# Automatic sociophonetics: Exploring corpora with a forensic accent recognition system

Georgina Brown[a] and Jessica Wormald[b]
*Department of Language and Linguistic Science, University of York, Heslington, York, YO10 5DD, United Kingdom*

(Received 30 June 2016; revised 22 February 2017; accepted 13 March 2017; published online 31 July 2017)

This paper demonstrates how the Y-ACCDIST system, the York ACCDIST-based automatic accent recognition system [Brown (2015). *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, UK], can be used to inspect sociophonetic corpora as a preliminary "screening" tool. Although Y-ACCDIST's intended application is to assist with forensic casework, the system can also be exploited in sociophonetic research to begin unpacking variation. Using a subset of the PEBL (Panjabi-English in Bradford and Leicester) corpus, the outputs of Y-ACCDIST are explored, which, it is argued, efficiently and objectively assess speaker similarities across different linguistic varieties. The ways these outputs corroborate with a phonetic analysis of the data are also discovered. First, Y-ACCDIST is used to classify speakers from the corpus based on language background and region. A Y-ACCDIST cluster analysis is then implemented, which groups speakers in ways consistent with more localised networks, providing a means of identifying potential communities of practice. Additionally, the results of a Y-ACCDIST feature selection task that indicates which specific phonemes are most valuable in distinguishing between speaker groups are presented. How Y-ACCDIST outputs can be used to reinforce more traditional sociophonetic analyses and support qualitative interpretations of the data is demonstrated.
 [http://dx.doi.org/10.1121/1.4991330]

[CGC] Pages: 422–433

## I. INTRODUCTION

Variationist research in sociolinguistics often focuses on the investigation and exploration of a single linguistic variable at any one time (e.g., Podesva, 2007; Nance *et al.*, 2016). The researcher will examine the data for evidence of different variants of a given variable and interpret what these might mean in terms of the structure of the variety being studied and the identities of the individuals being considered. The examination of multiple variants and findings can also be combined to enable researchers to gauge the similarities and differences between varieties (e.g., Multicultural London English project, Kerswill *et al.*, 2010). While we have accumulated a lot of detailed information about some varieties in this way (e.g., Wells, 1982a,b; Labov *et al.*, 2006; Hughes *et al.*, 2012), this paper introduces one specific automatic tool which could bring different analytical qualities to research, in combination with other more established auditory, acoustic and articulatory methods.

Automatic tools have started to seep into phonetic research. One example is the Forced Alignment and Vowel Extraction (FAVE) suite provided by the University of Pennsylvania (Rosenfelder *et al.*, 2011). Rather than a phonetician manually segmenting speech samples into phone segments through auditory judgment and inspection of a spectrogram, a forced aligner like FAVE can achieve an automatic time estimation of where each phone segment is in a speech sample. This assists in the analysis process as the researcher is able to quickly identify all tokens of a given variable. It is then possible to automatically extract a range of acoustic information (such as estimated formant values) using these time alignments. Although forced aligners are not entirely accurate they provide a good starting point to conduct research, especially with larger corpora.

While sociophonetic researchers are beginning to take advantage of these automatic methods, there is still a wealth of untouched speech technology techniques that could contribute further to our research. We demonstrate this through exploring sociophonetic variation using the York ACCDIST-based automatic accent recognition system (Y-ACCDIST) (Brown, 2014, 2015). The primary intention for Y-ACCDIST is to be used as a supporting tool in forensic casework. In cases where a speech recording of an unknown speaker is evidence, it might be useful to identify the speech community that the speaker may belong to. Given the previous success of Y-ACCDIST, we consider it worthwhile to investigate whether its algorithm could be transferred to sociophonetic analyses.

While other ACCDIST-based systems have been studied for this purpose (Huckvale, 2007b; Ferragne and Pellegrino, 2010), we aim to conduct this exploration of a system's capabilities in more detail by applying it to a corpus of speakers that has already undergone an extensive phonetic analysis. In this way, we can interrogate the system to assess how this tool may or may not complement a phonetic analysis by comparing the output to an already existing analysis. To do this, we apply the Y-ACCDIST system to the PEBL

[a]Electronic mail: gab514@york.ac.uk
[b]Current address: J. P. French Associates, York, UK.

(Panjabi-English in Bradford and Leicester) corpus, which has been studied in great detail by the second author (see Wormald, 2016). Specifically, this paper addresses three key questions and compares the outputs with findings that have arisen through phonetic analysis.

(1) In what ways does a simple accent recognition task corroborate a phonetic analysis in reflecting similarities and differences between varieties?
(2) Can we use Y-ACCDIST to reveal more localised groups of speakers that go beyond labels of geographical origin or language background?
(3) Can Y-ACCDIST indicate which phonemes are most pertinent in distinguishing between different spoken varieties?

We address these questions by utilising different possible outputs of the Y-ACCDIST system. The first two research questions above deal with the idea of assessing speaker and accent similarity, which can then inform us about the accents and the speakers in a corpus. The final research question looks into how we might determine which features are most "useful" in distinguishing between a given set of accent varieties.

This paper first reviews and examines a range of sociophonetic studies that have measured speaker or accent similarity in different ways. We then give an overview of state-of-the-art automatic accent recognition systems to show where Y-ACCDIST is placed within this area of research. The paper then moves onto the methodological details of the analysis we present, which includes a description of the corpus and the inner workings of the Y-ACCDIST system. Following this, we show three types of output Y-ACCDIST can offer, alongside sociophonetic explanations which have been informed by analyses conducted through other methodologies. Together, these explanations illustrate its potential to complement sociophonetic research.

## II. BACKGROUND

### A. Current methods in sociophonetics

Sociolinguistics involves examining variation at all linguistic levels, with sociophonetics largely focusing on the auditory and acoustic analysis of individual phonemes. There are increasingly innovative methodologies being adopted to enable us to more comprehensively analyse these segments (see, e.g., McDougall and Nolan, 2007; Fox and Jacewicz, 2009; Palo *et al.*, 2015; Strycharczuk and Scobbie, 2015; Spinu and Lilley, 2016). Of relevance to the present study is work which has focused on different ways of assessing similarity among speakers in a dataset. Similarity observations are explored in this paper. This section touches on some ways this has been done in previous studies from the perspectives of both production and perception.

#### 1. Distance metrics

The system demonstrated in this paper includes a variant of a specific similarity metric: the ACCDIST metric (Accent Characterisation by Comparison of Distances in the Inter-segment Similarity Table metric) (Huckvale, 2004, 2007a). The ACCDIST metric will be discussed in relation to the whole accent recognition system in Sec. III. The current subsection offers an insight into the use of distance metrics in other studies aiming to use them for sociophonetic research.

Work in dialectometry has sought to measure distances between dialects based on aggregate calculations which incorporate phonetic and acoustic information. Nerbonne (2009) provides a theoretical argument for why aggregate type studies, which consider many variables across different varieties, are crucial if we are to really understand patterns of linguistic variation and change. Nerbonne argues that although we can learn a lot about dialects by looking at one or two variables, it is only by looking at many that we can really begin to understand the relations between them and as such, properly characterise varieties.

Heeringa *et al.* (2009), Wieling *et al.* (2011), and Wieling *et al.* (2012) all measure dialect distance using the Levenshtein distance metric. Phonetic transcriptions of different words across a number of varieties are taken from atlas data. These are then compared by calculating the number of insertions, deletions and substitutions between different transcriptions. For example, as illustrated in Wieling *et al.* (2012), the difference between the Dutch word "autos" pronounced [ɑʊtos] and [othos] is 3. This value corresponds to the number of steps required to get from the first to the second transcription; the first segment [ɑ] is deleted, [ʊ] is substituted for [o], and [h] is inserted (Wieling *et al.*, 2012, pp. 309, 310). Wieling *et al.* (2011, 2012) also then weigh these distances. This weighting results in frequently aligned sounds being assigned a low distance, and infrequently aligned sounds generating a larger distance (Wieling *et al.*, 2011, p. 3).

