The search for linguistically coherent accents

Unsupervised clustering of diphthong variation in Southeast England

Amanda Cole and Patrycja Strycharczuk
University of Essex | University of Manchester

Linguistic research refers to many related accents in Southeast England: Standard Southern British English (SSBE), Received Pronunciation (RP), Estuary English (EE), Cockney and Multicultural London English (MLE). However, there is inconsistency and imprecision in the demarcation of these accents based on linguistic and social factors. This paper delineates accents in Southeast England based on patterns of linguistic co-variation which we then relate to social predictors. We applied functional Principal Component Analysis to F1 and F2 measurements for diphthongs extracted from wordlist and passage productions for 193 young, south-eastern speakers. Principal Components were entered into a clustering analysis that identified patterns of linguistic co-occurrence. Three clusters emerge, broadly aligning with SSBE, MLE and EE for both linguistic and social factors. We illustrate the linguistic centre of gravity of the three diphthong systems for use as reference points in future research, and we discuss the need to make explicit how accents are defined.

1. Introduction

A central tenet of what constitutes an accent is co-variation between multiple linguistic features. Guy (2013: 64) considers accents, or more broadly lects, as characterised by a "cluster of variables" which are coherent within a single accent. The question of how coherent these features must be and whether a lect is a

coherent object at all is referred to by Beaman and Guy (2022) as the "unity dilemma". Their work builds on Guy and Hinskens's (2016) double issue of *Lingua* addressing linguistic coherence, and specifically Weinreich, Labov and Herzog's (1968:100) notion of "orderly hetereogeneity" which presumes coherence and co-variation among sociolinguistic variables. The central premise of work on linguistic coherence is that linguistic features should act in unison within speakers of a single accent: variants (or rates of use of variants) within a single lect should be correlated (Guy and Hinskens 2016:2). Guy (2013:63) writes that if there is not linguistic coherence "the cognitive and social reality of the 'sociolect' is problematic". Though the concept of linguistic coherence intuitively feels like a foundation of a single lect, several studies have found relatively limited linguistic coherence in various linguistic varieties (Brazilian Portuguese: Oushiro and Guy 2015; New York City English: Becker 2016; Copenhagen Danish: Gregersen and Pharao 2016).

When limited co-variation is found between linguistic features in a single lect, it is possible that other factors mediate rates of co-variation such as salience (Erker 2022) or social meaning (Cole 2020). Alternatively, it could also be the case that the speaker groups which we presume to constitute a single linguistic variety based on social and/or regional factors include subpopulations with important linguistic differences. In these studies, linguistic coherence was tested within predetermined groups of speakers who were presumed to form a unified and identifiable speech community despite sometimes being selected from large populations such as whole cities or regions. It seems extreme to redefine our concept of a lect such that linguistic coherence is not necessarily a requirement (as do Gregersen and Pharao 2016). Instead, it is worth considering that the way we choose speech communities from a top-down approach may not in fact represent the most logical and verifiable split between speaker groups in terms of linguistic content.

1.1   Demarcating and defining accents

All accents occur on a continuum and include internal variation. Linguists are tasked with determining where one continuum ends and another begins, constituting distinct accents, and how this interacts with the social and regional makeup of speakers. Altendorf (2003:6) suggests that the defining element of a single variety is a "centre of gravity which is shared by all speakers of the same variety". However, what remains unclear is where the centre of gravity should fall, how much internal variation from this centre of gravity is permitted or expected within a single accent, and the degree of difference required from distinct accents. Central to this task is analysing and categorising patterns of co-variation between multiple linguistic features. In one such approach, Trudgill (1990:32) draws isoglosses on a

single composite map of England for eight linguistic features. He then splits England into thirteen dialect areas based on where the most "abrupt transitions" or overlapping isoglosses occur (Trudgill 1990: 32). He suggests that, though dialects occur on a continuum, we can determine dialect areas by dividing this continuum into "areas at points where it is least continuum-like" (Trudgill 1990: 6).

However, linguists typically do not use just linguistic data to delineate and demarcate separate accents. Linguists often also use information about speakers such as demographic information (e.g. age, gender), where they are from/live, as well as their language attitudes and evaluative norms. Correspondingly, both linguistic and social components constitute an accent. We adopt the working definition of an accent as a set of linguistic features shared by a particular demographic group. A classic example is Labov's (1966) work in which he selected speakers from New York City and explored patterns of variation in their speech. The concept of a speech community is very similar to the concept of an accent: a group of speakers whose linguistic production can be defined as a single, coherent system. Labov found that, firstly, speakers in NYC had linguistic uniformity which he suggested is a prerequisite of a speech community. Nonetheless, the speakers clearly did not all speak identically, but instead, structured variation existed which mapped onto social factors. Secondly, the speakers also shared evaluative norms, elucidated through style shifting patterns and language attitudes. Labov (1966) believed that the combination of these two factors confirmed that the speakers did indeed stem from a single speech community (see Patrick 2002 for further discussion).

Nonetheless, this approach is not without problems. Firstly, both structured linguistic variation and shared evaluative norms are not binary measures that either exist or do not. Instead, they both exist on a continuum, and there is no objective threshold for when a group of speakers can be considered a speech community. Secondly, several speakers were excluded from Labov's (1966) analysis based on social eligibility criteria such as non-native speakers of English. Eligibility criteria in linguistic studies pre-determine who is considered an authentic speaker. Structured, orderly and predictable patterns of variation in a community is an important finding, but it may be that some variation has already been filtered out with the preselection of speakers. In addition, Labov analysed patterns of linguistic production for African American speakers in NYC separately to other New Yorkers in his sample because he suggested that they participate in a different phonology. This approach is open to criticisms of circularity: linguistic features were the target of investigation but also used in the selection of speakers. Preselecting supposedly "authentic" speakers based on social and linguistic criteria leads to potential variability between linguists in the perceived linguistic centre of gravity of an accent. The same issue surfaces in the context of research on

Standard Southern British English (SSBE), one of the focal varieties in the current study, as discussed below.

1.2   Who speaks SSBE?

SSBE (sometimes referred to instead as Standard Southern British: SSB) is the term often used to refer to broad patterns of speech spoken in southern England. However, it is an accent which has been defined imprecisely and inconsistently in linguistic work. There are no clear or empirically determined boundaries between SSBE and other accents such as Received Pronunciation (RP), Estuary English (EE), Cockney or even Multicultural London English (MLE) in terms of linguistic content or group membership. Many linguists draw similarities between RP and SSBE but stop short of entirely equating them. For instance, Harrington, Kleber and Reubold (2008: 2825) refer to SSBE as "a variety spoken by the majority of RP speakers". Others suggest that SSBE is simply a new term which has replaced RP. This line of thought suggests that the term SSBE is used due to RP having acquired a "rather dated – even negative – flavour in contemporary British English" (Hughes, Trudgill, and Watt 2012: 3) and an ideological link with "the past and the upper class" (Lindsey 2019: 4). If we subscribe to this stance, the ways RP and SSBE speakers are identified should be identical or at least strongly comparable, with both terms describing the speech of the highest social classes. However, in reality, in linguistic research the terms are often used to refer to overlapping but not identical groups of speakers.

The defining features of RP are, firstly, that it is spoken by the highest social classes and, secondly, that it contains "no regional features whatsoever" (Trudgill 2001: 172). In addition, RP is associated with the British private, and particularly public (elite private) boarding school system (Agha 2003). Nonetheless, though being from the higher social classes is a prerequisite for an RP speaker, private schooling is not an eligibility requirement in all studies on RP. Badia Barrera (2015: 86) recruits speakers of RP from both private and comprehensive schools but indicates that the latter group live in a "prosperous rural area in the South of England". Much like for RP, many linguists include notions related to class in their eligibility criteria for determining an SSBE speaker. Similarly to Bada Barrera (2015: 86), Alderton (2022: 291) recruits speakers from both private and comprehensive schools but believes they speak SSBE because they live in the southern county of Hampshire where "the proportion of people in professional and managerial occupations in the county is higher than the national average", regardless of the speakers' own background. Holmes-Elliott and Levon (2017: 1049) also refer to class when identifying speakers of "upper middle-class Standard South-

ern British accent (not unlike Received Pronunciation)" on the scripted reality TV show *Made in Chelsea*, which is set in a very affluent district of West London.

Nonetheless, there are many other studies in which social class or proxies for social class are not included in the eligibility criteria for SSBE speakers. In a large number of linguistic studies, being from southern England is the only criterion that determines an SSBE speaker. However, even this criterion is not applied uniformly. While some studies require SSBE speakers to be from any area of southern England, others stipulate that they be specifically from Southeast England or even from the Home Counties (e.g. Williams and Escudero 2014). Chládková et al. (2017: 381) identify SSBE speakers as those who have spent "most of their lives in a geographical area where SSBE is spoken". The speakers are split into two age groups and are recruited from geographically disparate places in the south of England: Dorset, London, East Sussex and Hertfordshire. As no older speakers were recruited from Dorset, and no younger speakers were recruited from East Sussex or Hertfordshire, their apparent-time analysis implicitly assumes that SSBE is a single, homogenous accent that does not include regional variation, including between the southwest and the southeast of England. This is in spite of the division between the west and the east of southern England being one of the major English dialect isoglosses identified by Trudgill (1990). Perhaps this is why the speakers were also only selected if "their accent was judged as SSBE by the experimenters" (Chládková et al. 2017: 381).

