PHONETIC VARIATION AS A CUE TO REGIONAL IDENTITY: AN EXPERIMENTAL APPROACH

DANIEL LAWRENCE

The University of Edinburgh

Abstract

This paper reports on a web-based experiment which investigated British listeners' perceptual sensitivity to phonetic variation in the vowels of BATH, STRUT, FACE, and GOAT. While these vowels are widely reported to be 'shibboleths' of northern/southern regional origin in Britain, perceptual evidence for such a claim has been notably lacking. The present study aimed to address this gap, measuring the extent to which a diverse sample of British listeners could use variation in the target vowels as a perceptual cue to the regional origin of a speaker. A heterogeneous sample of participants from across Britain took part in a web-based regional identification task. Listeners were asked to place recordings of speakers reading isolated words on a clickable map. Among the stimuli were tokens of words containing the target variables, read by four speakers from South Yorkshire. These tokens were digitally manipulated to create two guises: one using vowel variants typical of northern varieties of British English, and one using variants typical of southern varieties. Listeners were more likely to place northern vowel variants in a northern location, with southern variants more likely to be placed in southern locations, confirming the prediction that these vowels function as cues to regional origin among speakers of British English. A comparison of the four vowels using regression analyses and classification methods allows an exploration of quantitative and qualitative differences in the way these variable forms can be used by listeners as socio-perceptual cues.

1. Introduction

Patterned variation in speech has been linked to a range of social categories, including regional background, gender, age, ethnicity, and the membership of subgroups (e.g. Labov, 2001; Eckert, 2000; Milroy, 1980; Trudgill, 1974). While the majority of evidence for this comes from the study of group-level variation in speech production, the often tacit claim made in such work is that phonetic variation serves a social function – group-level differences in production are equated with indexing *social meaning*, available to listeners as a perceptual cue to the identity of a speaker, and available to speakers as a way to express different aspects of their identity (Eckert, 2000; Campbell-Kibler, 2007). Despite this claim, relatively few studies have directly tested listeners' ability to make social judgments based on phonetic variation. Although the ubiquitous evidence of socially-patterned variation suggests that such an ability exists, developing a complete understanding of the interface between language and society requires researchers to grapple with more explicitly cognitively-oriented questions regarding sociolinguistic competence. Rather than simply asking "How does language vary in its social context?", the relevant question becomes "How much do language users know about socially-structured variation, and to what extent can this knowledge impact on their language production?". These issues cannot be addressed by the analysis of production alone – rather, they require evidence from perception.

Proceedings of the second Postgraduate and Academic Researchers in Linguistics at York (**PARLAY 2014**) conference
 ISSN 1758-0315 Issue 2

A small number of studies have investigated listeners' ability to use phonetic detail as a cue to the social background of the speaker, mainly in a US context. Fridland, Barlett and Kruez (2004) conducted an experiment in which listeners heard a series of words containing vowels implicated in the Southern Vowel Shift in Memphis. Listeners heard pairs of words including the variable vowels which had been digitally-manipulated to represent southern-shifted or non southern-shifted variants. The participants were asked to identify which token was the more southern-sounding. The results suggest that the SVS-participating vowels vary in the extent to which they mark regional identity for Memphis speakers – those changes which are more widely adopted by Memphis speakers than southern US speakers in general appear to be the most perceptually relevant. Plichta and Preston (2005) investigated US listeners' perceptual sensitivity to the monophthongisation of the PRICE vowel using an online experiment. Listeners were presented with words containing vowels taken from a synthetic continuum representing seven degrees of monophthongisation, and were asked to identify the regional origin of the speaker on a clickable map. The results indicate that listeners were sensitive to individual continuum steps, placing more diphthongal realisations in more northern locations. Interestingly, listeners' comments on the task suggest that they were unaware of their ability to discriminate the steps of the continuum – while they may be aware of PRICE monophthongisation as a northern-southern stereotype, their ability to distinguish the variants in a gradient manner appears to be mainly unconscious. Focussing on Glaswegian English, Macfarlane and Stuart Smith (2012) demonstrated that Glaswegian listeners can distinguish phonetic detail in a range of variable features, including final syllable length in *-er* compounds, the NURSE vowel, and syllable-final /l/. The researchers adopted an innovative elicitation technique where listeners were asked to identify the speaker's identity in relation to a set of locally socially-meaningful brand logos – one set representing a (middle class) 'Glasgow Uni' speaker, and one representing a (working class) 'Glaswegian' speaker. As well as demonstrating that listeners can assign social judgments to the variable forms under study, a comparison of responses across the different variables allowed the researchers to estimate the strength of each form as a social-indexical cue.

The present study extends the perceptual work discussed above, focusing on four variable features of British English – the vowels in BATH, STRUT, FACE and GOAT (used here in reference to Wells' (1982) lexical sets). These vowels are often mentioned in descriptions of differences between northern and southern varieties, notably in Wells (1982) and Chambers and Trudgill (1998). The relevant differences across northern and southern dialects with regard to these vowels are listed below (this is by no means an exhaustive list of the ways in which these vowels can vary in British English):

- The vowel in the GOAT lexical set is a usually a diphthong [əʊ] in southern varieties, but may be realised as a monophthong [oː] in northern varieties.
- The vowel in the FACE lexical set is a usually a diphthong [eɪ] in southern varieties, but may be realised as a monophthong [eː] in northern varieties.
- Southern speakers may realise the vowel in BATH as a low back vowel [ɑː], and the vowel in TRAP as higher and further forward [æ]; speakers of northern varieties, however, often realise both BATH and TRAP as [a], with [ɑː] used only in the PALM and START lexical sets.
- Speakers of southern varieties contrast the vowel in FOOT /ʊ/ with the that in STRUT /ʌ/. Northern speakers, on the other hand, may merge /ʊ/ and /ʌ/ completely, with speakers

realising both as [ʊ], or as an intermediate [ə] - like variant. Wells (1982) reports style-shifting on an [ʊ] - [ə] - [ʌ] continuum for many northern speakers.

