate accurate formant values and procedures are taken to avoid tracking 'false formants' (Mielke et al., 2019), it is simply not possible with data at this scale to be manually validated. Similarly forced aligned segments have a minimum time duration (often 30ms) and a minimum time resolution (often 10ms), particularly for vowels which may have undergone substantial reduction. We attempted to account for this by applying lower and upper-limits for vowel durations to be included in the study; it remains possible that biases or inaccuracies in vowel duration exist within the dataset.

Acknowledgements

We acknowledge the crucial contribution of The SPADE Consortium, which comprises the Data Guardians who generously shared their datasets which are reported in this paper, and/or whose datasets were used in the development of the Integrated Speech Corpus Analysis tool. No SPADE research would have been possible without The SPADE Consortium (https://spade.glasgow.ac.uk/the-spade-consortium/). Additional thanks to Jeff

Mielke, Rachel Macdonald, Vanna Willerton, and SPADE research assistants. This research was supported by SPeech Across Dialects of English (SPADE): Large-scale digital analysis of a spoken language across space and time (2017-2020), ESRC Grant ES/R003963/1, NSERC/CRSNG Grant RGPDD 501771-16, SSHRC/CRSH Grant 869-2016-0006, NSF Grant SMA-1730479 (Digging into Data/Trans-Atlantic Platform), and SSHRC #435-2017-0925.

References

Jean Anderson, Dave Beavan, and Christian Kay. 2007. The Scottish corpus of texts and speech. In J. C. Beal, K. P. Corrigan, and H. L. Moisl, editors, *Creating and Digitizing Language Corpora*, pages 17–34. Palgrave, New York.

Charles-James Bailey. 1968. Segmental length in Southern States English: an instrumental phonetic representation of a standard dialect in South Carolina. In *PEGS Paper No. 20*. Center for Applied Linguistics, Washington DC.

Charles Boberg. 2010. *The English Language in Canada: Status, History and Comparative Analysis*. Cambridge University Press, Cambridge.

David J. Broad and Frantz Clermont. 2017. Target-locus scaling for modeling formant transitions in vowel + consonant + vowel utterances. *Journal of the Acoustical Society of America*, 141:EL192–EL198.

Amanda Beth Cardoso. 2015. *Dialectology, phonology, diachrony: Liverpool English realisations of price and mouth*. Ph.D. thesis, University of Edinburgh.

Nagaratna B Chittaragi and Shashidhar G Koolagudi. 2019. Acoustic-phonetic feature based kannada dialect identification from vowel sounds. *International Journal of Speech Technology*, 22(4):1099–1113.

Cynthia G. Clopper, David B. Pisoni, and Kenneth de Jong. 2005. Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, 118:1661–1676.

John Coleman, Ladan Baghai-Ravary, John Pybus, and Sergio Grau. 2012. Audio BNC: the audio edition of the Spoken British National Corpus. Technical report, Oxford. Http://www.phon.ox.ac.uk/AudioBNC.

Thomas H. Crystal and Arthur S. House. 1982. Segmental durations in connected speech signals: preliminary results. *Journal of the Acoustical Society of America*, 72:705–716.

Gerry Docherty, Simon Gonzalez, and Nathaniel Mitchell. 2015. Static vs dynamic perspectives on the realization of vowel nucleii in West Australian English. In *Proceedings of the 18th International Congress of Phonetic Sciences*.

Robin Dodsworth. 2013. Retreat from the Southern Vowel Shift in Raleigh, NC: social factors. *University of Pennsylvania Working Papers in Linguistics*, 19:31–40.

Robin Dodsworth and Mary Kohn. 2012. Urban rejection of the vernacular: The SVS undone. *Language Variation and Change*, 24:221–245.

Medford Evans. 1935. Southern 'long i'. *American Speech*, 10:188–190.

A. H. Fabricius. 2000. *T-glottalling between stigma and prestige: a sociolinguistic study of Modern RP*. Ph.D. thesis, Copenhagen Business School, Copenhagen, Denmark.

Charlie Farrington, Tyler Kendall, and Valerie Fridland. 2018. Vowel dynamics in the southern vowel shift. *American Speech*, 93:186–222.

Robert Allen Fox and Ewa Jacewicz. 2009. Cross-dialectal variation in formant dynamics of American English. *Journal of the Acoustical Society of America*, 126:2603–2618.

Robert Allen Fox and Ewa Jacewicz. 2017. Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English. *Journal of the Acoustical Society of America*, 142:444–459.

Valerie Fridland, Tyler Kendall, and Charlie Farrington. 2014. Durational and spectral differences in American English vowels: dialect variation within and across groups. *Journal of the Acoustical Society of America*, 136:341–349.