Other studies have not required SSBE speakers to even be from southern England, but again, this is predicated on them speaking in a certain way. *The Dynamic Variability in Speech Corpus* (DyViS) is described as including "100 homogeneous speakers" (Nolan et al. 2009: 31), and it explicitly invokes the notion that SSBE represents "a single English-speaking speech community" (Hudson et al. 2007: 1809). Speakers recorded in this corpus were male, aged 18–25 and were mostly students at the University of Cambridge. They were not required to be exclusively from southern England but "there was no question that they were native speakers of SSBE" (Nolan et al. 2009: 37). Speakers' eligibility as SSBE speakers was vetted based on their linguistic productions. Potentially eligible speakers were asked to ring a dedicated phone number and leave a message so their accents could be covertly assessed. Speakers were considered eligible if they did not produce rhoticity, ʟ-vocalisation, ᴛʜ-fronting, ʏᴏᴅ-dropping, ɢ-dropping, the same vowel in ꜱᴛᴀʀᴛ as ᴛʀᴀᴘ, but could produce occasional ᴛ-glottalling despite this also being a non-standard feature. However, the final verdict about speakers' eligibility was made by an "SSBE native speaker" (Nolan et al. 2009: 39).

Similarly, Deterding (1997), a study often used as a baseline for synthesising SSBE vowel productions (e.g. Escudero and Chládková 2010), did not require SSBE speakers to be from southern England or, at least, this criterion is not

mentioned. Deterding selected speakers from professional occupations such as 1980s BBC newsreaders. This criterion alone could have permitted speakers with accents that are notably distinct to SSBE such as, for instance, 1980s BBC broadcaster Terry Wogan who speaks with an Irish accent. Though not made explicit, linguistic productions did seemingly form part of Deterding's speaker selection processes. He writes: "All the speakers have what might be termed a Standard Southern British accent (similar to RP) though there is inevitably a little variation between them" (Deterding 1997: 48). We are not told about the linguistic centre of gravity, or the degree or nature of linguistic variation permitted.

It is perhaps untenable to recruit a group of speakers of a single, coherent accent based only on their background characteristics without linguistic productions feeding into decisions even in a covert, unacknowledged or supplementary fashion. As we have seen, the way linguistic productions feed into decision-making about who is an eligible SSBE speaker is often based on intuitions and is not explicitly described. In this way, linguists are, much like non-linguists, subject to processes of enregisterment (see Agha 2003), in which ideological links are made between an accent label/nominalised entity, a set of social/regional expectations, and a set of linguistic features. Accents are arbitrarily and not objectively demarcated entities, and linguists hold intuitions and expectations about the linguistic and social makeup of specific accents. We may recruit speakers of SSBE, as well as other accents, based, in part, on intuitions about who is an authentic speaker. As a result, the way SSBE speakers are identified is often not reproducible and studies may not be replicable due to potentially differing intuitions among linguists as to the linguistic make-up of SSBE. It is perhaps not surprising that the linguistic material contributing to decision making around speaker eligibility is often not made explicit in much research – as we will see, there is currently no clear consensus on which linguistic features constitute SSBE.

1.3   Features of SSBE and related varieties

To date, Lindsey (2019) has provided the most complete description of SSBE linguistic features. Though many studies provide detail on SSBE speakers' social and regional background with no or little detail on their linguistic productions, Lindsey takes the opposite approach, presenting a detailed picture of the linguistic features, but very little explanation of who the sample speakers are or how the amalgamated linguistic forms were obtained. He describes SSBE as having changed over decades from RP. Lindsey's version of SSBE diverts from the IPA symbols used to describe RP vowels which were first proposed by Alfred C. Gimson (1962). He describes the diphthongs in "modern SSB" as having tenser endpoints than RP's [ɪ] or [ʊ]: FLEECE [ɪj] GOOSE [ʉw] FACE [ɛj] GOAT [əw] PRICE

[ɑj] ᴍᴏᴜᴛʜ [aw] ᴄʜᴏɪᴄᴇ [oj]. Lindsey's SSBE vowel productions have become more similar to the shifted vowel system observed in Cockney in which /ʊ/- and /ɪ/-diphthongs are rotated clockwise and anti-clockwise respectively (see Cole and Strycharczuk 2022).

Cockney is an accent associated with the white working class in East London (Fox 2015:1). Cockney is an endpoint on a south-eastern continuum of accents in opposition to RP. Cockney has traditionally been considered the most vernacular accent in both London and the southeast of England (Wells 1982:302). The accent in Essex, the county that borders East London, has been most heavily influenced by Cockney (see Cole and Evans 2020) but all of Southeast England has felt the perennial influence of Cockney for at least five hundred years (Wells 1997:47). Lindsey's description of SSBE reflects this. However, Lindsey's account of SSBE's linguistic make up is in opposition to other studies (e.g. Nolan et al. 2009; Arvaniti and Atkins 2016). The discrepancies in the linguistic material permitted in SSBE speakers is due, in part, to patterns of language change which have rendered SSBE a moving target. Fabricius (2002) and Trudgill (2001) account for ongoing language change in RP/SSBE by using caveated terms such as "modern RP" which they distinguish from the idealised set of linguistic productions associated with RP. However, ongoing language change has already surpassed the linguistic benchmarks they outline as "modern RP" at the start of the 20th century. The pan regional standard accent in southern England – whether it is termed RP, modern RP, SSBE or SSB – is in a constant process of change, particularly as a result of the enduring diffusion of Cockney features, which has complicated and diversified definitions.

Definitions of SSBE are particularly complicated because the accent falls at an imprecisely defined midpoint on a continuum between Cockney and the idealised set of linguistic values for RP (such as those first proposed by Alfred C. Gimson (1962) or described as "construct-RP" by Fabricius (2002)). Cockney and RP are accents that are emblematic of the lowest and highest classes respectively. As a result, the linguistic continuum between these two accents parallels the class continuum. Theoretically, the higher a person's class the more RP-like their accent; the lower a person's class, the closer towards the Cockney endpoint their accent is likely to be. Another accent that falls in this linguistic space is "Estuary English" (EE). Estuary English is a term which emerged in attempts to describe and capture processes of language change, and particularly the confluence of RP and Cockney in south-eastern England. It is widely presumed that EE falls closer towards the Cockney end of the continuum than SSBE does, and its name derives from the Thames Estuary that runs out of East London and along the southern border of Essex where Cockney has been heavily influential (Cole and Evans

2020). However, there is no common consensus on the exact social and linguistic delimiters between SSBE and EE.

Wells (1998) defines EE as "standard English spoken with an accent that includes features localizable in the southeast of England" (as cited in Fabricius 2002: 118).

Trudgill (2001) writes that Estuary English

> actually refers to the lower middle-class accents of the Home Counties which surround London: Essex and Kent, which do border on the Thames Estuary, but also parts or all of Surrey, Berkshire, Buckinghamshire, and Hertfordshire, which do not.   (Trudgill 2001: 178)

Agha (2003) writes

> Estuary English is an accent that hybridizes Cockney and RP features. Its speakers exhibit a greater tendency towards traditional Cockney patterns […] superimposed on General RP patterns. The accent is prevalent in the region of the Thames estuary in Southeastern England.   (Agha 2003: 265)

Wells (1997) and Altendorf (2003) describe the linguistic content of EE as having, to some extent, a shifted diphthong system as found in Cockney. All these definitions of EE are not only imprecise, but they could also be applied to SSBE which has also often been described in similarly inexact ways.

Another accent that potentially overlaps with SSBE is Multicultural London English (MLE). MLE emerged as a result of high rates of immigration to London which led to ethnically diverse, multilingual, and multidialectal communities (Cheshire et al. 2008). Though by no means exclusively, MLE features are most frequent in the speech of young ethnic-minority speakers in East London. It is also worth noting that the young MLE speakers of previous studies (e.g. Cheshire et al. 2008; Fox 2015) fulfil the selection criteria of many of the previously mentioned studies on SSBE (for some studies, simply being from southern England would ostensibly classify them as SSBE speakers). In spite of this, there is overlap but also notable differences between the linguistic content that has been documented in MLE and SSBE. MLE coincides with other southern varieties in many ways and includes the Cockney features that have diffused across much of southern England such as having ʟ-vocalisation, ᴛ-glottaling and ᴛʜ-fronting (Cheshire et al. 2008). MLE also includes features not previously documented as part of RP, Cockney, SSBE or EE, such as, in particular, an innovative diphthong system. In MLE, ᴘʀɪᴄᴇ is a narrow diphthong [aɪ] or [ɐɪ], or even monophthongal [æ]; ᴍᴏᴜᴛʜ is now typically a lowered, mid-front monophthong or an innovative back diphthong [ɑʊ]; ꜰᴀᴄᴇ is a narrow diphthong [ẹɪ] or [ɛ̣ɪ]; and ɢᴏᴀᴛ has a raised backed onset [oʊ ~ oː] (Kerswill, Torgersen and Fox 2008; Fox 2015).