Both Wieling *et al.* (2012) and Heeringa *et al.* (2009) also include acoustic measures and compare these to the distances calculated based on phonetic transcription. Wieling *et al.* (2012) measure vowel formant frequencies and use the Euclidean distance metric to assess the distance between the different varieties. Heeringa *et al.* (2009) included normalised formant tracks of entire words in addition to zero crossing calculations as means of characterising different varieties. The distances between varieties were then calculated using the Levenshtein distance, with each word being transformed into a series of frames, which were then compared. Heeringa *et al.* (2009) comment that the acoustic measures they use (normalised formant tracks and zero crossings) corroborate with distance measures calculated using tagged and transcribed data.

The key advantage to using distance metrics for this kind of research is that they combine multiple variables, rather than focusing on one or very few to arrive at our conclusions. This same principle is employed in the work presented in this paper.

#### 2. Perceptual similarity

Naive listener perception has also been used to observe similarity among speakers of different accent varieties. To do this, we can ask listeners to assign speech samples to

different categories. Observing these perceptually determined groupings and "errors" that listeners make can reveal relative degrees of similarity between speakers, as well as findings about the listeners themselves. This approach can be seen in Clopper and Pisoni (2004) and in a strand of Hanani *et al.* (2013). However, Clopper and Pisoni (2007) and Clopper and Bradlow (2009) took this a step further, being motivated by the limitations presented by a forced-choice perceptual categorisation task. They invited speakers to undertake a "free classification" task where they could effectively cluster speakers in ways that were not necessarily dictated by traditional, or researcher-led, dialect labels. Results from these tasks can then be visualised through cluster analyses.

The analyses we conduct in this study draw on elements from the principles of the studies just discussed. We hope to contribute an additional method for assessing similarity between speakers of different accent varieties. We do this in combination with techniques employed in speech technology. To address the speech technology component of this paper, Sec. II B offers an overview of automatic accent recognition systems.

## B. Automatic accent recognition

### 1. Text-independent accent recognition systems

The main motivation behind developing automatic accent recognition systems in the past has been to improve the performance of automatic speech recognition systems. Using automatic accent recognition technology before passing a speech sample through an automatic speech recognition system tends to increase automatic speech recognition rates (Najafian *et al.*, 2014).

Until recently, GMM-UBM (Gaussian Mixture Model Universal Background Model) systems were seen as the "de facto reference method" in automatic *speaker* recognition (Kinnunen and Li, 2010), and the study of Chen *et al.* (2001) looked at the performance of a GMM-UBM automatic *accent* recognition system on dialect varieties of Mandarin Chinese. This approach is a way of modelling multidimensional feature vectors, forming an overall probability distribution of the training data. Given unknown data (or test data), a likelihood can be calculated to reflect how likely the unknown data belongs to the same group as a particular training model. Nowadays, systems that implement i-vectors are viewed as the standard model in speaker recognition technologies (Dehak *et al.*, 2011). I-vectors are another type of model. They form a compressed representation of the training data by estimating the components of a GMM supervector that are best for the task of distinguishing between speakers. Naturally, automatic accent recognition research has followed suit and i-vector-based systems have been tested for accent recognition tasks (e.g., Behravan *et al.*, 2015; Bahari *et al.*, 2013; and Najafian *et al.*, 2016).

One advantage of both GMM-UBM and i-vector types of accent recognition systems is that they are text-independent, meaning that they do not require a transcription of the spoken content to accompany the speech sample for processing. This is of course a crucial requirement when considering these systems to assist with automatic speech recognition. The Y-ACCDIST system we present below is text-dependent, requiring a transcription for processing. While this limits the number of applications it can be used for, a comparison of different systems (discussed in more detail below) suggests that text-dependent systems might provide a level of performance that is suitable for sociophonetic research purposes.

### 2. ACCDIST-based accent recognition

Recently, automatic accent recognition systems have been considered for forensic applications (Brown, 2014, 2015, 2016a,b). Forensic analysts are sometimes faced with speaker profiling tasks, which aim to extract information about an unknown speaker in a recording, which might help investigative parties identify a perpetrator. It is possible that automatic accent recognition technologies could assist with these kinds of cases, particularly when the accent varieties in question are not well known or under-researched. Little research exists on possible technologies which could assist with speaker profiling tasks. This has been the motivation behind the research done so far on the Y-ACCDIST system (York ACCDIST-based automatic accent recognition system) (Brown, 2014, 2015, 2016a). Y-ACCDIST is an automatic tool that takes a speech sample, and its corresponding phonemic transcription, and aims to assign an accent label to the speaker. It is based on the ACCDIST metric (Huckvale, 2004, 2007a) which focuses on intra-speaker vowel distances to capture the pronunciation system of speakers of different accents. A more detailed description of how an ACCDIST-based system works is given in Sec. III.

Other ACCDIST-based accent recognition systems have been tested in past studies (Huckvale, 2007a; Hanani *et al.*, 2013), not necessarily with the forensic application as the key consideration. Both studies tested them on the Accents of the British Isles (ABI) corpus (D'Arcy *et al.*, 2004), which contains speakers of accents from 14 locations across the British Isles. On this 14-way accent recognition task, both studies observed performances with classification rates above 90% correct. Brown (2014, 2015) built on this work in two main ways.

(1) The ABI corpus that previous ACCDIST-based accent recognisers had been tested on contained rather dissimilar accent varieties. The work in Brown (2014, 2015) tested Y-ACCDIST on the *Accent and Identity on the Scottish/English Border* (AISEB) corpus of geographically proximate accent varieties (Watt *et al.*, 2014). This corpus contains speakers from four locations along the Scottish-English Border: Berwick-upon-Tweed, Eyemouth, Carlisle and Gretna. The assumption is that the accent varieties here are more similar to one another. Using reading passage recordings from 30 speakers per location Y-ACCDIST achieved a recognition rate of 86.7% correct for this four-way accent recognition task, where the rate expected by chance is 25% (Brown, 2014, 2015).
(2) The system architecture was altered to allow for the processing of content-mismatched (spontaneous) speech data by conducting segmental analysis at the level of the

phoneme, rather than segmental levels more specific than this. Huckvale's system required vowel analysis to be conducted at the level of word-specific vowels, so the vowels in *trap* and *cat* were treated as different vowel segments. This specificity meant that the spoken content of the unknown utterance needed to be identical to the spoken content of the training speech recordings. This is impractical when targeting forensic applications. Y-ACCDIST, on the other hand, collapses vowels into traditional phoneme classes, so these two vowels would be combined and averaged to represent the TRAP (Wells, 1982a) vowel phoneme.[1]

In relation to (2), Brown (2015) tested different versions of ACCDIST-based systems when implementing different degrees of segmental specificity. Using reading passage data from speakers of the four different varieties in the AISEB corpus, three different types of ACCDIST models were formed and tested on the same dataset:

(1) Word-specific vowels (as in, e.g., Huckvale, 2004, 2007a),
(2) Triphone-specific vowels (as in, e.g., Hanani *et al.*, 2013),
(3) Context-independent phoneme categories (the key element distinguishing Y-ACCDIST from other ACCDIST-based systems).

The third implementation is the most versatile in terms of the data it can potentially process (i.e., spontaneous content-mismatched speech), but Brown (2015) compares all three implementations on the same reading passage data to test whether performance is compromised. It is reasonable to hypothesise that, by collapsing phone segments into their phoneme categories, we might see a reduction in accent classification rates because this collapsing of different phones from different phonological environments might lead to more unstable representations (and therefore models). The results in Brown (2015) revealed, however, that this is not necessarily the case. Recognition rates showed that performance was more or less the same, with even slight increases in performance when context-independent phonemes were used.