These vowels are often described as 'shibboleths' of northern-southern regional identity in Britain. The STRUT and BATH vowels feature widely in linguistic stereotypes of northern and southern speakers in England, and were even the subject of a well-known poem by Leeds-born Tony Harrison, which includes the following passage:

'We say [ʌz] not [uz], T.W.!' That shut my trap.
I doffed my flat a's (as in 'flat cap')
my mouth all stuffed with glottals, great
lumps to hawk up and spit out... E-nun-ci-ate!

Harrison (1978)

Here Harrison, from a working-class family but privately educated, documents his experience of speech stigmatisation from his teacher. The mention of 'flat a's' is a reference to the northern realisations of the BATH vowels discussed above. While variation in FACE and GOAT is perhaps not part of such a well-known folk stereotype as variation in BATH and STRUT, monophthongal variants of these vowels are common in northern varieties of British English, and are widely reported as being connected to northern regional identity (Watt & Milroy, 1999; Haddican et al., 2013). In all cases, these vowels are claimed index regional identity; however, this claim is based either a) on production data; b) metalinguistic discourse from interviews with speakers (e.g. Haddican et al., 2013; Watt, 2002) or evidence from folk-linguistic discourse, as in the above poem. As yet, there exists no direct empirical evidence that British English speakers associate these vowels with the social categories 'northern' and 'southern', nor evidence that listeners can draw on these variables as a cue to regional origin. Thus, the aim of the present study was to test the extent to which these vowels serve as indices of northern/southern regional origin among speakers of British English.

2. Methodology

2.1. Creation of stimuli

Four speakers were recruited from Sheffield, Northern England. They were all between the ages of 20 and 30, and could be described as 'adoptive' speakers of a standard variety. All were capable of producing northern-like and southern-like variants of the target variables. Speakers were recorded reading a list of eight words in the frame "The next word is _" on a Zoom H2n portable digital recorder. The words are summarised in Table 1:

FACE	**GOAT**	**BATH**	**STRUT**
waste	*snow*	*pass*	*bus*
late	*gross*	*disaster*	*butter*

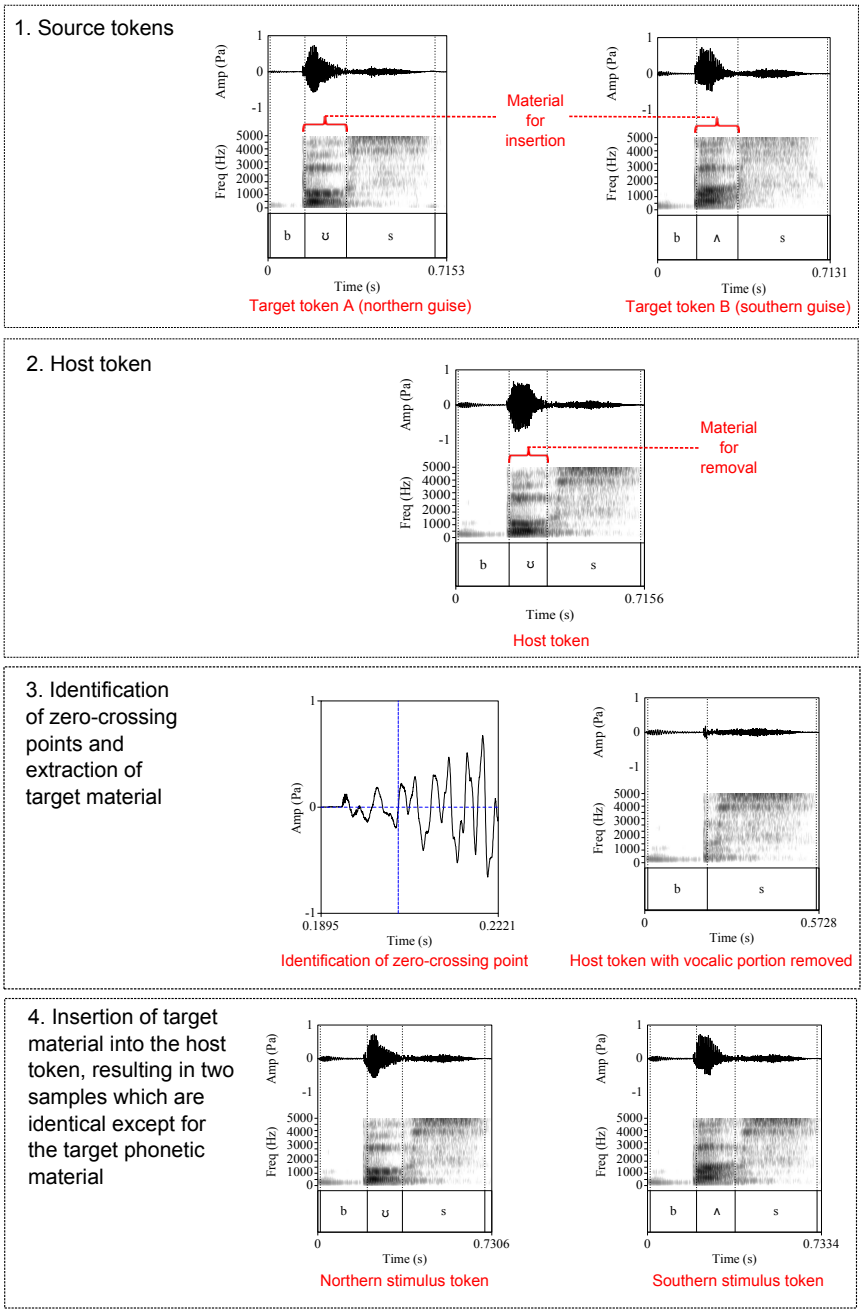
Table 1: Lexical items used as stimuli

Participants were asked to produce these words in northern and southern guises, following an example given by the researcher. All tokens were judged by the researcher to be appropriately representative of the variants of interest.

In order to make claims about the role of an individual phonetic variant in participants' judgments of regional identity, it was desirable to control for any other features which might vary between northern and southern guises. These might include the realisations of the other segments in each token, as well as the pitch contour of each guise. To attempt to control for a possible effect of such differences, the tokens were digitally manipulated in *Praat* (Boersma & Weenink, 2009), following the cross-splicing techniques described by Campbell-Kibler (2007). A secondary purpose in adopting this approach was to evaluate the viability of cross-splicing for potential future studies of the social perception of vowel variation.