The variation between MLE and SSBE is unfalteringly presumed by linguists to be too great to constitute internal variation within a single accent. These accents are always described in the literature as separate entities. In the same way, through the very process of naming them, EE, RP, Cockney and SSBE are often presumed to represent distinctive and discrete varieties despite having overlapping linguistic features and speaker groups. In practice, however, there is no clearly defined set of linguistic features or a set of social characteristics that indisputably demarcate these varieties. The challenges faced in documenting patterns of linguistic variation and defining the accents of Southeast England is further complicated by the pre-selection of speakers that we have discussed above – in order to describe a linguistic variety, we must identify representative speakers, but this process of speaker selection often presupposes the variety's constituent linguistic features.

The objective of this study is to explore patterns of linguistic variation and co-variation in a group of young, south-eastern speakers in a way that is guided by the structure in the data rather than by any pre-conception about the linguistic content of each linguistic variety or who represents it. We work from the tenet that an accent contains both a linguistic and a social component. An accent should be linguistically coherent and should include internal linguistic variation from a centre of gravity. In addition, accents include a social component: they are a set of linguistic features spoken by a certain speaker group. We have chosen the diphthong system as the object of study as they are the most notable delimiter between different south-eastern accents. We draw out subpopulations of south-eastern speakers based on their diphthong systems which we then relate to speakers' social information. For each cluster that emerges we illustrate the centre of gravity for the diphthong system which can be used as a reference point for future research. We also compare the linguistic centres of gravity when an accent is defined with a top-down approach (based on speakers' shared social characteristics) and a bottom-up approach (based on patterns of linguistic coherence).

2.   Methods

2.1   Speakers

Speech production data of a wordlist and passage reading were collected from 193 speakers. The speakers were aged 18–33 years (mean age = 21.8) (born 1986–2001), and they had lived in Southeast England for at least half of the years between the ages of 3 and 18. Southeast England was defined very broadly, permitting speakers from the following places: London, Essex, Hertfordshire, Bedfordshire,

Cambridgeshire, Buckinghamshire, Oxfordshire, Surrey, Berkshire, eastern Hampshire, West Sussex, East Sussex, Kent or southern Suffolk. The speakers' linguistic productions were not taken into account in determining if they were eligible to participate.

As data were collected at the University of Essex, the vast majority of participants were University of Essex staff or students. With few exceptions, speakers were university educated or, for most, were in the process of completing a degree. The participants came from the following demographic groups: 100 female, 93 male; 20 Asian British, 50 Black British, 123 White British; 118 middle class, 75 working class; 12 from fee paying schools, 24 from grammar schools, 157 from state schools; 67 from London, 126 from the Home Counties (defined here as any county in the Southeast except London).

Participants identified their class from fixed choice categories and defined their ethnicity in their own terms. Following this, the speakers were grouped according to the most prevalent groups on the 2021 UK Census: White British, Black British and Asian British. For instance, speakers who considered themselves "British Indian", "British Bangladeshi", "Pakistani British" were grouped as "Asian British" for the purpose of this study. Speakers who used terms such as "Black European", "Black Caribbean", "Black African" or "Black South African" were classified as "Black British". All White British participants used this exact term in their self-identification of ethnicity which the exception of two who identified as "White". In this study, we recognise that, of course, ethnic identities are varied and complex, and that these categories are not monolithic. However, for the purpose of this study, we have followed the broad categories outlined in the 2021 UK Census.

2.2   Items

The stimuli consisted of readings of a wordlist and passage designed to cover the English vowel space (adapted from Chicken Little: Blackwood Ximenes, Shaw and Carignan 2017). Although using productions from read speech as opposed to spontaneous speech restricts us to analysing relatively more formal speech styles, this methodological decision has the advantage of controlling for the phonological environment, when comparing data from different speakers. The wordlist included words in the /h/-V-/d/ or /b/-V-/d/, as well as some high frequency words, for example, *mouth* and *toad*. The wordlist and passage are included in full in the appendix.

Participants were recorded individually while seated in an empty laboratory. They read the passage and then the wordlist from laminated sheets of paper. All recordings were made with a sampling rate of 44.1 kHz, 16-bit resolution

using a Marantz solid state recorder with a lapel microphone. The word list and passage productions were transcribed in ELAN (Max Planck Institute for Psycholinguistics 2019; ver. 5.4) to exclude disfluencies and reading errors. The ELAN files were then used as input for automatic segmentation into time-aligned text-grids using FAVE (Rosenfelder et al. 2014; ver. 1.2). The text-grids were manually checked for major issues with alignment and any errors were hand corrected. Individual vowel boundaries were not checked but alignment was, on the whole, very accurate.

F1 and F2 measurements were then extracted dynamically at 1 ms intervals using FAVE (Rosenfelder et al. 2014). F1 and F2 values were extracted with FAVE default settings which included a maximum of 5 formants up to 5000Hz for males and 5500Hz for females. Function words from the passage productions or words without primary stress were not included in the analysis. As FAVE is based on US varieties of English, the lexical sets corresponding to each vowel were checked and adjusted to reflect typical productions in Southeast England.

2.3   Formant data processing

Statistical analysis was performed in R (R Development Core Team 2016; ver. 4.1.0). The input data, code and complete results are available in an Open Science Framework (OSF) repository at https://osf.io/3azdk/?view_only=8a726f5336b64 fdaa87cdeaa0a6f087a.

Outlier values of F1 and F2 were removed from the data, as they likely represent tracking errors. Outliers were defined as data points that were more than 2.5 standard deviations from the mean. The formant data were then down sampled to 10 per cent intervals (11 equidistant measurements per formant trajectory) and z–scored within speaker (modification of Lobanov 1971). Following this, we extracted by-speaker mean formant trajectories for each diphthong. The mean trajectory values were calculated using a generalised additive model (GAM) and defined as GAM-predictions for normalised F1 and F2, using normalised time as the sole predictor within each group.

By-speaker mean formant trajectories for the PRICE vowel are illustrated in Figure 1. Corner vowels, FLEECE, THOUGHT and TRAP are also included in the figure for reference (the formant values correspond to across-speaker means). As seen in the figure, the realisation of PRICE shows considerable inter-speaker variation, even though the segmental context and style are controlled for. We observe a continuum of variation between fronted and lowered PRICE and retracted-raised realisations.
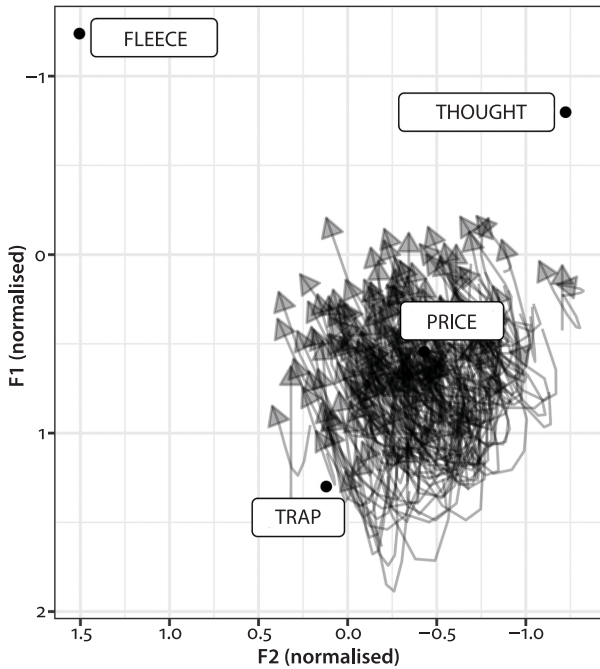
Figure 1. By-speaker mean trajectories of the PRICE vowel (GAM-smoothed), projected onto a normalised vowel space

2.4   Dimensionality reduction

In order to analyse the observed dynamic variation, we reduced it using two-dimensional functional Principal Component Analysis (fPCA; Gubian, Torreira, and Boves 2015). This method parametrises the variation in the shape of formant trajectories, reducing it to a set of principal components (PCs). The input to the analysis were time-varying F1 and F2 trajectories, analysed simultaneously, with a view to capturing co-variation between the first and the second formant. The output of the fPCA is a set of PCs, each of which can be visualised and interpreted in terms of specific dynamic properties of vowel formants. Thus, the PCs correspond to specific linguistic variables that are particularly prominent in the data. As an example, the first two PCs for PRICE are shown in Figure 2.