Studies have also compared ACCDIST-based systems against other types of automatic accent recognition system. First, Hanani *et al.* (2013) compared two ACCDIST-based systems (following the triphone-specific segmental implementation in their models) against a number of different types of automatic accent recognition system, testing them all on the same corpus of accents (the ABI corpus). Influenced by Hanani *et al.* (2013), Brown (2016a) compared two similar ACCDIST-based systems (Y-ACCDIST systems which take on a context-independent segmental modelling approach) with other types of accent recognition system, all on the same corpus of geographically proximate accents (the AISEB corpus). In these studies, a combination of text-dependent systems and text-independent systems were compared. Both studies unsurprisingly found that the text-dependent ACCDIST-based systems outperformed text-independent GMM-based systems. In the study of Hanani *et al.*, a 14-way classification task on the ABI accent varieties, their highest-performing ACCDIST-based system achieved a recognition rate of 95.18%, whereas their GMM-UBM system achieved 61.13% on the same task. In Brown (2016a), on a four-way classification task on the AISEB varieties, her highest-performing Y-ACCDIST system achieved a recognition rate of 87.5%, whereas her GMM-UBM system achieved a rate of 37.5%. The low performance of the GMM-UBM system was put down to the nature of the AISEB data. The AISEB varieties are expected to pose more of a challenge to the systems with respect to an increased degree of similarity among the accent varieties, compared to the varieties found in the ABI corpus.

Given the success of some of these systems when distinguishing between accent varieties, it is reasonable to contemplate whether some of these technologies could offer analytical tools to sociophonetic research. In past studies, we have witnessed the use of ACCDIST-based methodologies applied to more sociolinguistic studies. Huckvale (2007b) showed that through conducting an agglomerative hierarchical cluster analysis with individual speaker ACCDIST models, we can observe expected or meaningful clusters of speakers. A cluster analysis of this kind is demonstrated in this paper using Y-ACCDIST and the PEBL corpus to discover in more detail what these analyses could potentially reveal about speaker populations.

Ferragne and Pellegrino (2010) similarly demonstrated an ACCDIST-based approach to a sociophonetic analysis. Using the same corpus of accents that Huckvale (2004, 2007a,b) and Hanani *et al.* (2013), the ABI corpus, Ferragne and Pellegrino conducted a cluster analysis among these varieties using an ACCDIST-based modelling technique, as well as applying multidimensional scaling (MDS) to be able to see how the different accents cluster within space. They report that their cluster analysis and MDS outputs did not necessarily corroborate one another. For example, the dendrogram from the cluster analysis showed Scottish and Irish accents clustering together, apart from the other accents in the corpus. However, in the MDS output of Ferragne and Pellegrino, this observation did not seem to re-emerge. It is quite possible that different types of analysis serve different functions and pick up on different aspects of the data. This might explain the lack of corroboration between these two types of analysis. By working on a different corpus of accent varieties, where a thorough phonetic analysis has been conducted, this paper aims to assess ACCDIST-based outputs against the expectations and findings that have arisen from the phonetic analysis. This will allow us to discover how an ACCDIST-based approach to sociophonetic analysis corroborates with a phonetic approach. Thus, we can learn more about how it is modelling and classifying speakers, as well as finding out more about the corpus.

## III. METHODOLOGY

### A. The corpus

A subset of 66 speakers from the PEBL corpus[2] was used in the analysis. Speakers from the two British cities of Bradford and Leicester were interviewed as part of a sociophonetic project exploring whether a single heritage

language (in this case, Panjabi) could be said to account for similar patterns of variation observed in the two locations (Wormald, 2016). Within each region, "Panjabi-English" (PE) and "Anglo-English" (AE) speakers took part in paired sociolinguistic interviews, with speaker pairs matched by region, language background and speaker sex. Speakers were evenly distributed between 19 and 53 years of age. The sociolinguistic interview included a number of tasks to gather a variety of linguistic information for analysis.[3] In this article, we use the recordings from the reading passage task, this equates to about two to three minutes of speech per speaker. A modified version of the reading passage "Fern's Star Turn"[4] was used in the analysis. This passage provides a fair duration of speech from each of the participants and was specifically designed for sociolinguistic research and thus incorporates many of Wells' (1982a) keywords.

PE refers to the native English variety spoken by second- and future-generation individuals with Panjabi language heritage. Speakers with Panjabi language heritage here are individuals who have at least one parent who is a first generation migrant from the Panjab region (North-West India and Northern Pakistan) and that the parent is a native Panjabi speaker. PE speakers themselves may not necessarily speak Panjabi, although all participants had some knowledge of the language. In contrast, AE speakers are defined as those with no heritage language other than English, with both parents and grandparents being born in the UK. See Wormald (2016) for a more in depth discussion of this and consideration of the diversity and complexity associated with Panjabi. Table I includes a breakdown of the speakers by region, speaker sex, and language background.

To be able to train the system on enough data to model speaker groups, we initially only focus on the speaker groups determined by their regional and language backgrounds. This has meant that variations in speaker sex and age are not considered.

## B. Y-ACCDIST development details

Brown (2016a) presented two versions of Y-ACCDIST. Which version we select to use is dependent on the nature of the corpus. The version shown and demonstrated in this paper is the first version of Y-ACCDIST, which is more versatile when it comes to analysing corpora of a moderate size and with accent groups containing unbalanced numbers of speakers. This is the Y-ACCDIST-Correlation system described in Brown (2016a). The second version of Y-ACCDIST (Y-ACCDIST-SVM) is recommended for larger corpora with balanced numbers of speakers in each class and incorporates a machine learning method for classification: support vector machines (SVMs) (Vapnik, 1998). SVM classifiers will not work well on smaller corpora because the number of features used in the models would be greater than the number of speakers in our training set (Batuwita and Palade, 2012).

TABLE I. Breakdown of participants from the PEBL corpus included in this analysis.

| | City | |
|---|---|---|
| Speaker group | Leicester | Bradford |
| Male AE | 4 | 5 |
| Female AE | 8 | 5 |
| Male PE | 14 | 9 |
| Female PE | 10 | 11 |
| Totals | 36 | 30 |

Y-ACCDIST-Correlation is, therefore, being implemented for the experiments in this paper. The dataset we are using is of moderate size and we have an imbalance of speakers in each class. Y-ACCDIST-Correlation is much less sensitive to these data properties. Different aspects of the performance of the Y-ACCDIST-SVM system can be found in Brown (2014, 2015, 2016a,b).

We can think about the inner workings of Y-ACCDIST-Correlation in two main stages. Taking our training speakers from our accent corpus (in this case, PEBL), we first *model* our speakers' pronunciation systems, and then move on to the *classification* of an unknown speaker.

### 1. Modelling

*a. Forced alignment.* For each speaker in the training data, the speech sample (the reading passage) is force aligned, using an aligner built using the Hidden Markov Model Toolkit (HTK) (Young *et al.*, 2009) and a Northern English English pronunciation dictionary that was prepared by the second author. This involved collapsing FOOT and STRUT vowels (Wells, 1982a), which are not distinguished in either location, and not including BATH-broadening. Although Leicester is not in the geographical north of England, it is in the linguistic north (e.g., Wells, 1982b) and thus this transcription more appropriately characterises both varieties than would a Received Pronunciation one. The result of the alignment process was a time-aligned phonemic transcription of the speech sample for each speaker. It should be kept in mind that these are just estimated time-alignments of the speech signal, but we find that they serve as sufficient markers for our purpose.

*b. Construction of Y-ACCDIST matrices.* Once each speaker's sample is aligned, we can extract midpoint acoustic features to represent each vowel phone. In more traditional phonetics, formant values might be what is chosen to represent a speech segment. Here, however, we are using mel-frequency cepstral coefficients (MFCCs), which are widely used across speech technology applications. They are short-term spectral features that take the log of the magnitude spectrum, which is then mel-filtered to approximate the shape of the vocal tract at the time the signal is produced.