R. Fromont and J. Hay. 2012. LaBB-CAT: an annotation store. In *Australasian Language Technology Workshop 2012*, volume 113, pages 113–117.

Josef Fruehwald. 2013. *The Phonological Influence on Phonetic Change*. Ph.D. thesis, University of Pennsylvania.

T. Gay. 1968. Effects of speaking rate on dipthong formant movements. *Journal of the Acoustical Society of America*, 44:1570–1573.

John J. Godfrey, Edward C. Holliman, and Jane McDaniel. 1992. SWITCHBOARD: telephone speech corpus for research and development. In *Proceedings of the 1992 IEEE international conference on Acoustics, speech and signal processing - Volume 1*, pages 517–520.

E. Grabe. 2004. Intonational variation in English. In P. Gilles and J. Peters, editors, *Regional Variation in Intonation*, pages 9–31. Niemeyer, Tubingen.

S. Greenbaum and G. Nelson. 1996. The International Corpus of English (ICE project). *World Englishes*, 15:3–15.

Bill Haddican, Paul Foulkes, Vincent Hughes, and Hazel Richards. 2013. Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change*, 25:371–403.

Robert Hagiwara. 1997. Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*, 102.

Abualsoud Hanani, Martin J Russell, and Michael J Carey. 2013. Human and computer recognition of regional accents and ethnic groups from british english speech. *Computer Speech & Language*, 27(1):59–74.

Jonathon Harrington and Stephan Cassidy. 1994. Dynamic and target theories of vowel classification: evidence from monophthongs and diphthongs in Australian English. *Language and Speech*, 37:357–373.

James Hillenbrand and R. T. Gayvert. 1993. Vowel classification based on fundamental frequency and formant frequencies. *Journal of Speech, Language, and Hearing Research*, 36:694–700.

Sophie Holmes-Elliott. 2015. *London calling: assessing the spread of metropolitan features in the southeast.* Ph.D. thesis, University of Glasgow.

A. S. House and G. Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America*, 25:105–113.

Arthur S. House. 1961. On vowel duration in English. *Journal of the Acoustical Society of America*, 33:1174–1178.

Ewa Jacewicz, Robert Allen Fox, and J. Salmons. 2007. Vowel duration in three American English dialects. *American Speech*, 82:367–385.

John D. Kelleher, Brian Mac Namee, and Aoife D'Arcy. 2015. *Fundamental of Machine Learning for Predictive Data Analytics*. MIT Press, Cambridge MA.

Sam Kirkham, Claire Nance, Bethany Littlewood, Kate Lightfoot, and Eve Groake. 2019. Dialect variation in formant dynamics: the acoustics of lateral and vowel sequences in Manchester and Liverpool English. *Journal of the Acoustical Society of America*, 145:784–794.

Max Kuhn and Davis Vaughan. 2020. *yardstick: Tidy Characterizations of Model Performance*. R package version 0.0.6.

William Labov. 1991. Three dialects of English. In Penelope Eckert, editor, *New ways of analyzing variation in English*, pages 1–45. Academic, New York.

William Labov, Sharon Ash, and Charles Boberg. 2006. *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Mouton de Gruyter, Berlin.

Peter Ladefoged and Ian Maddieson. 1993. *The Sounds of the World's Languages*. Wiley, Oxford.

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

Kirsty MacDougall. 2006. Dynamic features of speech and the characterization of speakers: Toward a new approach using formant frequencies. *The International Journal of Speech, Language and the Law*, 13:89–126.

Kirsty MacDougall and Francis Nolan. 2007. Discrimination of speakers using the formant dynamics of /u:/ in British English. In *Proceedings of the 16th International Congress of Phonetic Sciences*, pages 1825–1828. Saarbrucken.

Robert Mayr and Hannah Davies. 2011. A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh. *Journal of the International Phonetic Association*, 41:1–25.

Michael McAuliffe, Arlie Coles, Michael Goodale, Sarah Mihuc, Michael Wagner, Jane Stuart-Smith, and Morgan Sonderegger. 2019. ISCAN: A system for integrated phonetic analyses across speech corpora. In *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne.

Michael McAuliffe, Michaela Scolof, S. Mihuc, Michael Wagner, and Morgan Sonderegger. 2017a. Montreal forced aligner [computer program]. Https://montrealcorpustools.github.io/Montreal-Forced-Aligner/.

Michael McAuliffe, Elias Stengel-Eskin, Michaela Socolof, and Morgan Sonderegger. 2017b. Polyglot and Speech Corpus Tools: a system for representing, integrating, and querying speech corpora. In *Proceedings of Interspeech 2017*.