For each speaker and item, three tokens were extracted from the word list recording – two instances of the word read in a northern guise, and one in the southern guise. One of the two northern tokens served as the 'host', into which the vowel portion of each of the target northern and southern guises was to be inserted. The splicing process was semi-automated using a script which iterated through the set of recordings in sets of three – two source tokens (Figure 1, point 1) and a host token (Figure 1, point 2). For each token, the script prompted the researcher to identify the material to be deleted or inserted. The researcher manually identified the material to be removed from the host token, at which point the script would delete this material and prompt the researcher to select the desired material from the target token. This was inserted into a copy of the host token at the exact point of deletion (Figure 1, point 4). At every stage of insertion or deletion, the script automatically identified the nearest zero-crossing point of the waveform (Figure 1, point 3), reducing the risk of unnatural clicks or pops caused by the manipulation process. Pitch contours were matched across the stimuli pairs by estimating the pitch pattern of the 'host' token and modifying the stimuli pairs' fundamental frequency using the Praat 'Manipulation' tool. All tokens were mixed down to mono, and the mean intensity of all recordings was scaled to 70dB-SPL.

Figure 1: Creation of minimally-contrastive speech stimuli by cross-splicing natural speech samples



When choosing exactly which phonetic content to select for cross-splicing, it was desirable to replace any material which might provide a cue to vowel quality, whilst modifying the host token as little as possible. The strategy used to achieve this varied depending on the phonetic environment of the vowel:

- *pass*: All material from directly after the burst of [p] up until the beginning of high-frequency turbulence of [s] was replaced.

- *bus*: The steady-state vowel portion was replaced, including the decay of the vowel up until the onset of high-frequency turbulence from the fricative.
- *gross*: It was necessary to replace all material from the release of the velar stop up until the onset of [s] turbulence, as formant structure in the [ɹ] made separating it from the following vowel sound highly unnatural.
- *waste*: It was necessary to include [w] in the segment to be spliced due to differences in the trajectory of the glide between the recordings, as well as the difficulty of segmenting the glide from the diphthong.
- *snow*: The entire vowel portion was replaced, identified as distinct from the nasal by higher-intensity formant structure.
- *disaster*: The entire vowel portion was replaced, from the beginning of visible high-intensity formant structure following the voiced fricative, up until the onset of high-frequency turbulence from the following fricative.
- *butter*: The steady-state vowel portion was replaced, up until the high-frequency turbulence from the release of the alveolar stop.
- *late*: The entire vowel portion was replaced, identified as distinct from the liquid by its high-intensity periodicity. All material up until the high-frequency turbulence from the release of /t/ was included.

In addition to this set of 64 stimuli, 16 filler items were created using recordings from the *Speech Accent Archive* (Weinberger & Kunath, 2011). The aim of including these was to draw listeners' attention away from the fact that the speakers re-occurred, and to add variety in terms of different voices and lexical items. The filler samples were chosen to represent a range of regional varieties, genders, ages, and lexical items, in an attempt to add to the apparent validity of the task. Information about these is provided in Table 2:

SAR Recording Title	Gender; Region; Age	Word	Transcription
english2	F; Birmingham, England; 30	*store*	[stɔː]
		frog	[fɹɒg]
english413	F; Manchester, England; 20	*kids*	[kɪdz]
		frog	[fɹɒg]
english488	F; Irvine, Scotland; 35	*store*	[stɔˠ]
		snake	[sneɪk]
english460	F; Rutland, England; 19	*store*	[stɔː]
		bags	[bagz]
english13	M; Oxfordshire, England; 69	*store*	[stɔə]
		cheese	[tʃiːz]
english57	M; Birmingham, England; 34	*store*	[stɔː]
		cheese	[tʃɨiz]
english141	M; York, England; 19	*kids*	[kɪdz]
		station	[steːʃən]
english319	M; Grimsby, England; 29	*peas*	[pʰːiz]
		bags	[bagz]

Table 2: Filler items

The complete set of 64 target items was split across two conditions, such that each listener heard each speaker reading one of the words for each lexical set in both northern and southern guises. The stimuli were divided in this way to reduce the length of the experiment, in an attempt to avoid listeners' becoming bored or frustrated and exiting early – an important consideration for an online task. The full set of fillers were added to the 32 target items in each group, meaning that a total of 48 responses were collected from each listener. The stimuli were pseudo-randomised in such a way that no two instances of the same speaker or variable would be heard in succession.

2.2. Experimental task

Listeners were presented with the digitally-manipulated speech tokens and were asked to guess the origin of the speaker by placing a pin on a clickable map. Previous work adopting a similar approach has used a range of methods to collect social judgments from listeners. These include comparative questions (e.g. *Which speaker sounds more southern?*) (Fridland et al., 2004) and rating scales (Fridland et al., 2005), as well as check-boxes corresponding to categorical features (Campbell-Kibler, 2007). In the present work, an issue which arose at the early stages of experimental design was the 'slipperiness' of 'North' and 'South' as regional identifiers in Britain. Despite widespread acknowledgment of the 'North-South divide' in public discourse, the actual locations or dialects which are considered to be northern or southern are a matter of debate among speakers (see Wales, 2000). It was thought that including a binary question (e.g. 'Do you think this speaker is northern or southern?') might seem overly simplistic to participants, and would prompt requests for a definition of northern and southern in terms of specific places. On the other hand, providing a list of individual locations meant assuming *a priori* which locations would be relevant choices for listeners. The clickable map, based on Plichta and Preston (2005), was an attempt to solve this problem. The benefit of using a clickable map was that it potentially allowed listeners to identify locations which had not been

pre-specified by the researcher; on the other hand, if a listener could only identify the speaker as northern or southern, but was unable to place a sample in a specific location, they would still be able to place the sample in a broad geographical area. Labels were provided for 14 major cities. Participants were informed that these labels were there as a guide, but that they could click anywhere on the map if they were unsure, or if they felt that none of the labeled locations matched the origin of the speaker. A further benefit to the map interface was that it provided participants with a visually engaging and challenging task, resulting in a high rate of completion.