Having extracted a midpoint MFCC vector (consisting of 12 coefficients) for every vowel in a speaker's sample, we can then compute an average MFCC vector to represent each vowel phoneme in the inventory. Using these average representations, we then organise a table (a matrix) which holds all the vowel phoneme pair combinations that are possible (this is illustrated using just three vowel phonemes in Fig. 1). This allows for the Euclidean distance to be calculated between each pair of phonemes (represented by the averaged

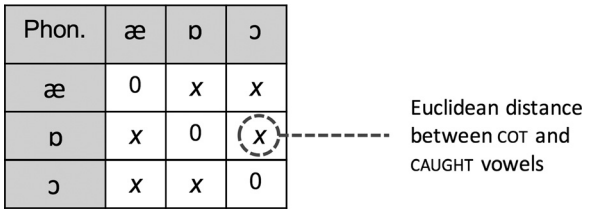


FIG. 1. Illustration of part of a Y-ACCDIST matrix.

12-element MFCC vectors), which effectively aims to capture the degree of similarity between them. This matrix of distance values is expected to characterise the speaker's individual pronunciation traits. To explain, we can take the vowels in COT and CAUGHT in North American English. If we were modelling the speech sample of a typical Pittsburgh English speaker, we would expect a small Euclidean distance between these two vowels, as these two vowels are realised similarly in this variety (e.g., Labov *et al.*, 2006). The Euclidean distance between these two vowels of a New York English speaker, however, is expected to be larger than this because these two vowels are realised differently in this variety. By calculating these distances between all vowel pairs possible, the model should capture a number of these kinds of differences which characterise a speaker's accent.

### 2. Classification

Now we have modelled each of our training speakers' pronunciation systems as Y-ACCDIST matrices, we can use these models to classify an unknown speaker. The first step is to create average Y-ACCDIST matrices to represent each of the accents in our corpus. This was achieved by calculating the mean of each Y-ACCDIST matrix element across all of the speaker matrices in each accent class. In the case of PEBL, we end up with four average Y-ACCDIST matrices each representing a single variety: Bradford PE, Bradford AE, Leicester PE and Leicester AE. Together, these average Y-ACCDIST matrices act as a reference system.

To classify an unknown speaker's speech sample, we then convert the sample into a Y-ACCDIST matrix, in the same way described in the *Modelling* section above. Using this newly formed Y-ACCDIST matrix, we then calculate the Pearson r product-moment correlation (as per Ferragne and Pellegrino, 2010) between our unknown matrix and each of the averaged reference matrices which represent each variety. The correlation measure is intended to indicate the degree of similarity between the unknown matrix and each of the reference matrices. The unknown matrix is therefore assigned the same accent label as the reference matrix with which it generates the highest correlation value.

Figure 2 displays a flow diagram to sum up the overall process.

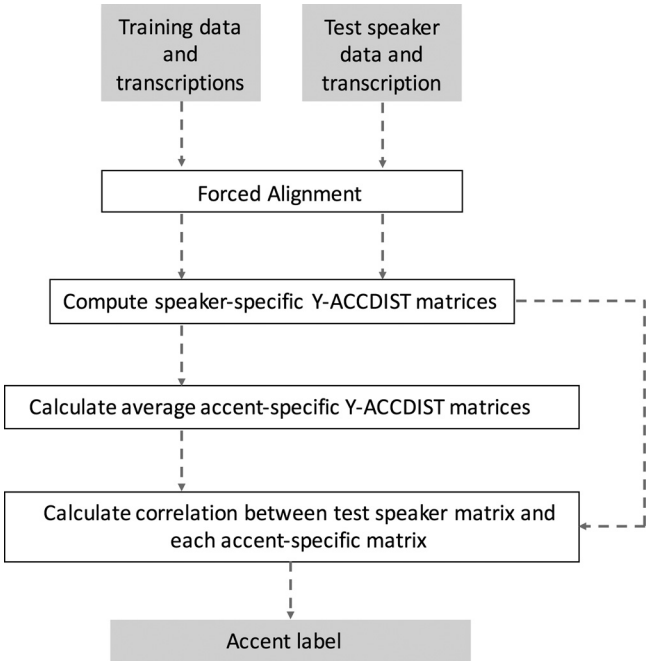


FIG. 2. Y-ACCDIST system flow diagram.

## IV. ANALYSIS

### A. Accent recognition

Simply running Y-ACCDIST as an accent recogniser on the data can indicate the relative degrees of similarity between the varieties we have in our corpus. This approach is loosely similar to that taken in the perceptual similarity experiments by Clopper and Pisoni (2004), where human listeners were asked to make "forced choice" responses to speaker classification tasks.

Y-ACCDIST was trained to classify the PEBL speakers into groups characterised by speaker language background and region. We did this in a *leave-one-out* cross-validation setup, where each of the speakers in our corpus became the "unknown" test speaker on rotation, while the rest of the speakers were used to train the system. On a four-way classification task like this, the recognition rate we can expect if the system was working by chance is 25% correct. On this particular task, Y-ACCDIST achieves a classification rate of 72.7% correct, which is well above chance level. We can take a closer look at this recognition task by inspecting the confusion matrix of Table II showing which categories of speakers were confused for another.

With only one exception, the speakers are categorised into the correct region. The majority of errors within the system arise as a consequence of the system miscategorising the speaker's language background. Using our detailed knowledge of the corpus, we can account for the errors that occur with sociolinguistic reasoning.

The overall behaviour of the system and how it has categorised the speakers is concurrent with the more traditional

TABLE II. Confusion matrix from the Y-ACCDIST classification task. Correct classifications are shown in bold.

| Speaker group | Bradford AE | Bradford PE | Leicester AE | Leicester PE | Total |
|---|---|---|---|---|---|
| Bradford AE | **7** | 3 | 0 | 0 | 10 |
| Bradford PE | 4 | **16** | 0 | 0 | 20 |
| Leicester AE | 1 | 0 | **8** | 3 | 12 |
| Leicester PE | 0 | 0 | 7 | **17** | 24 |

sociolinguistic findings presented in Wormald (2016). In her thesis, Wormald presents the results of analyses from a number of linguistic variables (voice quality, the vowels FACE, GOAT, and GOOSE, and /r/). For each of the linguistic features examined, similarities are observed between the PE speakers in Bradford and the PE speakers in Leicester. Throughout the thesis, however, Wormald argues that although similar patterns are observed in the two PE groups, it is the relationship that these varieties have to the AE in each location which is more prominent. In other words, there is more linguistic similarity between AE and PE in a given region than there is between Bradford PE and Leicester PE. For example, all speakers in Bradford retain monophthongal realisations of FACE and GOAT, whereas in Leicester, all speakers retain a diphthong for these vowels. Within each location, PE speakers have closer and fronter realisations of FACE, and closer but more retracted realisations of GOAT. However, it is not that the two PE varieties have similar realisations, their realisations are locally positioned—it is the relationship to AE which is consistent.

Similarly, with /r/, Wormald (2016) demonstrates that PE speakers in both locations have increased variability and less categoricity in their realisations. However, the type of variation is constrained by the locality, with Bradford PE speakers favouring post-alveolar approximants [ɹ] and taps [ɾ], and Leicester PE speakers more frequently adopting labiodental and post-alveolar approximants [ʋ ɹ]. Wormald (2016) argues that it is the AE varieties which help to predict which additional variants will be found in the PE variety, with [ɾ] found only amongst older Bradford AE speakers, and [ʋ] observed among some Leicester AE females. This idea is supported by the number of intra-region confusions we see in the confusion matrix above, compared to the number of inter-region confusions, and is also consistent with other studies exploring contact varieties in the UK (e.g., Stuart-Smith *et al.*, 2011). There are many more intra-region confusions than inter-region ones, reinforcing that AE and PE speakers within a given region are more linguistically similar than PE speakers across different locations.