Figure 2 shows how the formant trajectories are perturbed by variation in PC1 and PC2 values. Perturbations associated with rise in the PC values are illustrated in red. As we can see from the figure, when the PC1 has a positive value, this is associated with overall increase in F1 value, but especially around the vowel acoustic midpoint. Simultaneously, an increase in PC1 value corresponds to an increase in F2 values across the whole F2 trajectory. In contrast, negative val-
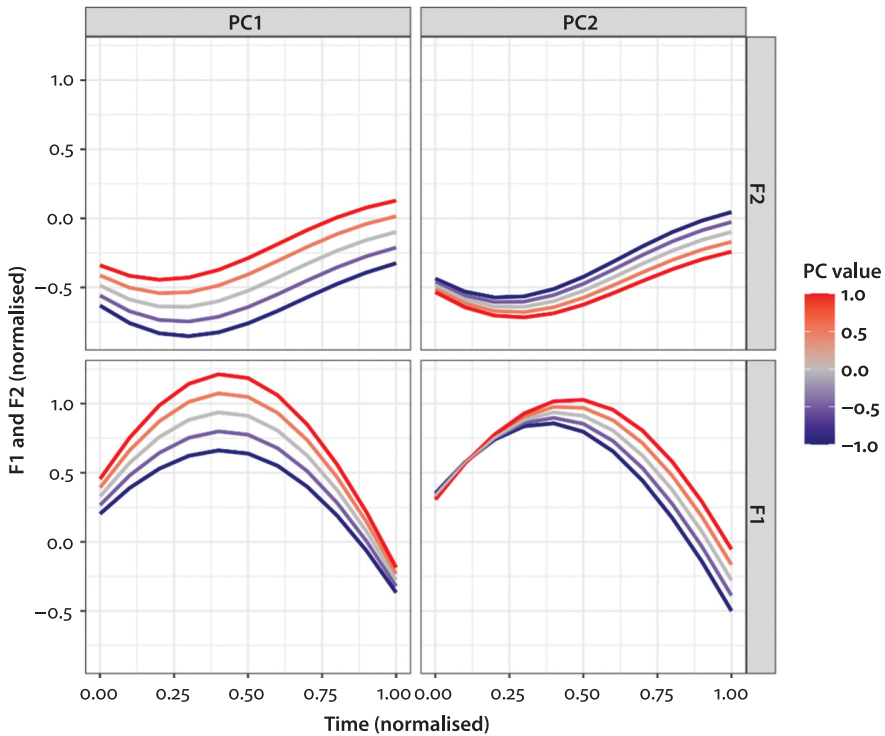
Figure 2. Change in formant trajectories as a function of change in the values of the first two principal components for the PRICE vowel

ues of PC1, marked in blue, correspond to lowering of F1 (especially at the midpoint), and overall lowering of F2. Given the standard interpretation of formants for the auditory properties of vowels, we can thus generalise that the first principal component, PC1, for PRICE corresponds to a pattern of variation between fronted and lowered realisations of PRICE, marked by high PC1 values, and raised and retracted realisations of the same vowel, marked by low PC1 values. PC1 captured 60.82 per cent of variance in mean PRICE trajectories.

The second principal component, PC2, captured 18.70 per cent of variance in PRICE formant trajectories. As seen in the right panel of Figure 2, PC2 picks up on a pattern of variation, in which relatively higher PC2 values are associated with retraction and lowering of the PRICE offglide. This is distinct from the pattern of fronting and lowering (versus retraction and raising) captured by PC1, which affected the entire formant trajectory.

The example of PRICE illustrates that the fPCA translates continuous dynamic variation in formant trajectories to a set of linguistically interpretable variables. These variables are expressed as continuous PC scores. As such, we can use the

PCs as dependent variables to explore the social structure of the observed variance in the realisation of vowels.

## 2.5   Analysis of individual features

While the fPCA tells us which aspects of vowel dynamics vary in the sample, we need additional analysis to understand how these variables map onto social predictors. In order to understand this relationship, we used conditional inference trees (Hothorn, Hornik, and Zeileis 2006). We built a series of trees with PC scores as predictor variables. More specifically, we modelled by-speaker PC means as a function of social information about the speakers: speaker age, gender, ethnicity, social class identity, the type of school they attended and whether or not they came from London (see Section 2.1 for more information on the social predictors included in the study). We used conditional inference trees because they are recursive, which provides an easy way to explore all potential hierarchical relationships (interactions) between the different predictor variables. Figure 3 illustrates the outcome of the analysis for the first two PCs of PRICE.

The nodes in conditional inference tree plots, as shown in Figure 3, represent the independent variables that lead to significant splits in dependent variable. PC1 (top of Figure 3) was significantly affected by ethnicity, although a significant difference was only observed between Asian British and Black British speakers versus White British speakers. The boxplots show the distributions of PC1 values for these sets of speakers. Asian British and Black British speakers show a higher median PC1 value, compared to White British speakers. Combining this information with the interpretation of PCs illustrated in Figure 2, we can generalise that PRICE is fronted and lowered for Asian British and Black British speakers, whereas White British speakers tend to have a relatively more retracted and raised PRICE vowel. The model predicting PC2, shown at the bottom of Figure 3, suggests that PC2 is mainly predicted by speaker gender. Male speakers have higher PC2 values than female speakers. In the context of Figure 2, this can be interpreted as a pattern of fronting and raising of PRICE offglide for females.

The same procedure of analysis and interpretation was repeated for the other diphthongs and their PC scores. We provide a summary of the main findings in Section 3.1.

2.6   Clustering analysis

The method described in Section 2.5 above allows us to explore the social structure of the variation for the individual variables identified in the data. However, we are also interested in potential co-variation between the different variables,
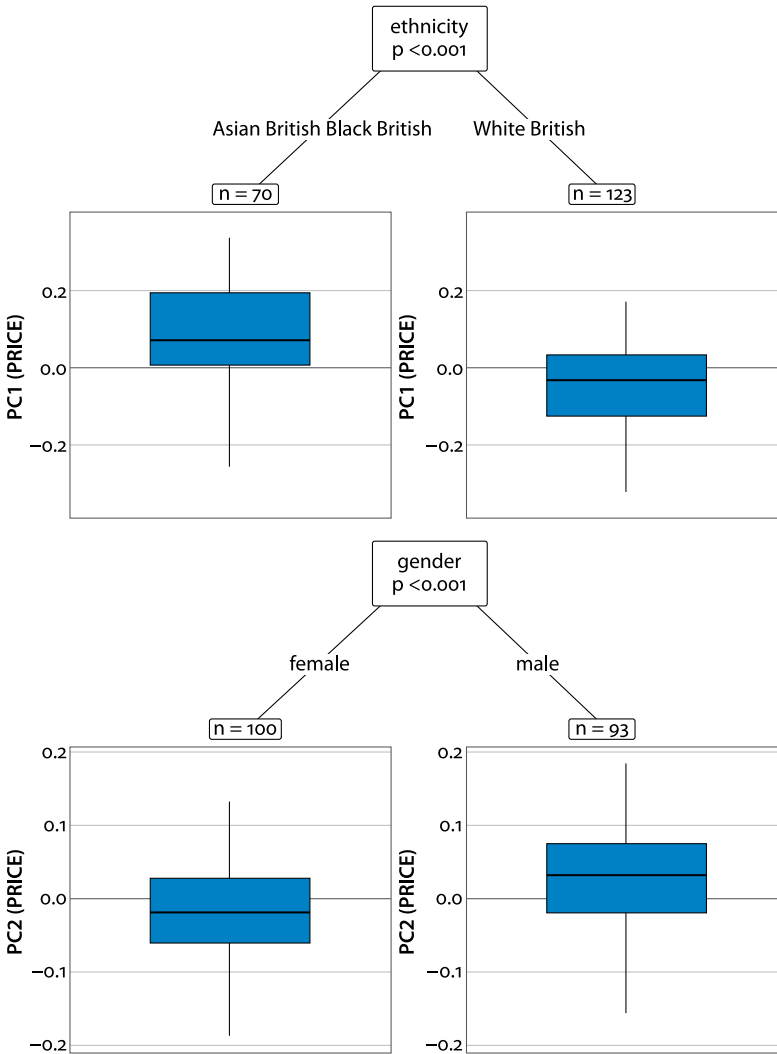
Figure 3. Conditional inference trees predicting PC1 (top) and PC2 (bottom) values for PRICE

considering that accents are typically defined by co-occurrence of specific features. In order to explore such co-variation, we performed a clustering analysis, using Gaussian Mixture Modelling (Fraley and Raftery 2002; Fraley et al. 2012; ver. 4). The input to the analysis were mean by-speaker PC scores for each diphthong. As in Section 2.5, we used the first two PCs for each vowel. The PCs form a 14-dimensional space (7 vowels, 2 PCs), in which an individual speaker is a data point. The goal of the analysis is to establish whether the speakers cluster in this

space (which corresponds to co-occurrence of specific vowel features quantified as PC scores). The speakers are then assigned to different clusters, depending on the optimal number of clusters, as determined by the clustering algorithm.