The experiment was delivered on a custom-built web-based interface which was designed and programmed by the researcher. The programming language used was ActionScript 3.0, based on the guidelines provided in Weinstein (2012). This programming language, based in the Adobe *Flash* platform, is widely used in online behavioural experiments due to its wide cross-platform compatibility (see e.g. Reimers & Stewart, 2007; Simcox & Fiez, 2014). The experiment was hosted on a University of Edinburgh web server, and responses were stored in a private text file on this server. The experiment was promoted through social media by friends and colleagues of the researcher, adopting a convenience sampling approach which allowed a relatively large number of responses to be collected in a short period of time.

On starting the experiment, participants were first presented with a consent form which explained the purpose and nature of the experiment, and required participants to provide basic personal information and confirm their consent to participate in order to continue. They were then prompted to test their sound settings by clicking a button which played a short sound sample, allowing the participant to adjust the volume of the experiment as appropriate.

The experimental task proceeded as follows: participants were presented with a map of Britain, and were prompted to click on a button to listen to the next speaker. After listening, participants were invited to click anywhere on the map to guess the regional origin of the speaker, placing a pin on the map which they were free to move until they chose to continue. They were explicitly told that they did not have to click only on the marked locations. Listeners were allowed to hear each recording as many times as they desired, and were asked to add any comments they had regarding social characteristics of the speaker after choosing a location.

Figure 2 provides an impression of the experiment.

Figure 2: Experimental interface

The experiment ran in a window of fixed size and resolution, reducing cross-participant variation which might occur due to different screen sizes or zoom levels. The responses recorded included the pixel coordinates of each regional placement, any comments about the social identity of the speaker, and the number of times the participant listened to the recording. Detailed background information was collected from each participant, including their age, gender, regional origin, socioeconomic status, and ethnicity. Finally, participants were asked to provide any general comments they had regarding the task, with a view to collecting feedback for future developments of the experiment.

3. Results

3.1. Visualising the data

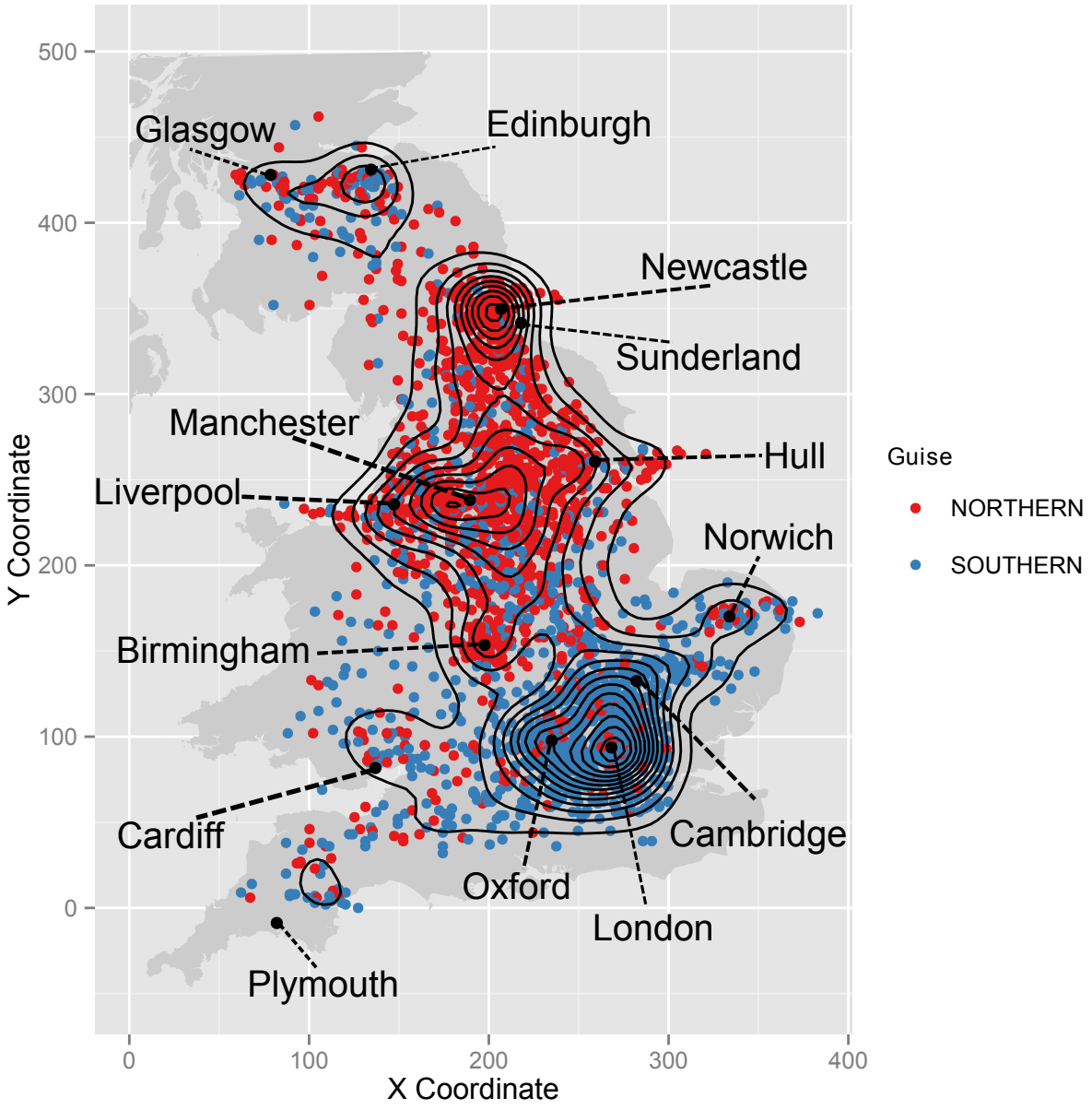
The experiment was published in March 2014 and ran for two months. A total of 86 listeners took part, resulting in a complete dataset of 2752 responses to the target items. This dataset was imported into the statistical package R (R Core Team, 2014) for analysis. Respondents' biographical data were carefully checked, and any respondents who listed their regional origin as somewhere outside of the British Isles were excluded. In addition, any respondents who reported that they had a background in linguistics were not included in the following analyses. 6 participants (192 responses) were excluded in this manner. Participants reported their regional origin by entering it in a text field as part of an exit questionnaire. The table below summarises the responses, using categories adapted from the regions listed on the UK Office of National Statistics website (http://www.ons.gov.uk/).