Interestingly, we can also sociolinguistically account for the one speaker who has been miscategorised by region. We argue that this is an understandable error made by the system. This particular speaker has parents and grandparents from Yorkshire and describes his accent as "Leicester with a northern edge." Thus, it "makes sense" that the system has miscategorised him for a Bradford AE speaker. Bradford is in Yorkshire, the county in which the miscategorised speaker has family, with this speaker being more linguistically similar to Bradford AE speakers than other Leicester AE speakers are to Bradford AE speakers. Thus, although he has been miscategorised as being from Bradford, it is almost expected, given the linguistic similarity between this speaker and Bradford AE speakers. Running this type of analysis tells us something about the performance of the system—that it is performing well and that "errors" are not necessarily a reflection of the inadequacy of the system, but can be explained by knowledge of the data. This type of analysis also tells us something about the speakers—that this male is indeed quite linguistically different from the other Leicester AE speakers.

We have shown here that the system's performance corroborates findings which were found through more traditional methods, showing Y-ACCDIST's potential as a sociophonetic research tool. In this instance, the Y-ACCDIST classification task has been undertaken after a large amount of more traditional sociolinguistic analyses have been conducted, so we can show how its performance compares with analysis done in Wormald (2016). It might be useful to run Y-ACCDIST classification as a data screening stage prior to more traditional analyses to fuel research hypotheses. The speed with which this can be undertaken makes it an appealing additional method.

### B. Y-ACCDIST cluster analysis

As well as the recognition outputs above, we can use Y-ACCDIST's modelling technique of representing individual speakers' pronunciation systems in the form of Y-ACCDIST matrices to perform a cluster analysis. We can liken this analysis to the sorts of "free classification" tasks human listeners have been asked to do in perceptual similarity experiments (e.g., Clopper and Pisoni, 2007; Clopper and Bradlow, 2009). The type of cluster analysis we have used here is an agglomerative hierarchical cluster analysis, the same which can be seen in Huckvale (2007b). It is a *bottom-up* analysis where it starts at the individual speaker level, and gradually makes larger and larger clusters by pairing up clusters based on the highest degree of similarity. We can then inspect these clusters in the form of a dendrogram for possible sociolinguistic relationships.

The dendrogram in Fig. 3 displays some potentially meaningful groupings. The work undertaken by the second author who collected the corpus has, up to now, been primarily phonetic with a theoretical approach often taken to the interpretation of the patterns. Little in-depth consideration of within-group deviations and identity-based variation has been done. However, the fieldwork was conducted at a number of different centres within each location, with different communities of practice reflected in the 66 speakers included. Although some of what the dendrogram shows does not, at present, make a great deal of sense, according to the knowledge that we have, there are several clusters which seem to reflect different communities of practice and more localised networks.

In Bradford, the majority of participants, both AE and PE, were sourced from a single community centre. Leading on from this, one potential criticism of this study could be that similarities that are logged between speakers are to do with the fact that they were recorded in the same room, and this might have had effect on the overall system performance. Indeed, all of the participants recorded from a single community centre were interviewed in the same room. However, not all of these "same-centre" speakers cluster together on the dendrogram. Instead, the clusterings reflect internal within-group variations and correspond to either small communities of practice or linguistic similarity.

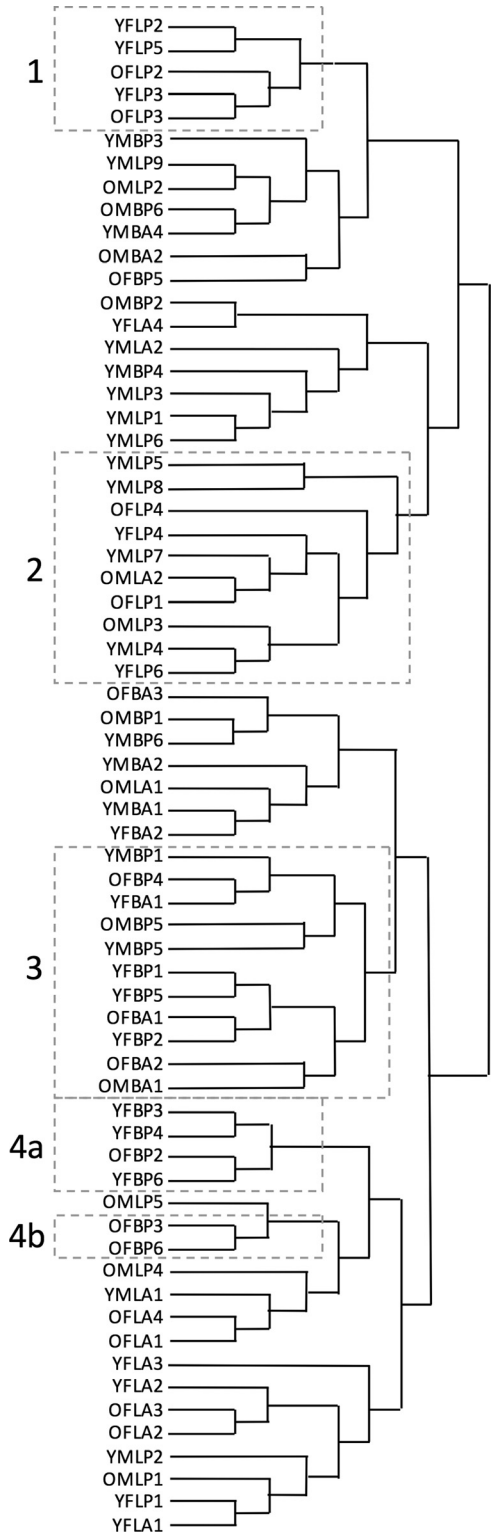


FIG. 3. Dendrogram illustrating speaker clusters. Speakers are identified by their codes, the first letter corresponds to age, Y ≤ 30 and O ≥ 30; the second letter corresponds to speaker sex, M = male and F = female; the third letter corresponds to region, B = Bradford and L = Leicester; and the final letter corresponds to language background, P = Panjabi-English and A = Anglo-English. Numbers were attributed based on the order in which speakers were recorded. The numbered boxes indicate clusters of interest that are discussed below.

Figure 3 shows separate clusters that appear to reflect a combination of linguistic and social factors (clusters 3, 4a, and 4b). All of the speakers in these clusters were recorded in the same room. However, the clusters seem to reflect further sub-groups based on linguistic similarity and correspond to speakers who spend more time together. Other speakers who were recorded in the same room but who interact with different members of the community centre and have slightly different patterns of linguistic variation appearing in a different cluster. This has not, as yet, been quantified and this assertion is based on the second author's experience during the fieldwork collection.

In Leicester, the majority of PE participants were recruited from a single location, a large Gurdwara (Sikh temple) in the city. In this instance, participants were interviewed in an array of different rooms throughout the Gurdwara. However, the different clusterings of Leicester PE speakers once again appear to reflect more nuanced patterns of linguistic variation and correspond to who people actually spend time with. Clusters 1 and 2 in Fig. 3 include 14 different speakers recorded in eight different rooms. However, the separate clusters appear to correspond to two different groups, defined by linguistic similarity and more localised speaker networks.

We believe these preliminary results highlight new and exciting prospects for those of us looking at linguistic variation. The clusters identified and discussed appear to correspond to within-group patterns of variation, with smaller communities of practice becoming apparent. There are currently a number of unexplained clusters in the dendrogram, with this being partly a reflection of the second author not having yet fully explored the qualitative data which might illuminate the more nuanced within-group variation in the corpus. This is also likely to be a consequence of the macro-analytical nature of the methodology, where noise in the analysis is inevitable.

### C. Feature selection

In speech technology, the primary purpose of integrating feature selection into a system like this is to improve recognition rates or to lower computational cost. It is a means of calculating the most valuable "features" in a process, so we can just include those and remove the features which do not add any value to the analysis. For accent recognition, instead of a linguist deciding which features are most diagnostic of accent varieties, we can input all features we have available into the modelling phase of a system, and then compute the ranking of these features that are expected to distinguish the given varieties. This can be executed in a number of ways, and this section presents just one.