3.   Results

3.1   Individual variables and their social predictors

Following the method described in Sections 2.4 and 2.5, we identified a number of variables that are systematically conditioned by the social predictors included in the analysis. Significant findings are summarised in Table 1, where "significant" is defined as statistically significant predictors according to the conditional inference tree model. The relative terms in the table should be interpreted in terms of comparison to the other groups of speakers, e.g. "vowel raising in Black British and Asian British speakers" is meant as "raising relative to White British speakers". We do not discuss aspects of variation that did not correspond to any social predictors. The linguistic interpretation of the variables listed in the table is based on a combination of analysing the fPCA outcome (as in Figure 2 and the partitioning analysis (as in Figure 4). The full details of the analysis, including plots illustrating all the individual PCs, as well as the outputs of the partitioning models are available in the supplementary material on OSF.

The two predictors that most systematically condition dynamic vowel variation in our data are ethnicity and gender. Mean vowel formant trajectories (GAM-smoothed), as a function of gender and ethnicity, are plotted in Figure 4. Overall, we can observe a pattern of FACE raising and PRICE fronting and lowering in Asian British and Black British speakers. PRICE-fronting is correlated with variation in height: the fronted PRICE realisations also tend to be lowered. In addition, there is another pattern of height variation in PRICE, conditioned by gender and correlated with PC2 in PRICE. Across all ethnicities, female speakers show a raised offglide in PRICE, compared to males. Similarly, female speakers show a raised offglide in CHOICE across the different ethnicities.

The MOUTH vowel shows a complex pattern of variation conditioned by ethnicity and gender. Asian British and Black British speakers show retraction of MOUTH onglide accompanied by slight raising of the vowel offglide. The same pattern is further differentiated by gender within White British speakers: MOUTH fronting in White British females is less advanced than in White British males, but more advanced than for Asian British and Black British speakers.

The GOAT vowel seems to show some variation in height, similar to FACE, and there is variable GOAT-fronting. These correlate with ethnicity and gender: we find

Table 1. Summary of fPCA-defined variables and their conditioning social predictors. Only significant predictors are listed and described

Variable	Significant predictors	Interpretation
PC1 in FACE	ethnicity	raising of onglide and overall fronting in Black British and Asian British speakers
PC2 in FACE	ethnicity	raising of onglide and offglide retraction in Black British and Asian British speakers
PC1 in PRICE	ethnicity	overall fronting and lowering in Black British and Asian British speakers
PC2 in PRICE	gender	raising and fronting of offglide in females
PC2 in CHOICE	gender	overall raising and slight fronting of offglide in females
PC1 in MOUTH	ethnicity : gender	overall fronting and offglide lowering in White British males, retraction and offglide raising in Asian British and Black British speakers. White British females show an intermediate pattern between these two groups.
PC2 in MOUTH	ethnicity	lowering of the acoustic vowel target in Asian British and Black British speakers
PC1 in GOAT	gender : region	overall fronting and lowering in female speakers, retraction and raising in London males. Intermediate pattern in males from the Home Counties.
PC2 in NEAR	ethnicity	overall raising in Asian British and Black British speakers
PC1 in SQUARE	gender : ethnicity	fronting, raising and monophthongisation in males. Retraction and lowering in White females, and intermediate pattern in Asian British and Black British females

GOAT-fronting and lowering in female speakers, and to a degree also in males from the Home Counties. In contrast, London males tend to have more retracted and raised GOAT.

The centring diphthongs, NEAR and SQUARE, are also subject to conditioning by ethnicity and gender. The NEAR vowel is overall raised in Asian British and Black British speakers. Males generally show raised and fronted SQUARE vowels compared to females. Furthermore, within the female speakers, White females have a particularly lowered and centralised SQUARE.

Some of the features we observe are consistent with previously described variation in the Southeast. For example, MOUTH-fronting and PRICE-retraction, as seen in White British speakers, is a feature of Cockney and thus would theoretically be more common in Estuary English than RP or SSBE. In contrast, PRICE
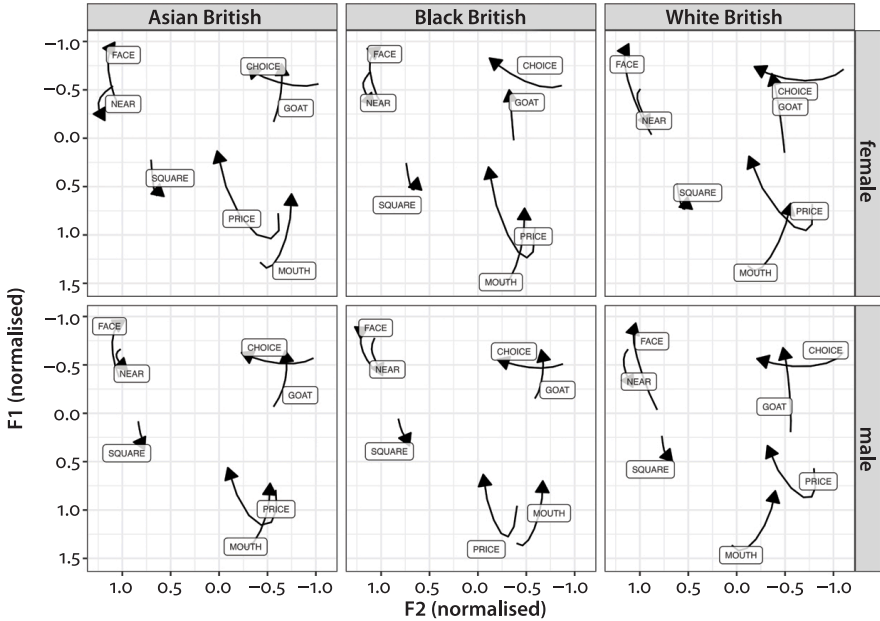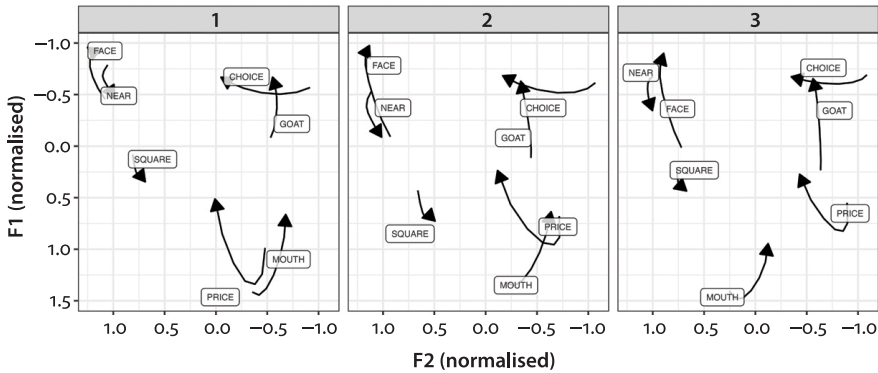
Figure 4. Formant trajectories for the seven diphthong vowels, depending on speaker gender and ethnicity. The trajectories are GAM-smoothed means for each combination of ethnicity and gender in the data

fronting and lowering, particularly prominent in Black British males, and the accompanying retraction of MOUTH are features of MLE. Similarly, FACE raising and fronting is also an MLE feature.

3.2   Clustering results

According to the clustering analysis, the optimal number of clusters in the data is three, which suggests there are three hidden subpopulations of speakers. The mean vowel trajectories for each cluster are visualised in Figure 5.

Cluster 2 represents a combination of features consistent with SSBE, as described by Lindsey (2019). It is closely related to RP but with some influences of Cockney. For example, it is characterised by slight fronting of MOUTH and retraction of PRICE (compared to traditional description of RP), such that the two trajectories are crossed. The GOAT vowel is somewhat fronted, and the NEAR vowel is clearly lowered relative to FACE, and it has a clearly diphthongal quality. Similarly, SQUARE is centralised and clearly diphthongal, showing a change in height throughout the vowel trajectory.

Figure 5. Formant trajectories for the three clusters identified in the data. The trajectories are GAM-smoothed means for each cluster

In comparison, Cluster 3 shows stronger Cockney influences compared to Cluster 2. The most prominent features in this regard are advanced fronting of MOUTH, and accompanying retraction of raising of PRICE. The CHOICE vowel is also raised, and the FACE and GOAT vowels have a lowered onglide relative to Cluster 2. We can generalise that these features reflect Cockney influences but are less extreme than in Cockney such at the diphthongs produced by speakers from working-class, East London families (see, for instance, Cole and Strycharczuk 2022). In addition, speakers in this group do not hail from just London or Essex, but from across Southeast England. Therefore, Estuary English appears to be the most appropriate label for this cluster. Further, NEAR and SQUARE vowels are both raised, compared to Cluster 2, and more monophthongised.

While Clusters 2 and 3 share some similarities, Cluster 1 diverges to a greater degree. In Cluster 1, PRICE is more front than MOUTH, and PRICE is also considerably lowered. FACE are GOAT have raised onglides, with a less diphthongal quality than in Cluster 2. All these features are consistent with MLE, although PRICE and FACE are still clearly diphthongs. However, Cluster 1 (MLE) and Cluster 3 (EE) do share some features. Though the closing diphthongs seem to pattern distinctly, Cluster 1 and Cluster 3 both include raised and somewhat monophthongised centring diphthongs which are not shared by Cluster 2.