Region	Number of listeners
East Anglia	3
London	14
South East	12
North West	7
East Midlands	2
West Midlands	3
North East	2
Scotland	16
Wales	3
Yorkshire & The Humber	16
South West	2

It is evident that the sample is biased towards Scotland, Yorkshire and The Humber, and London and the South East, reflecting the convenience sampling approach adopted. The analyses presented in this paper group together all responses, on the basis that the variables under study are generally claimed to be recognised supra-regionally as indexical of northern/southern regional identity in Britain; thus, it would be expected that a patterned response to the vowel variants would be observable regardless of the regional origin of the listener. Due to time and space constraints, an investigation of the role of listener regional origin is left for a future analysis.

The coordinates of each click were plotted on a 400x500 grid, superimposed on a map of the British Isles. Figure 3 visualises the regional placements of the digitally-manipulated recordings; northern variants are displayed in red and southern variants are displayed in blue.[1]

Figure 3: Regional placements separated by guise



It is evident from the above figure that participants responded very differently to tokens from each guise, with northern variants tending to be placed further north on the map, and southern variants further south. The contour lines indicate the areas which received the highest density of clicks, estimated using two-dimensional kernel density estimation – the closer the contours, the higher the density of placements in that area.

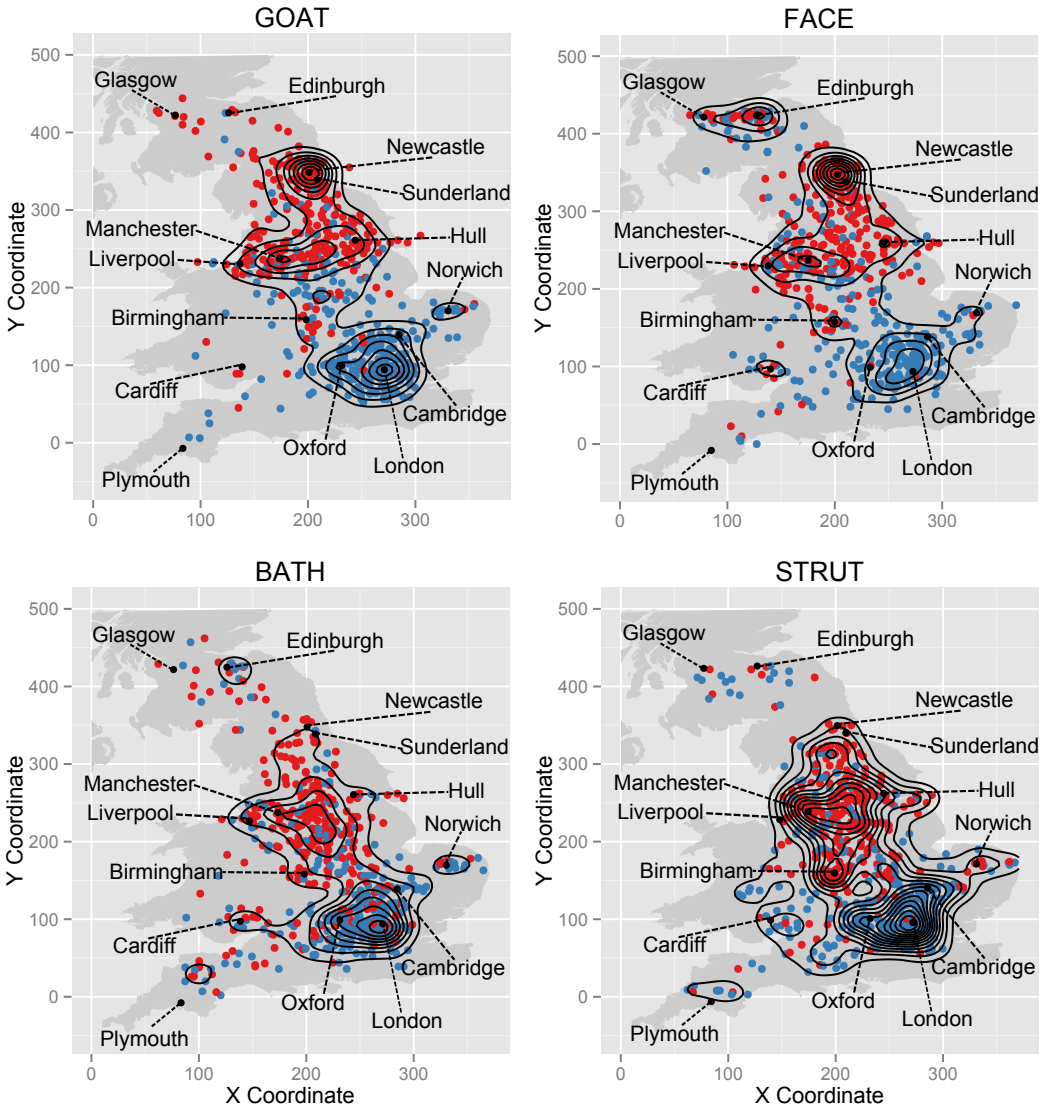
Responses appear to cluster strongly in at least four areas, roughly corresponding to London

[1]Note that although the labeled points marked in Figures 3-6 have the same coordinates as the points provided on the clickable map, in some cases participants seem to have clicked on the label text itself rather than the points – this may explain the occasional slight misalignment of clusters of responses and the labeled points.

and the South East, a central-northern area corresponding to the Midlands, Yorkshire and Lancashire, a north-eastern area corresponding to Tyneside, and Scotland. The colour-coding of guises clearly shows that a larger proportion of northern variants were placed around northern locations and that the converse is true for southern variants. It is interesting to note, however, that respondents' judgments are far from categorical; despite the overall tendency of northern and southern clustering, there is a still a considerable 'error rate' for each guise, with northern variants being placed in southern locations and vice-versa. Rather than determining a categorical judgment, the variable realisations of the vowels shift the probability that a speaker will be identified as coming from a particular region.

Considering the patterns for each variable, it becomes clear that the four vowel variables differ in terms of the regional placements their variants are likely to cue, with the largest differences between the northern variants. This is shown in Figure 4:

Figure 4: Regional placements separated by variable



Northern variants of FACE and GOAT seem to behave differently to the northern variants of STRUT and BATH – the latter generally cluster around central-northern locations, where the former cluster strongly in the Tyneside area. There also appears to be a strong cluster of both

northern and southern FACE variants in Scotland.