David Meyer, Evgenia Dimitriadou, Kurt Hornik, Andreas Weingessel, and Friedrich Leisch. 2019. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*. R package version 1.7-3.

Jeff Mielke, Erik R. Thomas, Josef Fruehwald, Michael McAuliffe, Morgan Sonderegger, Jane Stuart-Smith, and Robin Dodsworth. 2019. Age vectors vs. axes of intraspeaker variation in vowel formants measured automatically from several English speech corpora. In *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia.

Geoffrey Stewart Morrison. 2009. Likelihood-ratio forensic voice comparison using parametric representations of the formant trajectories of diphthongs. *Journal of the Acoustical Society of America*, 125:2387–2397.

G. E. Peterson and H. L. Barney. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 24:175–184.

Gordon E. Peterson and Ilse Lehiste. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, 32:693–703.

M. A. Pitt, L. Dilley, K. Johnson, S. Kiesling, W. Raymond, E. Hume, and E. Fosler-Lussier. 2007. *Buckeye Corpus of Spontaneous Speech*, 2 edition. Ohio State University, Columbus.

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Margret E. L. Renwick and Joseph A. Stanley. 2020. Modeling dynamic trajectories of front vowels in the American South. *Journal of the Acoustical Society of America*, 147:579–595.

Megan L. Risdal and Mary E. Kohn. 2014. Ethnolectal and generational differences in vowel trajectories: Evidence from African American English and the Southern vowel system. In *Selected papers from NWAV 42*, pages 139–148, Philadelphia. University of Pennsylvania.

Nicole Rosen and Crystal Skriver. 2015. Vowel patterning of Mormons in Southern Alberta, Canada. *Language & Communication*, 42:104–115.

Ingrid Rosenfelder, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, and Jiahong Yuan. 2014. FAVE (Forced Alignment and Vowel Extraction) program suite v1.2.2 10.5281/zenodo.22281.

F. Schiel. 1999. Automatic Phonetic Transcription of Non-Prompted Speech. In *Proc. of the ICPhS*, pages 607–610, San Francisco.

Heike Schoorman, Wilbert Heeringa, and J *o*rg Peters. 2015. Regional variation of Saterland Frisian vowels. In *Proceedings of the 18th International Congress of Phonetic Sciences*. University of Glasgow.

Victor Skretkowicz and Susan Rennie. 2005. *Scottish National Dictionary*. Dictionaries of the Scottish Language.

Morgan Sonderegger, Jane Stuart-Smith, Michael McAuliffe, Rachel Macdonald, and Tyler Kendall. 2022. Managing data for integrated speech corpus analysis in SPeech Across Dialects of English (SPADE). In *Open Handbook of Linguistic Data Management*, pages 195–207. MIT Press, Cambridge.

Jane Stuart-Smith, B. Jose, Tamara Rathcke, Rachel MacDonald, and E. Lawson. 2017. Changing sounds in a changing city: An acoustic phonetic investigation of real-time change over a century of Glaswegian. In C. Montgomery and E. Moore, editors, *Language and a Sense of Place: Studies in Language and Region*, pages 38–65. Cambridge University Press, Cambridge.

James Tanner. 2020. *Structured phonetic variation across dialects and speakers of English and Japanese*. Ph.D. thesis, McGill University.

Joshua Tauberer and Keelan Evanini. 2009. Intrinsic vowel duration and the post-vocalic voicing effect: some evidence from dialects of North American English. In *Proceedings of Interspeech*.

Charalambos Themistocleous. 2017. Dialect classification using vowel acoustic parameters. *Speech Communication*, 92:13–22.

Erik R. Thomas. 2001. *An acoustic analysis of vowel variation in New World English*. American Dialect Society.

Peter Trudgill. 1999. *The Dialects of England*. Blackwell, Oxford.

Sander Van der Harst, Hans Van de Velde, and Roeland Van Hout. 2014. Variation in standard dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin. *Language Variation and Change*, 26(2):247–272.

Catherine I. Watson and Jonathon Harrington. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America*, 106:458–468.

John C. Wells. 1982. *Accents of English*. Cambridge University Press, New York.

Brett Wetzell. 2000. Rhythm, dialects, and the Southern Drawl. Master's thesis, North Carolina State University.

Daniel Williams, Jaydene Elvin, Paola Escudero, and Adamanitos Gafos. 2019. Multidimensional variation in English diphthongs. In *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia.

Daniel Williams and Paola Escudero. 2014. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *Journal of the Acoustical Society of America*, 136:2751–2761.

Cécile Woehrling, Philippe Boula de Mareüil, and Martine Adda-Decker. 2009. Linguistically-motivated automatic classification of regional french varieties. In *Tenth Annual Conference of the International Speech Communication Association*.