In the case of accent recognition with Y-ACCDIST, the "features" are the distance values in the Y-ACCDIST matrix (the Euclidean distance between each phoneme-pair combination). By only including the phoneme-pair distances which are distinctive in the particular accent recognition task we are conducting, it is expected that recognition rates will increase because we remove the phoneme-pairs which do not add anything to the task. This has been shown using the AISEB corpus in Brown (2016a). Keeping phoneme-pair distances which do not help to distinguish between varieties is expected to introduce "noise" to the analysis.

In addition, by integrating feature selection, we can remove the number of assumptions the system makes about the particular varieties involved. Past ACCDIST-based systems (as well as past experiments using Y-ACCDIST) have only included vowel segments to construct the accent models (Huckvale, 2004, 2007a; Hanani *et al*., 2013; Brown, 2015). Consonants have been discarded. Of course, vowels are expected to play a key part in distinguishing between varieties of British English, but sociophonetic research demonstrates that consonants also assist in distinguishing between accent groups. Feature selection allows us to avoid making these kinds of segmental assumptions *a priori* and include all segments that exist, both vowels and consonants, and then let the system indicate which phoneme-pairs are most likely to contribute in any subsequent classification task. Consonants have therefore been included in this analysis in the same way as vowels have: each phoneme is represented by an average midpoint MFCC vector.

While including a feature selection step sets out to improve system performance, it could also provide a useful guide for sociophonetic research. Using the output ranking of matrix elements from feature selection, we can achieve a general picture of which phonemes might be most valuable in distinguishing between the varieties. Thus, this provides a preliminary analysis and could act as an additional tool to guide future research when considering which features to examine in detail.

Feature selection can be conducted in numerous ways, but in this paper we present just one: analysis of variance (ANOVA). ANOVA was demonstrated as a suitable method of feature selection for automatic accent recognition in Wu *et al*. (2010). ANOVA can be used to assess each Y-ACCDIST matrix element to see whether it is significantly different between the accent groups. A p-value can be generated for each matrix element to indicate the degree of significance. Plenty of other ways to implement a feature selection step exist. In particular, when using the Y-ACCDIST-SVM version of the system, we could apply SVM-specific methods (see Brown, 2016a). Figure 4 is the resulting heatmap after performing ANOVA on the Y-ACCDIST matrices representing speakers from each of the four PEBL varieties. The darker the cell, the higher the ranking that matrix element obtained from the feature selection process. Thus, those cells which appear darker represent the features which the system has determined to be most useful in distinguishing between Bradford PE, Bradford AE, Leicester PE and Leicester AE. Tables III and IV provide a mapping of the phoneme symbols used to their relevant IPA symbols.

Of course, it is important to note that a heatmap like this can only offer approximate indications of which phonemes might be of interest. In the heatmap above, we look for the darker rows and columns of the matrix to achieve a rough idea of which segments might be more valuable in distinguishing between the given accent varieties. When running feature selection on the PEBL corpus, a number of linguistic features were identified as potentially useful. Some of these will be discussed later.

The GOAT vowel and the consonant /r/ rank relatively highly on the feature selection output, indicating that these features are useful when distinguishing between groups.

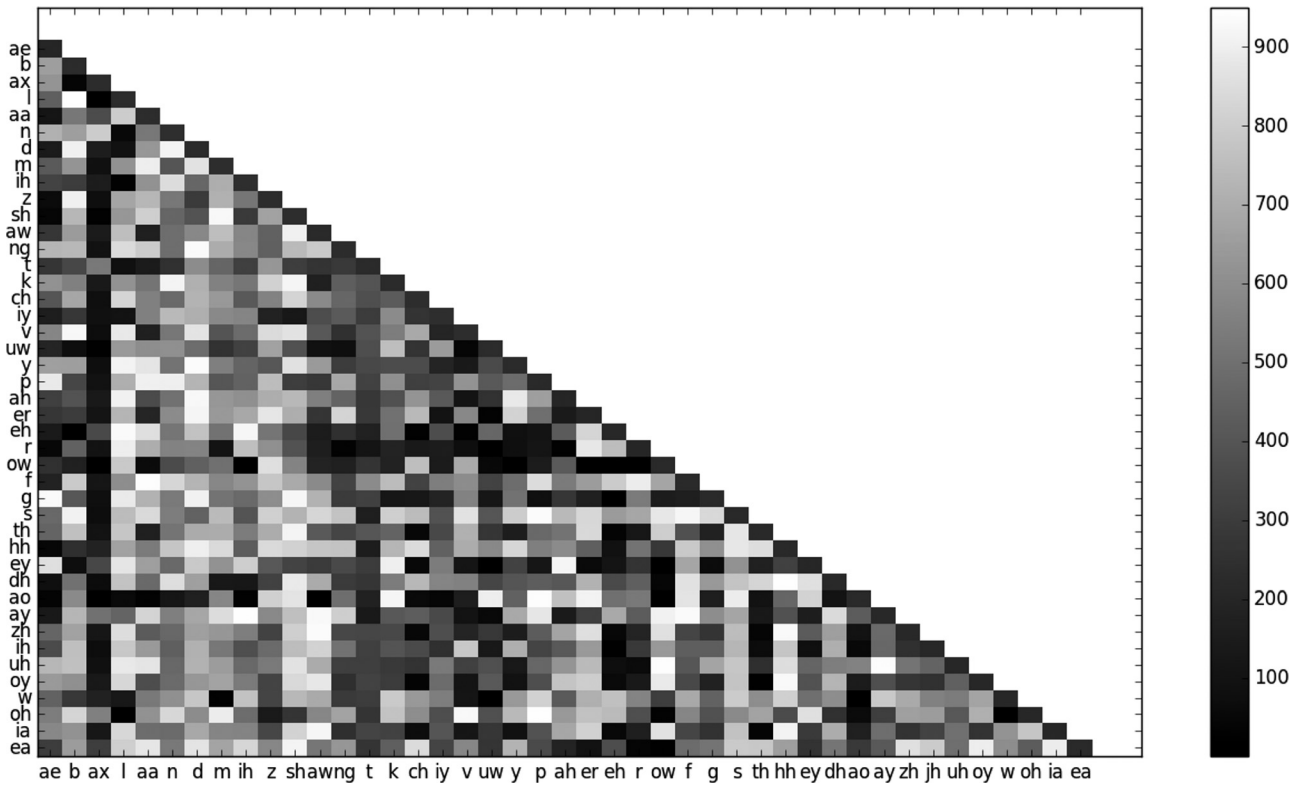


FIG. 4. Resultant heat map after performing ANOVA-based feature selection on Y-ACCDIST matrices.

TABLE III. Mapping of IPA symbols to the consonant symbols used in Fig. 4.

| IPA symbol | Symbol |
|---|---|
| /p/ | {p} |
| /b/ | {b} |
| /t/ | {t} |
| /d/ | {d} |
| /k/ | {k} |
| /g/ | {g} |
| /m/ | {m} |
| /n/ | {n} |
| /ŋ/ | {ng} |
| /f/ | {f} |
| /v/ | {v} |
| /θ/ | {th} |
| /ð/ | {dh} |
| /s/ | {s} |
| /z/ | {z} |
| /ʃ/ | {sh} |
| /ʒ/ | {zh} |
| /tʃ/ | {ch} |
| /dʒ/ | {jh} |
| /j/ | {y} |
| /w/ | {w} |
| /l/ | {l} |
| /ɹ/ | {r} |
| /h/ | {hh} |

This concurs with the findings reported in Wormald (2015, 2016) who undertook a more traditional acoustic and auditory analysis of these features to characterise group patterns.