3.2. Statistical analysis

To identify statistically-significant differences between the placement patterns of the speech samples, the distributions of each vowel variant were compared in all relevant combinations and tested for independence – first comparing the distributions of variants *within* variables (e.g. [ʊ] vs [ʌ], [eɪ] vs [eː]), then comparing distributions *between* variables (e.g. [eɪ] vs [əʊ]). The statistical test used was Peacock's (1983) two-dimensional extension of the Kolmogorov-Smirnov test. This test estimates the largest difference between the cumulative probability distributions of two sets of points, and estimates the probability of a difference of a given size under a null hypothesis that both sets of points come from the same distribution. Before statistical analyses were conducted, a small number of responses which were placed outside of the map area were excluded (n=63).

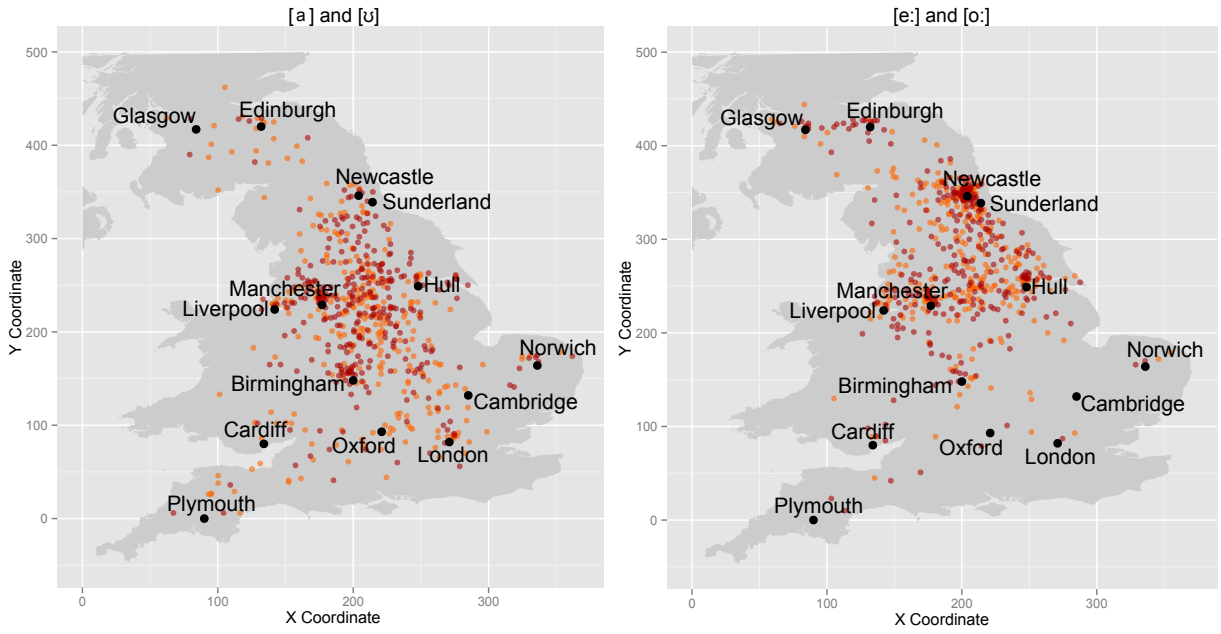
	Variants		K-S	p - value
Within variables, between guises	[ɑː]	[a]	0.4992	2.7313e-29 ***
	[ʌ]	[ʊ]	0.5914	2.4881e-42 ***
	[eɪ]	[eː]	0.6128	2.4850e-45 ***
	[əʊ]	[oː]	0.7333	2.5155e-66 ***
Between variables, within southern guises	[ʌ]	[eɪ]	0.1700	0.0099
	[ʌ]	[əʊ]	0.1138	0.20
	[ɑː]	[eɪ]	0.1027	0.20
	[ɑː]	[əʊ]	0.1410	0.0821
	[ʌ]	[ɑː]	0.1027	0.20
	[eɪ]	[əʊ]	0.2083	2.9078e-04 ***
Between variables, within northern guises	[ʊ]	[eː]	0.3683	2.2901e-15 ***
	[ʊ]	[oː]	0.3429	3.8012e-13 ***
	[a]	[eː]	0.3944	8.9890e-18 ***
	[a]	[oː]	0.4364	3.9175e-22 ***
	[ʊ]	[a]	0.1587	0.0223
	[eː]	[oː]	0.0857	0.20

Table 3: Kolmogorov-Smirnov tests within and between variables

The first four rows of Table 3 demonstrate the highly-significant effect of the vowel manipulation on each variable. When listeners heard words from the BATH, STRUT, FACE and GOAT lexical sets, variants typical of northern varieties cued a placement in a more northern location than their southern counterparts. The next five rows show that there are no significant differences between placements of the majority of southern variants, assuming a Boneferroni-corrected threshold p-value of 0.003. A highly significant difference was found between placements of the southern variants of FACE and GOAT, suggesting that while [əʊ] and [eɪ] cue different regional placements from their northern counterparts, participants interpret them as carrying a different regional meaning. Returning to Table 4, this effect seems to be carried by a higher frequency of [eɪ] placements in Scotland, which will be explored further below. The final six rows of table 3 show that northern variants of BATH and STRUT seem to pattern differently

from northern variants of FACE and GOAT. A comparison of [ʊ] and [a] returns no significant difference, nor does a comparison of [eː] and [oː]; however, comparing northern STRUT and BATH placements to northern FACE and GOAT placements returns highly significant effects in all cases. This exploratory analysis verifies the intuition captured in Figure 4, suggesting that while northern variants of all four variables cue a placement in a general northern location, northern variants of FACE and GOAT are stronger indices of locations in the North-East. Figure 5 highlights this difference:

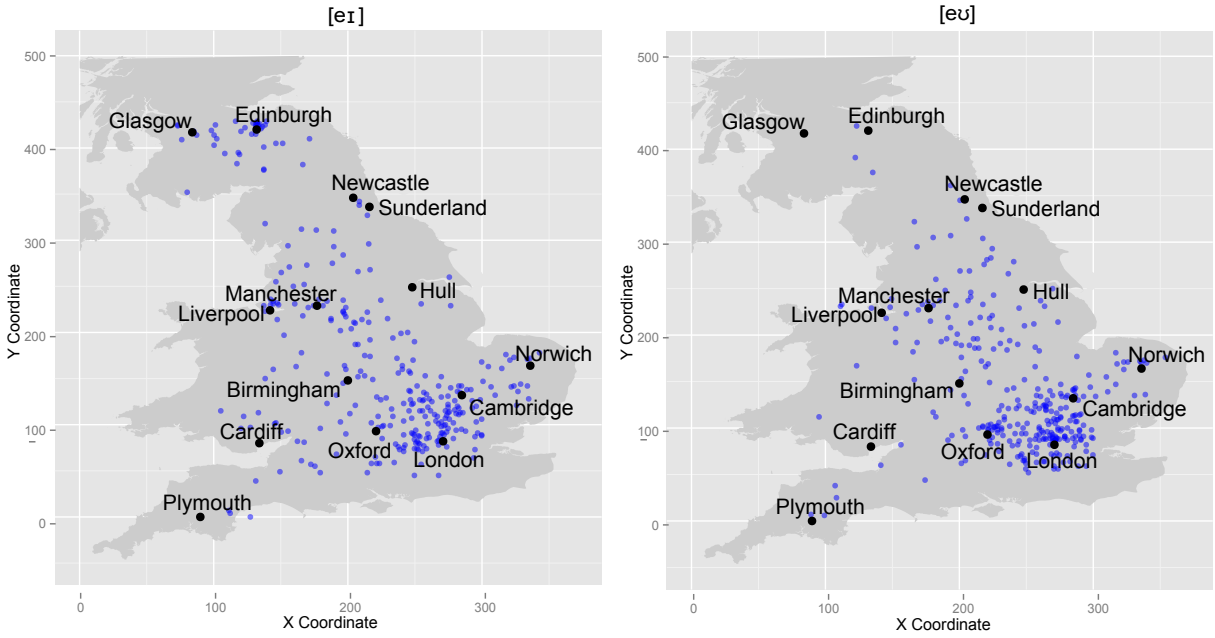
Figure 5: Left: Placements of northern BATH (orange) and STRUT (red); Right: Placements of northern GOAT (orange) and FACE (red)



As well as appearing to be stronger predictors of a placement in the North-East, [eː] or [oː] seem to reduce the probability that a speaker will be placed in a southern location – participants seem to have an intuition that there is some possibility that a speaker who uses [a] in BATH or [ʊ] in STRUT might be from a southern location, but that this is much less likely if the speaker uses [eː] in FACE and [oː] in GOAT.

In addition to the difference in the placement of northern variants discussed above, the tests of independence in Table 3 revealed a highly significant difference in the placements of southern FACE and GOAT variants [eɪ] and [əʊ]. These are compared in Figure 6:

Figure 6: Comparison of southern FACE and GOAT placements



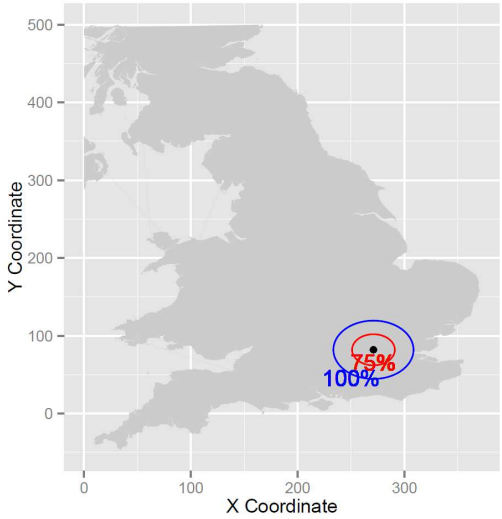
The significant difference between [əʊ] and [eɪ] placements appears to be carried by a tendency for [eɪ] tokens to be placed in Scotland (more accurately, Edinburgh) at a slightly higher frequency than [əʊ]. A future analysis might usefully explore whether this tendency is specific to Scottish listeners – it might be hypothesised that this trend of mapping a southern English form to the Scottish capital represents a perception of SSBE influence on Edinburgh English.

This exploratory analysis shows that the vowels in BATH, STRUT, FACE, and GOAT can be used by listeners as cues to regional identity. While this has generally been taken for granted in the sociolinguistic literature, the present analyses represent the first experimental evidence for their perceptual salience as 'shibboleths' of regional origin. An additional finding is that the different vowel variants seem to contribute in different ways to regional placements, rather than simply being associated with broad northern or southern stereotypes. The 'regional meaning' carried by the vowels can be thought of as a probability that the speaker will come from a particular area, given their use of a particular variant. The presence of this variant appears to shift the probability of a placement in a particular location, and the way that this probability shifts appears to differ depending on the variant – for example, the likelihood of a speaker being identified as from the North-East is higher if a listener hears [eː] and [oː] than if they hear [ʊ] or [a]; furthermore, the likelihood of a speaker being identified as southern even though they use typically northern variants is much higher for [ʊ] or [a] in BATH and STRUT than it is for [eː] and [oː] in FACE and GOAT.

So far, these analyses have ignored the fact that a range of factors might influence a regional placement. Firstly, it is likely that the background of the listener and their linguistic experience play a key role in their judgments. Secondly, the stimuli vary in terms of the lexical items used and phonetic contexts in which the target vowels appear, which might effect the salience of the sociolinguistic variant. In addition, the speakers exhibit natural variation in their pronun-

ciation of both the target variants and other segments in the stimuli; they also vary in terms of their F0 and voice quality. Finally, it is possible that individual participants have idiosyncratic preferences for their placements. In order to control for the effect of these factors, it would be desirable to fit a mixed effects model, as is standard practice in sociolinguistic work. Typical production studies deal with one-dimensional continuous variables (e.g. formant values) or categorical variables (e.g. the realisation of /t/ as [t] or [ʔ]), and attempt to model the linguistic and non-linguistic factors which influence individual production events. However, the independent variable in the present data consists of a set of two-dimensional points. The feature of interest is the extent to which the location of each point can be predicted by aspects of the perceptual events presented in the experiment – either the characteristics of the stimuli or the background of the listener. One notable observation is that listeners' placements of the speech samples generally cluster around regional labels. As Figure 7 illustrates, all points lie within 75 pixels of a regional label, with half of them within 25 pixels of the nearest label.