In order to explore how the clusters of features map onto social predictors, we fitted a conditional inference tree with the same predictors as described in Section 2.5 but predicting cluster membership, i.e. whether a particular speaker was classified as Cluster 1, 2 or 3 by the clustering algorithm. The result of the model is illustrated in Figure 6. Similarly to the models of individual features reported in Section 3.1, the two significant predictors were ethnicity and gender. Asian British and Black British speakers were most likely to be characterised as

Cluster 1, consistent with the presence of MLE influences in this cluster. White British females were most likely to be classified as Cluster 2, the SSBE cluster. White British males were equally likely to be grouped with Cluster 2 (SSBE) or Cluster 3 (EE). However, despite these tendencies, the clusters are not entirely separable by social predictors: each cluster is a mixture of different ethnicities and genders, albeit in different proportions.
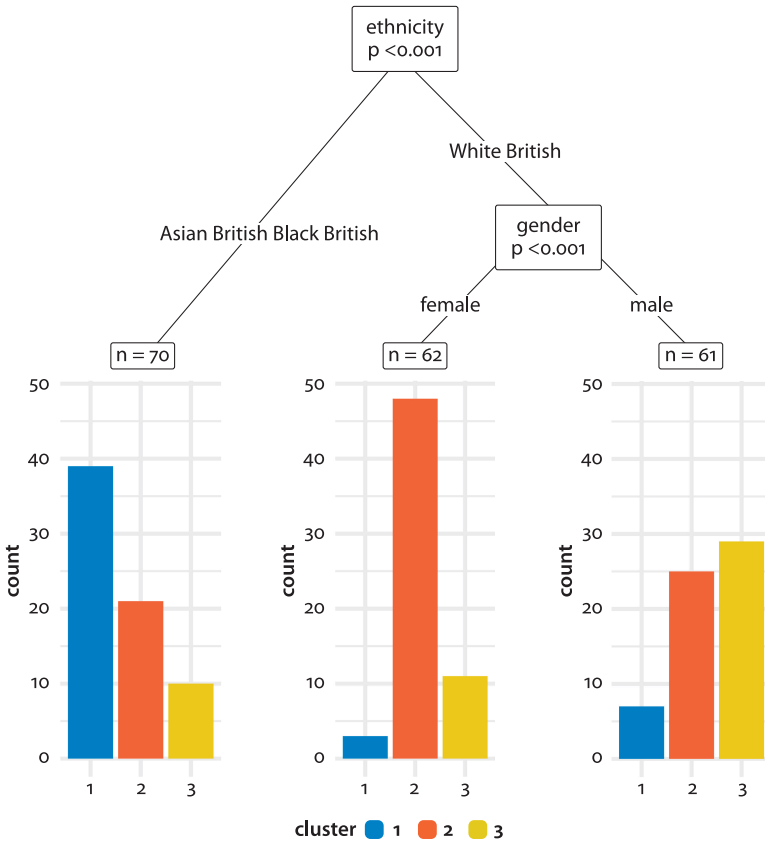
Figure 6. Results of a conditional inference tree predicting cluster membership based on demographic data about speakers

4.   Discussion

In this paper, we have worked from the principle that an accent contains both a linguistic and a social component. The central premise of the linguistic component is that accents should be linguistically coherent: there should be co-variation

between multiple linguistic features (see Guy and Hinskens 2016). This clearly does not require all speakers of a single accent to speak identically. Instead, an accent includes internal variation from a linguistic centre of gravity. In addition, accents include a social component: they are a set of linguistic features spoken by a certain speaker group. Based on co-variation in the diphthong system, we find that within a group of young speakers from Southeast England there are linguistic subpopulations which are broadly related to social predictors. We illustrate the linguistic centre of gravity for these three accents which can be used as reference points in future research (Cluster 1: MLE, Cluster 2: SSBE and Cluster 3: EE).

Cluster 2 (SSBE) certainly differs to the idealised set of linguistic values for RP (such as those first proposed by Alfred C. Gimson (1962) or described as "construct-RP" by Fabricius (2002)). We have labelled Cluster 2 as "SSBE", as it represents the linguistic productions which are most standard and closest to RP within our sample, and the diphthong system closely resembles (some) previous accounts of SSBE (e.g. Lindsey 2019). Relative to RP, there is slight fronting of MOUTH, retraction of PRICE, slight fronting of GOAT, NEAR has a diphthongal quality and is lowered relative to FACE, and SQUARE is centralised and clearly diphthongal. White British females were most likely to be categorised as Cluster 2 and White British males were split between Cluster 2 and Cluster 3. The other two clusters do not align as closely with Lindsey's (2019) account of SSBE. Compared to Cluster 2, in Cluster 3 (EE) MOUTH is lowered and fronted, CHOICE is raised, FACE and GOAT have lowered onglides, NEAR and SQUARE are both raised and more monophthongised, and the retraction and raising of PRICE is more extreme. Cluster 3 is furthest towards the Cockney extreme of the south-eastern continuum. We describe Cluster 3 as "EE" because it fulfils the most consistent definitions of this accent: the observed linguistic productions of Cluster 3 sit somewhere between RP and Cockney and it is spoken by speakers from across the Southeast (e.g. the accent is not just localised to London). In contrast, the vowel system in Cluster 1 is most comparable to previous accounts of MLE (see Kerswill, Torgersen, and Fox 2008; Cheshire et al. 2011). Compared to Cluster 2, FACE are GOAT have raised onglides with a less diphthongal quality, PRICE is considerably lowered and is also more front than MOUTH. Though some Asian British and Black British speakers were categorised as Cluster 2 (SSBE), and an even smaller number were categorised as Cluster 3 (EE), they were most likely to be in Cluster 1 (MLE).

Clusters broadly corresponding to MLE, SSBE and EE have emerged which have collections of co-occurring linguistic features and social predictors in line with previous work. The MLE cluster is most common for Asian British and Black British speakers, the SSBE cluster is most common in White British females, and most White British males are in the EE cluster, with some also falling in the SSBE

cluster. Class did not emerge as an important predictor of cluster membership. It may be that class is not an important linguistic predictor for young people in Southeast England or, at least, not for our group of speakers who were predominantly university students. It is also possible that the way class was defined in this study based on self-identification is not capturing linguistic variation. It is also important to note that the relationship between cluster membership and social factors is only a trend. For instance, though White British women are most likely to be in Cluster 2, speakers from other social groups also fall into Cluster 2, and White British women also sometimes fall into Clusters 1 or 3. It is logical that the most coherent groupings of speakers based on linguistic content does not categorically align with macro-social categories. There is not a direct one-to-one mapping between linguistic and social factors. Nonetheless, the mapping between the diphthong systems and social factors for our three clusters broadly aligns with previous research.

Our findings were also largely as expected for our other stream of analysis in which we investigated variation between social predictors and individual variables. Variation was most notable for gender and ethnicity (which was closely tied to location: 68 per cent of those in London were Asian British or Black British compared to 19 per cent in the Home Counties). In our data, those with productions closest to Cockney tended to be white, male and from the Home Counties. MOUTH-fronting and PRICE-retraction were most prevalent in White British speakers, and the former was also more common in White British males. We also find that MLE-like features are most common among Asian British and Black British speakers, those from London and males. The fronting and lowering of PRICE, and the accompanying retraction of MOUTH and raising of FACE are significantly more likely in Asian British and Black British speakers. PRICE fronting and lowering was particularly prominent in Black British males, and a raised and backed onset for GOAT is common in London males. These are all features previously documented in MLE (see Kerswill, Torgersen and Fox 2008; Cheshire et al. 2011; Fox 2015).

These results concord with common assumptions about linguistic variation in Southeast England: White British speakers in the home Counties tend to use more Cockney-like variants while Asian British and Black British speakers in London tend to use more MLE-like features, and men tend to have more vernacular productions than women. However, like any variationist model, we cannot explain all the linguistic variation in the data with our pre-selected social factors. A top-down, variationist approach can tell us to what extent there is variation in pre-selected linguistic variables between pre-selected social groups. As Cheshire (1999: 65) points out, "in variationist analyses we are limited in what we discover by what we set out to look for." For instance, in previous studies of south-

eastern accents (e.g. EE, MLE, SSBE, Cockney) centring diphthongs are typically not included as variables of interest in the analysis. However, in our analysis we have found that, firstly, NEAR is raised in Asian British and Black British speakers. Secondly, SQUARE is fronter, more raised and monophthongised in males, is retracted and lowered in White females and is found in an intermediate position in Asian British and Black British females. We cannot compare these findings with previous studies as these linguistic variables have not been included in analyses, and we have found a previously undocumented pattern of variation in MLE.