Wormald (2016) discusses how for both of these features, each of the varieties have slightly different realisations. For GOAT, all Bradford speakers retain a monophthong with qualitative variation between the PE [oː]-like variant and the AE [o̟ː]. In Leicester, all speakers retain a diphthong and qualitative differences serve to distinguish between the PE [əʉ̟] and AE [ɜʉ̟] variants (Wormald, 2016). With regards to /ɹ/, Wormald (2016) describes a complex pattern of fine phonetic variation, with PE speakers exhibiting less categoricity in their realisations of /ɹ/ than AE speakers. Moreover, Bradford PE speakers are more likely to use the tapped [ɾ] in addition to the post-alveolar approximant [ɹ], whereas Leicester PE speakers are more likely to use labialised approximants like [ʋ] or [ɥ]. Thus, the results from the feature selection corroborate with the findings of more traditional analyses—these features are useful in distinguishing between and characterising the four groups considered here.

TABLE IV. Mapping of IPA symbols to the vowel symbols used in Fig. 4, along with their corresponding keyword from Wells (1982a).

| IPA symbol | Symbol | Keyword |
|---|---|---|
| /ɜ/ | {er} | NURSE |
| /ɒ/ | {oh} | LOT |
| /ə/ | {ax} | Schwa (+lettER +commA) |
| /æ/ | {ae} | TRAP |
| /eɪ/ | {ey} | FACE |
| /uː/ | {uw} | GOOSE |
| /ɔ/ | {ao} | THOUGHT/NORTH |
| /iː/ | {iy} | FLEECE |
| /əʊ/ | {ow} | GOAT |
| /ʌ/ | {ah} | STRUT |
| /ɪ/ | {ih} | KIT |
| /e/ | {eh} | DRESS |
| /aɪ/ | {ay} | PRICE |
| /ɑː/ | {aa} | BATH/PALM |
| /ɛ/ | {ea} | SQUARE |
| /ɔɪ/ | {oy} | CHOICE |
| /ɪə/ | {ia} | NEAR |
| /aʊ/ | {aw} | MOUTH |
| /ʊ/ | {uh} | FOOT |

The feature selection process could also be undertaken with more exploratory aims. Both THOUGHT and, in particular, /ə/ are ranked highly indicating that these are useful group discriminants. These have not been comprehensively examined by the second author, but the output of the feature analysis suggests that these vowels would be worth considering in further detail, with this potentially leading onto a more in-depth investigation.

Interestingly, /l/ is shown to have low rankings in the feature selection output, which suggests that this is not a useful feature when discriminating between groups. At first glance, this appears to be at odds with much research on South Asian Englishes spoken in the UK (e.g., Kirkham, 2017; Stuart-Smith *et al.*, 2011; Kirkham and Wormald, 2015; Heselwood and McChrystal, 2000). Initially, we considered whether the quantity of tokens may be too small, if this were to be the case its distinctiveness may be inhibited, purely as a reflection of the lack of information provided to the system. However, the reading passage used in the data collection contained 67 instances of /l/. Importantly though, this number includes all contexts of /l/ as these are not distinguished by the system in the feature selection, and this particular example might be revealing a potential flaw in Y-ACCDIST's modelling approach of collapsing segments into their phoneme categories. Onset and coda /l/ here are considered together. The allophonic variation associated with /l/ could mean that the model is unable to stabilise the variation. The result might also mean that the degree of variation associated with /l/ means that it would be more useful as a *speaker* discriminant rather than a *group* discriminant, although additional work would need to be carried out to fully explore this idea.

This section has demonstrated how feature selection, derived through the use of an automatic accent recognition system, can complement and support traditional sociophonetic analysis. It is pertinent at this point to highlight that although the feature selection can point us towards interesting directions which could be valuably pursued in future research, the discussion of /l/ demonstrates that there are still developments to be made in this area. It also leads us to consider potential further developments of Y-ACCDIST. It is important to note that potentially distinctive information has been missed by Y-ACCDIST only making use of midpoint acoustic features, ignoring dynamic information. Despite this, the observations for /ɹ/ and GOAT, seem to confirm

results from other sociophonetic studies, while the observations for THOUGHT and /ə/ suggest interesting directions for future research. Together, it appears that the feature selection stage might have something to offer sociophonetic research.

## V. DISCUSSION

Although originally intended for forensic applications, this paper has demonstrated how Y-ACCDIST corroborates with a number of sociophonetic findings by applying it to the PEBL corpus, which had already been thoroughly investigated through other more traditional phonetic methods. First, we showed how the results of the confusion matrix, derived from performing automatic accent recognition on 66 speakers in the PEBL corpus, was consistent with the interpretation put forward in Wormald (2016) about those particular varieties. The fact that there are many more confusions within location, rather than within heritage language groups reinforces the idea that there is a greater degree of similarity between the varieties within a given location. Following this, the Y-ACCDIST cluster analysis revealed groupings of speakers, with these potentially corresponding to within-group communities of practice and more fine-grained sociophonetic variation. Last, the output from feature selection indicated some of the phonemic segments that Wormald (2016) identified as key distinguishing variables through other methods. It also highlighted segments which were not investigated at all in Wormald (2016) (e.g., /ə/), and therefore may have instigated a research direction, which could have otherwise gone uninvestigated.

Of course, Y-ACCDIST can only provide a macro-level analysis, where it takes a number of linguistic variables at once and quickly generates a general picture of an accent corpus. It cannot provide the kind of micro-level analysis of individual variables that can be found in Wormald (2016). There are therefore interesting details that Y-ACCDIST overlooks about these varieties. For example, we are left with uncertainty about the discriminatory power of /l/ in the PEBL varieties. A micro-level analysis could alleviate this and establish /l/'s realisational distribution among these accents. Y-ACCDIST should not, therefore, directly replace these kinds of analyses, but be used in parallel or as an initial screening stage. We propose that it can play a complementary role in an analysis. In the same way that the continuing development of Y-ACCDIST aims to support forensic analysis, we have presented evidence that it can also support and enhance sociophonetic research.

## ACKNOWLEDGMENTS

The authors would like to thank Dr. Dominic Watt and Professor Peter French who provided valuable insights and guidance at various stages of our research. We would also like to thank the anonymous reviewers for their constructive input. Any errors remaining in this article are our own. The research presented here was supported by both the Economics and Social Research Council and the Wolfson Foundation.

[1]Throughout this paper, we refer to specific vowel sounds using Wells' keywords (Wells, 1982a). These are frequently used as reference points among sociolinguists. See Table IV for a list of these words with accompanying IPA symbols.

[2]The results reported here only include speakers who completed the reading passage, not all speakers in the corpus completed this task and as such, only a subset of the corpus has been used.

[3]Interested readers are referred to Wormald (2016, Chap. 3) for more details on the corpus.

[4]This passage was written by Dominic Watt and Carmen Llamas at the University of York. Interested readers are invited to contact these authors for more information and/or access to the passage.

Bahari, M., Saeidi, R., Van Hamme, H., and Van Leeuwen, D. (**2013**). "Accent recognition using i-vector, Gaussian mean supervector and Gaussian posterior probability supervector for spontaneous telephone speech," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Vancouver, Canada, pp. 7344–7348.

Batuwita, R., and Palade, V. (**2012**). "Class imbalance learning methods for support vector machines," in *Imbalanced Learning: Foundations, Algorithms and Applications*, edited by H. He and Y. Ma (Wiley-Blackwell, Oxford), pp. 83–100.

Behravan, H., Hautamäki, V., and Kinnunen, T. (**2015**). "Factors affecting i-vector based foreign accent recognition: A case study in spoken Finnish," Speech Commun. **66**, 118–129.

Brown, G. (**2014**). "Y-ACCDIST: An automatic accent recognition system for forensic applications," MA thesis, University of York, York, UK.

Brown, G. (**2015**). "Automatic recognition of geographically-proximate accents using content-controlled and content-mismatched speech data," in *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, UK, Paper No. 458.

Brown, G. (**2016a**). "Automatic accent recognition systems and the effects of data on performance," in *Proceedings of Odyssey: The Speaker and Language Recognition Workshop*, Bilbao, Spain, Paper No. 29.