Capacity to uncover such patterns is a key advantage of fPCA, as the method allows us to analyse variation in vowels without any prior expectation about processes that might be involved, such as fronting, retraction, raising, etc. Another advantage of this method is that it can be applied to study F1 and F2 simultaneously as we have done in this study, taking a holistic view of patterns of co-variation between the first two formants. Furthermore, fPCA provides a principled data-driven way of reducing dynamic information in formant trajectories which is especially advantageous in the case of diphthong vowels. As such, this method has potentially very fruitful applications for speech communities where little is known about the structure of sociolinguistic variation and can even provide new insights in the case of well-studied accents like SSBE.

The clustering analysis we have used provides a novel perspective on how distinct accents are in Southeast England. For each of our clusters, we have plotted the centre of gravity which is a defining element of a single accent (Altendorf 2003: 6). In spite of this, the three accents all include internal variation, as well as feature mixing and continuums that run between all our clusters. It is already known that a continuum runs between Cockney and SSBE, but our study is the first to evidence that a continuum also runs between MLE and SSBE. MLE, like RP and Cockney, is a linguistic endpoint while SSBE is a midpoint on a continuum that, like EE, falls somewhere within the space between MLE, Cockney and RP. Because SSBE and EE are midpoints on a continuum, they are not well defined and previous definitions have not been able to sufficiently delimit them. We suggest that the vowel systems observed in Clusters 2 and 3 can be taken as reference points for SSBE and EE diphthong systems respectively. It has previously been suggested that the existence of Estuary English –and perhaps also SSBE – can be falsified by the variation found within areas of the Southeast (Przedlacka 2002; Torgersen and Kerswill 2004). However, variation is an inevitability within single accents. In our approach, we have teased apart three diphthong systems that include internal variation but also include a sufficient degree of variation from each other that our model could delimit them.

As we have seen, linguistic research often recruits eligible speakers based on linguists' intuitions and observations about their accent, even if these are covert,

subsidiary and not made explicit. Though we have superficially provided some sustenance to this approach, the *a priori* exclusion/inclusion of certain speaker groups – be that based on social factors or linguistic productions – in the strive for "authentic" speakers may be problematic for two reasons. Firstly, it reduces the replicability of studies because linguists' intuitions may vary, and, secondly, it may give a view of linguistic variation as more structured and categorical than reality. The pre-selection of speakers in linguistic studies partly accounts for the difference between our MLE cluster (Cluster 1) and previous accounts of the MLE vowel system in the literature. Our MLE cluster is not as extreme as previous accounts of MLE and is closer towards SSBE (i.e. Lindsey's 2019 account). The likely reason for this discrepancy is that the linguistic features often documented and described in the literature as emblematic of an accent are the reference points, i.e. the most extreme speakers who are selected because they fall furthest on the continuum. The *a priori* selection of speakers may filter out some facets of variation and give a view of variation as more categorical than it may truly be. In contrast, we show the averages for each cluster. As a result, though some do, our speakers in our MLE cluster do not all speak the broad version of an MLE vowel system documented in previous research. In addition, our sample is mostly formed of staff and students at a university who may have more standard accents on average than other groups of south-eastern speakers. Further, we analysed read speech which may have produced more standard speech than spontaneous productions. In spite of this methodological approach which may have limited the potential for linguistic variation between speakers, the machine learning algorithm we have employed has still been able to find meaningful differences in the speakers' productions.

It is worth reiterating that the three clusters we have revealed do present coherent diphthong systems, but we cannot consider these fully coherent accents. In order to tease apart fully coherent accents, there should be correlations between diphthongs and monophthongs, as well as consonantal, morphological, syntactic and discourse features, etc. What we have presented, instead, is a representation of the different centres of gravity for the diphthong systems in Southeast England. However, these three clusters do broadly align, in terms of both linguistic content and corresponding social predictors, with SSBE, EE and MLE. In addition, we have worked from the tenet that an accent represents a mapping between social and linguistic factors. We have demonstrated that grouping speakers together using a top-down or bottom-up approach yields different centres of gravity. The three-way, broad concordance in results between, firstly, our cluster analysis, secondly, our more traditional variationist approach and, thirdly, previous research into accent variation in the Southeast, is a sanity check. However,

the optimisation in how social/regional and linguistic features feed into defining an accent is a potential direction for future research.

Funding

References

Agha, Asif. 2003. "The Social Life of Cultural Value". *Language and Communication* 23: 231–273.

Alderton, Roy. 2022. "T-Tapping in Standard Southern British English: An 'Elite' Sociolinguistic Variant?". *Journal of Sociolinguistics* 26: 287–298.

Altendorf, Urike. 2003. *Estuary English: Levelling at the Interface of RP and South-Eastern British English*. Tübingen: Gunter Narr Verlag.

Arvaniti, Amalia, and Madeleine Atkins. 2016. "Uptalk in Southern British English". In Jonathan Barnes, and Alejna Brugos. eds. *Proceedings of Speech Prosody*. Boston: Speech Prosody Special Interest Group, 153–157.

Badia Barrera, Berta. 2015. "A Sociolinguistic Study of T-Glottalling in Young RP: Accent, Class and Education". Ph.D. Dissertation, University of Essex.

Beaman, Karen V., and Gregory R. Guy. 2022. "The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning". In Karen V. Beaman, and Gregory R. Guy. eds. *The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning*. New York: Routledge, 1–13.

Becker, Kara. 2016. "Linking Community Coherence, Individual Coherence, and Bricolage: The Cooccurrence of (r), Raised BOUGHT and Raised BAD in New York City English?" *Lingua* 172: 87–99.

Blackwood Ximenes, Arwen, Jason Shaw, and Christopher Carignan. 2017. "A Comparison of Acoustic and Articulatory Methods for Analyzing Vowel Differences across Dialects: Data from American and Australian English". *The Journal of the Acoustical Society of America* 142: 363–77.

Cheshire, Jenny. 1999. "Taming the Vernacular: Some Repercussions for the Study of Syntatic Variation and Spoken Grammar". *Cuadernos de Filologia Inglesa* 8: 59–80

Cheshire, Jenny, Susan Fox, Paul Kerswill, and Eivind Torgersen. 2008. "Ethnicity, Friendship Network and Social Practices as the Motor of Dialect Change: Linguistic Innovation in London". *Sociolinguistica* 22: 1–23.

Cheshire, Jenny, Paul Kerswill, Sue Fox, and Eivind Torgersen. 2011. "Contact, the Feature Pool and the Speech Community: The Emergence of Multicultural London English". *Journal of Sociolinguistics* 15: 151–196.

doi Chládková, Kateřina, Silke Hamann, Daniel Williams, and Sam Hellmuth. 2017. "F2 Slope as a Perceptual Cue for the Front–Back Contrast in Standard Southern British English". *Language and Speech* 60: 377–398.

doi Cole, Amanda. 2020. "Co-Variation, Style and Social Meaning: The Implicational Relationship Between (h) and (ing) in Debden, Essex". *Language Variation and Change* 32: 349–371.

doi Cole, Amanda, and Bronwen Evans. 2020. "Phonetic Variation and Change in the Cockney Diaspora: The Role of Gender and Identity". *Language in Society* 50: 641–665.

doi Cole, Amanda, and Patrycja Strycharczuk. 2022. "Dialect Levelling and Cockney Diphthong Shift Reversal in South East England: The Case of the Debden Estate". *English Language & Linguistics* 26: 621–643.

doi Deterding, David. 1997. "The Formants of Monophthong Vowels in Standard Southern British English Pronunciation". *Journal of the International Phonetic Association* 27: 47–55.

doi Erker, Daniel. 2022. "How Social Salience Can Illuminate the Outcomes of Linguistic Contact: Data from Spanish in Boston". In Karen V. Beaman, and Gregory R. Guy. eds. *The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning*. New York: Routledge, 145–162.

doi Escudero, Paola, and Kateřina Chládková. 2010. "Spanish Listeners' Perception of American and Southern British English vowels". *The Journal of the Acoustical Society of America* 128: 254–260.

doi Fabricius, Anne. 2002. "Ongoing Change in Modern RP: Evidence for the Disappearing Stigma of T-glottalling". *English World-Wide* 23: 115–136.

Fox, Susan. 2015. *The New Cockney: New Ethnicities and Adolescent Speech in the Traditional East End of London*. London: Palgrave Macmillan.

doi Fraley, Chris, and Adrian E. Raftery. 2002. "Model-Based Clustering, Discriminant Analysis and Density Estimation". *Journal of the American Statistical Association* 97: 611–631.

Fraley, Chris, Adrian E. Raftery, Thomas Brendan Murphy, and Luca Scrucca. 2012. *Mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical Report*. Version 4. <https://www.webpages.uidaho.edu /~stevel/519/literatures/MCLUST%20for%20R.pdf> (accessed September 20, 2022)

Gimson, Alfred. C. 1962. *An Introduction to the Pronunciation of English*. London: Edward Arnold.

doi Gregersen, Frans, and Pharao, Nicolai. 2016. "Lects are Perceptually Invariant, Productively Variable: A Coherent Claim about Danish Lects". *Lingua* 172–173: 26–44.

doi Gubian, Michele, Francisco Torreira, and Lou Boves. 2015. "Using Functional Data Analysis for Investigating Multidimensional Dynamic Phonetic Contrasts". *Journal of Phonetics* 49: 16–40.

doi Guy, Gregory R. 2013. "The Cognitive Coherence of Sociolects: How do Speakers Handle Multiple Sociolinguistic Variables?" *Journal of Pragmatics* 52: 63–71.

doi Guy, Gregory R., and Frans Hinskens. 2016. "Linguistic Coherence: Systems, Repertoires and Speech Communities". *Lingua* 172–173: 1–9.

doi Harrington, Jonathan, Felicitas Kleber, and Ulrich Reubold. 2008. Compensation for Coarticulation, /u/-Fronting, and Sound Change in Standard Southern British: An Acoustic and Perceptual Study". *The Journal of the Acoustical Society of America* 123: 2825–2835.

doi Holmes-Elliott, Sophie and Erez Levon. 2017. "The Substance of Style: Gender, Social Class and Interactional Stance in Southeast England". *Linguistics* 55: 1045–1072.

doi Hothorn, Torsten, Kurt Hornik, and Achim Zeileis. 2006. "Unbiased Recursive Partitioning: A Conditional Inference Framework". *Journal of Computational and Graphical Statistics* 15: 651–674.

Hudson, Toby, Gea de Jong, Kirsty McDougall, Philip Harrison, and Francis Nolan. 2007. "F0 statistics for 100 young male speakers of Standard Southern British English." In Jürgen Trouvain, and William J. Barry. eds. *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany, 1809–1812. Hughes, Arthur, Peter Trudgill, and Dominic Watt. 2012. *English Accents and Dialects: An Introduction to Social and Regional Varieties of English in The British Isles*. London: Routledge.

doi Kerswill, Paul, Eivind Torgersen, and Susan Fox. 2008. "Reversing 'Drift': Innovation and Diffusion in the London Diphthong System". *Language Variation and Change* 20: 451–491.

Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

doi Lindsey, Geoff. 2019. *English After RP: Standard British Pronunciation Today*. Cham: Springer.

doi Lobanov Boris, M. 1971. "Classification of Russian Vowels Spoken by Different Speakers". *The Journal of the Acoustical Society of America* 49: 606–608.

Max Planck Institute for Psycholinguistics. 2019. *ELAN*. Version 5.4. <https://archive.mpi.nl/tla/elan> (accessed September 20, 2022)

Nolan, Francis, Kirsty McDougall, Gea de Jong, and Toby Hudson. 2009. "The DyViS Database: Style-Controlled Recordings of 100 Homogeneous Speakers for Forensic Phonetic Research". *International Journal of Speech Language and the Law* 16: 31–57.

Oushiro, Livia, and Gregory Guy. 2015. "The Effect of Salience on Co-Variation in Brazilian Portuguese?" *University of Pennsylvania Working Papers in Linguistics* 21: 154–166.

Patrick, Peter L. 2002. "The Speech Community". In J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes. eds. *The Handbook of Language Variation and Change*. Malden: Blackwell Publishing, 573–597.

Przedlacka, Joanna. 2002. *Estuary English? A Sociophonetic Study of Teenage Speech in the Home Counties*. Frankfurt am Main: Peter Lang.

R Development Core Team. 2016. *R: A Language and Environment for Statistical Computing*. R package version 4.1.0. <http://www.R-project.org.> (accessed September 20, 2022).

Rosenfelder, Ingrid, Josef Fruehwald, Keelan Evanini, Scott Seyfarth, Kyle Gorman, Hilary Prichard, and Jiahong Yuan. 2014. *FAVE (Forced Alignment and Vowel Extraction)*. Version 1.2. <https://github.com/JoFrhwld/FAVE> (accessed September 20, 2022)

doi Torgersen, Eivind, and Paul Kerswill. 2004. "Internal and External Motivation in Phonetic Change. Dialect Levelling Outcomes for an English Vowel Shift". *Journal of Sociolinguistics* 8: 23–53.

Trudgill, Peter. 1990. *The Dialects of England*. Oxford: Wiley-Blackwell.

Trudgill, Peter. 2001. *Sociolinguistic Variation and Change*. Edinburgh: Edinburgh University Press.

Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. "Empirical Foundations for a Theory of Language Change". In Winfred P. Lehmann, and Yakov Malkiel. eds. *Directions for Historical Linguistics*. Austin: University of Texas Press, 95–195.

Wells, J.C. 1982. *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

Wells, J.C. 1997. "What is Estuary English". *English Teaching Professional* 3: 46–47.

Wells, J.C. 1998. "Pinning Down Estuary English". <http://www.phon.ucl.ac.uk/home/estuary/est-lund.htm> (accessed September 20, 2022).

Williams, Daniel, and Paola Escudero. 2014. "A Cross-Dialectal Acoustic Comparison of Vowels in Northern and Southern British English". *The Journal of the Acoustical Society of America* 136: 2751–2761.

Appendix A. Wordlist

bout hoard hide bite
bead board booed that
out breathe bad earth
anything hit other bored
hat with bore better
little thumb bed tap
toad held though falling
telling heat bar bird
thought bid bud ball
book hat toad hit
hot together gonethick
mouth authority bell bait
better news happy

Appendix B. Passage

One morning Chicken Little was pecking at the earth under the acorn tree, looking for tasty worms, when he felt a sharp WHACK on his fluffy head. "The sky is falling!" cried Chicken Little. His head hurt, and he could feel a big, painful bump on it. "I'd better warn the others!" And off he raced, in a panicked cloud of fluff. "Get out of the bath! Alert the authorities!" squawked Chicken Little to Plucky Ducky. "The sky is falling! It's the end of the world". "Hello there!" quacked Plucky Ducky, puffing out her chest. "Stay calm. I'll save you. I'm not scared of anything." "No time to waste! Or bathe!" squealed Chicken Little, "A chunk of sky just bruised my head." Chicken Little wasn't playing around: Plucky Ducky could see the swelling under his feathers.

"Oh boy," she said. "Never fear, I'm no coward of course. I'll get us out of this alive. No time to think! We'd better grab the others." Hurrying down the path, the pair met Gloaty Goaty digging up the ground with her horns. "It's doomsday!" said Chicken Little "I thought it was Thursday," she wept, "I'm feeling a bit delicate." He began to writhe and tremble in pure, genuine, heartfelt fear. "Have courage," said Plucky Ducky sympathetically, "I'll protect you. No need to nominate me for a medal or reward me or anything, I'm just a good sport. This is my method of operation. Join us! Make haste!" he said proudly. "Oh help, I'm scared," wailed Gloaty Goaty. "I have faith that everything will be just fine," said Plucky Ducky as Gloaty Goaty trotted after them.

Down in the woods they found Perky Turkey, picking flowers. "Hello all! grinned Perky Turkey, who was always happy. "It's an international disaster!" bawled Gloaty Goaty. "We're all doomed." "Soothe yourselves." smiled Perky Turkey. "You ought not get so unsettled. Let's keep our chins up!" "The whole sky's about to crash down." "I'm coordinating the evacuation. You'd better come" assured loyal Plucky Ducky in a smooth voice. "Count me in, it sounds like a ball." said Perky Turkey joyously. Down by the bog, they found Naughty Horsey sitting in a bath of frothy mud. "Evacuate! Evacuate! The sky is falling!" screeched Chicken Little.

Naughty Horsey snorted. "Rubbish! You nutters make me feel normal. Not another drama you thick, paranoid fowl! You always go overboard. Apparently you got hay-fever once, and told everyone you'd caught bird flu." He hit his hooves on the floor, "You haven't thought this through." "But it's true –honestly," wept Gloaty Goaty. He was so rattled he could barely breathe. Naughty Horsey scowled. "Pathetic! It's nothing of the sort! Another false alarm." Gloaty Goaty's lip began to tremble. "It's an utter disaster" said Plucky Ducky as blithe as ever, "But fear not, I brought provisions. We're all in this together. We are brothers in arms. My plan is to relocate everyone to a safe underground shelter, with enough food to last for months."

Naughty Horsey struggled to stand up; he wasn't exactly fit. "Food for months?" he grunted. "Hmm, I do like my grub. Fine, I'll come, if these two bright sparks don't bother me. Oh how I loathe them!" "See!" said Perky Turkey, clapping her wings. "Every cloud has a silver lining!" "You can shut your beak too," glowered Naughty Horsey seething as they all ran away down the lane.

Address for correspondence

Amanda Cole
Department of Language and Linguistics
University Of Essex
Wivenhoe Park
Colchester CO4 3SQ
United Kingdom

amanda.cole@essex.ac.uk
https://orcid.org/0000-0002-8319-3543

Co-author information

Patrycja Strycharczuk
Linguistics and English Language
School of Arts, Languages and Cultures
The University of Manchester

patrycja.strycharczuk@manchester.ac.uk

Publication history