Brown, G. (**2016b**). "Exploring forensic accent recognition using the Y-ACCDIST system," in *Proceedings of the 16th Speech Science and Technology Conference*, Sydney, Australia, pp. 305–308.

Chen, T., Huang, C., Chang, E., and Wang, J. (**2001**). "Automatic accent identification using Gaussian mixture models," in *Proceedings of IEEE Workshop on Automatic Speech Recognition and Understanding*, Italy.

Clopper, C., and Bradlow, A. (**2009**). "Free classification of American English dialects by native and non-native listeners," J. Phon. **37**, 436–451.

Clopper, C., and Pisoni, D. (**2004**). "Some acoustic cues for the perceptual categorization of American English regional dialects," J. Phon. **32**, 111–140.

Clopper, C., and Pisoni, D. (**2007**). "Free classification of regional dialects of American English," J. Phon. **35**, 421–438.

D'Arcy, A., Russell, M., Browning, S., and Tomlinson, M. (**2004**). "The Accents of the British Isles (ABI) corpus," in *Proceedings of Modelisations pour l'Identification des Langues*', Paris, France, pp. 115–119.

Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., and Ouellet, P. (**2011**). "Front-end factor analysis for speaker verification," IEEE Trans. Audio Speech Lang. Process. **19**, 788–798.

Ferragne, E., and Pellegrino, F. (**2010**). "Vowel systems and accent similarity in the British Isles: Exploiting multidimensional acoustic distances in phonetics," J. Phon. **38**, 526–539.

Fox, R. A., and Jacewicz, E. (**2009**). "Cross-dialectal variation in formant dynamics of American English vowels," J. Acoust. Soc. Am. **126**(5), 2603–2618.

Hanani, A., Russell, M., and Carey, M. (**2013**). "Human and computer recognition of regional accents and ethnic groups from British English speech," Comput. Speech Lang. **27**, 59–74.

Heeringa, W., Johnson, K., and Gooskens, C. (**2009**). "Measuring Norwegian dialect distances using acoustic features," Speech Commun. **51**, 167–183.

Heselwood, B., and McChrystal, L. (**2000**). "Gender, accent features and voicing in Panjabi-English bilingual children," Leeds Work. Pap. Ling. Phon. **8**, 45–70.

Huckvale, M. (**2004**). "ACCDIST: A metric for comparing speakers' accents," in *Proceedings of the International Conference on Spoken Language Processing*, Jeju, Korea, pp. 29–32.

Huckvale, M. (**2007a**). "ACCDIST: An accent similarity metric for accent recognition and diagnosis," in *Lecture Notes in Computer Science:*

*Speaker Classification II*, edited by C. Müller (Springer, Berlin), pp. 258–275.
Huckvale, M. (**2007b**). "Hierarchical clustering of speakers into accents with the ACCDIST metric," in *Proceedings of the International Congress of Phonetic Sciences*, Saarbrücken, Germany, pp. 1821–1824.
Hughes, A., Trudgill, P., and Watt, D. (**2012**). *English Accents and Dialects*, 5th ed. (Hodder Education, London).
Kerswill, P., Cheshire, J., Fox, S., and Torgersen, E. (**2010**). "Multicultural London English: The emergence, acquisition and diffusion of a new variety," ESRC Research Project No. RES-062-23-0814.
Kinnunen, T., and Li, H. (**2010**). "An overview of text-independent speaker recognition: From features to supervectors," Speech Commun. **52**, 12–40.
Kirkham, S. (**2017**). "Ethnicity and phonetic variation in Sheffield English liquids," J. Int. Phon. Assoc. **47**(1), 17–35.
Kirkham, S., and Wormald, J. (**2015**). "Acoustic and articulatory variation in British Asian English liquids," in *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, UK, Paper No. 640.
Labov, W., Sharon, A., and Boberg, C. (**2006**). *The Atlas of North American English* (Walter de Gruyter, Berlin).
McDougall, K., and Nolan, F. (**2007**). "Discrimination of speakers using the formant dynamics of /uː/ in British English," in *Proceedings of the International Congress of Phonetic Sciences*, Saarbrücken, Germany, pp. 1825–1828.
Najafian, M., DeMarco, A., Cox, S., and Russell, M. (**2014**). "Unsupervised model selection for recognition of regional accented speech," in *Proceedings of Interspeech*, Singapore, pp. 2967–2971.
Najafian, M., Safavi, S., Weber, P., and Russell, M. (**2016**). "Identification of British English regional accents using fusion of i-vector and multi-accent phonotactic systems," in *Proceedings of Odyssey: The Speaker and Language Recognition Workshop*, Bilbao, Spain, Paper No. 44.
Nance, C., McLeod, W., O'Rourke, B., and Dunmore, S. (**2016**). "Identity, accent aim, and motivation in second language users: New Scottish Gaelic speakers' use of phonetic variation," J. Socioling. **20**, 164–191.
Nerbonne, J. (**2009**). "Data-driven dialectology," Lang. Ling. Compass. **3**, 175–198.
Palo, P., Schaeffler, S., and Scobbie, J. M. (**2015**). "Effect of phonetic onset on acoustic and articulatory speech reaction times studied with tongue ultrasound," in *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, UK, Paper No. 840.
Podesva, R. J. (**2007**). "Phonation type as a stylistic variable: The use of falsetto in constructing a persona," J. Socioling. **11**, 478–504.
Rosenfelder, I., Fruehwald, J., Evanini, K., and Yuan, J. (**2011**). FAVE (Forced Alignment and Vowel Extraction) Program Suite, http://fave.ling.upenn.edu (Last viewed June 29, 2016).
Spinu, L., and Lilley, J. (**2016**). "A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives," J. Phon. **57**, 40–58.
Strycharczuk, P., and Scobbie, J. M. (**2015**). "Velocity measures in ultrasound data. Gestural timing of post-vocalic /l/ in English," in *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, UK, Paper No. 309.
Stuart-Smith, J., Timmins, C., and Alam, F. (**2011**). "Hybridity and ethnic accents: A sociophonetic analysis of 'Glaswasian,' " in *Language Variation—European Perspectives 3: Selected Papers from ICLaVE 5*, edited by F. Gregerson, J. Parrott, and P. Quist (John Benjamins, Copenhagen), pp. 43–57.
Vapnik, V. (**1998**). *Statistical Learning Theory* (Wiley, New York).
Watt, D., Llamas, C., and Johnson, D. E. (**2014**). "Sociolinguistic variation on the Scottish-English border," in *Sociolinguistics in Scotland*, edited by R. Lawson (Palgrave Macmillian, London).
Wells, J. C. (**1982a**). *Accents of English 1: An Introduction* (Cambridge University Press, Cambridge), Vol. 1.
Wells, J. C. (**1982b**). *Accents of English 2: The British Isles* (Cambridge University Press, Cambridge), Vol. 2.
Wieling, M., Margaretha, E., and Nerbonne, J. (**2012**). "Inducing a measure of phonetic similarity from pronunciation variation," J. Phon. **40**, 307–314.
Wieling, M., Nerbonne, J., and Baayen, H. (**2011**). "Quantitative social dialectology: Explaining linguistic variation geographically and socially," PLoS One **6**(9), e23613.
Wormald, J. (**2015**). "Dynamic variation in 'Panjabi-English': Analysis of F1 & F2 trajectories for FACE /eɪ/ and GOAT /əʊ/," in *Proceedings of the International Congress of Phonetic Sciences*, Glasgow, UK, Paper No. 809.
Wormald, J. (**2016**). "Regional variation in Panjabi-English," Ph.D. thesis, University of York, York, UK.
Wu, T., Duchateau, J., Martens, J.-P., and Compernolle, D. (**2010**). "Feature subset selection for improved native accent identification," Speech Commun. **52**, 83–98.
Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. (**2009**). *The HTK Book (for HTK version 3.4)* (Cambridge University Engineering Department, Cambridge, UK).