Figure 7: Distance of placements from nearest regional label



The fact that responses cluster so strongly around regional labels suggests that the labels played a large role in participants' placements. Taking advantage of this, the data were grouped into 13 categories using a simple clustering method, with Newcastle and Sunderland treated as a single category due to their close proximity. Treating the location of each regional label as the centroid of a cluster, each point was assigned to the cluster whose centroid whose coordinates had the shortest Euclidean distance from the coordinates of that point.

Having split the data into 13 categories, a set of logistic regression models were fit for each location using *Rbrul* (Johnson, 2009). This program performs stepwise model selection – given a set of possible predictors, the program generates a set of regression models, selecting the model which accounts for the most variation in the dependent variable with fewest independent variables, and returning the best predictors for a given dataset and their coefficients. The dependent variable in each model was a factor representing the selection of the specified location vs the selection of any other location. Independent variables were included for *Vowel* (referring to the vowel variable that listeners heard on a given trial: FACE, GOAT, BATH or STRUT), *Guise* (whether the vowel was a northern or southern variant), *Variant* (presenting the interaction between Guise and Vowel – which of eight possible variants did listeners hear?) *Speaker* (which

speaker did listener hear?) and *Word* (which of the two possible words did listeners hear?), with random effects for *Item* (the specific recording heard) and *Participant* (the individual listener). Interactions between *Variant*, *Speaker* and *Word* were also checked. A key factor not explored in the present analysis is the role of the listeners' regional origin, which is excluded from the present paper due to space constraints.

Table 4 summarizes the predictors selected for each cluster, representing the factors which influenced participants' selection of each location.

Cluster	**N**	**Best model**
Birmingham	262	Name (Random) + File (Random) + Variant (p = 5.59e-16)
Cambridge	259	Name (Random) + File (Random) + Variant (p = 4.21e-16)
Cardiff	94	Name (Random) + File (Random) + Variant (p = 0.00277) + Speaker (p = 0.0294)
Edinburgh	87	Name (Random) + File (Random) + Variant (p= 0.0295)
Glasgow	51	Name (Random) + File (Random) + Variant (p = 0.0355)
Hull	221	Name (Random) + File (Random) + Variant (p = 2.8e-10)
Liverpool	153	Name (Random) + File (Random) + Variant (p = 0.0215)
London	311	Name (Random) + File (Random) + Variant (p = 4.32e-16)
Manchester	362	Name (Random) + File (Random) + Variant (p = 7.31e-12)
Newcastle	324	Name (Random) + File (Random) + Variant (p = 3.65e-07) + Speaker (p = 0.0148)
Norwich	95	Name (Random) + File (Random) + Variant (p = 8.88e-06)
Oxford	234	Name (Random) + File (Random) + Variant (p = 2.17e-15)
Plymouth	44	Name (Random) + File (Random) + Variant (p = 0.034)

Table 4: Best predictors of placements in each regional cluster

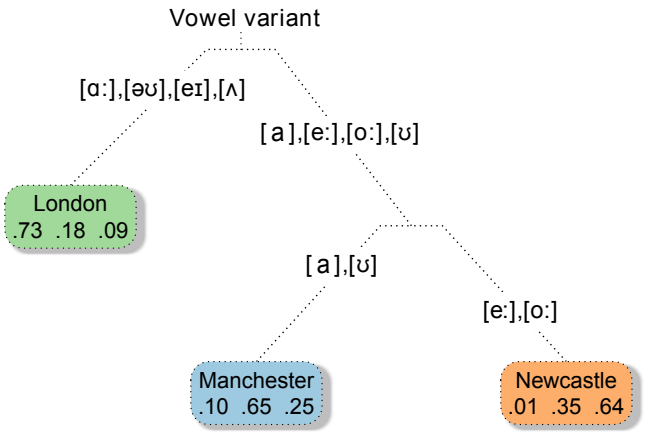
In all cases, the variant (i.e. the individual vowel variant that participants heard) was selected as the best predictor of a regional placement, suggesting that the vowel participants heard influenced the location of the regional placements they made. The northern or southern *Guise* was not selected in any model, suggesting that it is the individual vowel variant that drives placements. This confirms the observation made previously – listeners appear to treat monophthongal FACE and GOAT as indexically very different from the northern variants of BATH and STRUT, and this pattern appears to be robust even when controlling for other factors which might influence listeners' placements. In three cases, Cardiff, Newcastle, and Cambridge, the identity of the speaker appears to have influenced the regional placements. This suggests that listeners were not only drawing on the target vowels, but also used information from other aspects of the recordings. This might be related to the way the speakers realised the other segments in the recorded words, or perhaps some aspect of intonation or voice quality; however, such investigations go beyond the scope of the present paper.

While the extent to which listeners' intuitions represent actual production patterns remains to be seen, the present data provide strong evidence that monophthongal FACE and GOAT variants [eː] and [oː] are strongly associated with North-Eastern locations, whereas [a] and [ʊ] are associated with central-northern locations such as Manchester. These associations are by no means categorical, and patterns can be observed when the rate at which apparently northern variants are placed in southern locations – while a small proportion of [a] and [ʊ] tokens are placed in southern locations, the proportion of southern [eː] and [oː] placements is far lower.

As a final exploratory analysis, an attempt was made to apply the tree-based classification method described in Tagliamonte and Baayen (2012). The aim of this method is the same

as that of logistic regression – the prediction of a particular outcome based on a set of predictors. While regression models accomplish this analytically, classification trees work by splitting the data into partitions on the basis of a set of yes/no questions, then estimating the amount of information captured by a given split. The algorithm works through all specified predictors in the first iteration, then recursively partitions the selected subsets until no further splits are predictive of the data. The resulting *conditional inference tree* expresses the effect of the predictors on the outcome variable in the form of a decision tree. An example is given in Figure 8:

Figure 8: Conditional inference tree generated from a subset of the regional identification data



This tree was produced using the *rpart* package in R (Therneau, Atkinson & Ripley, 2014). Initially, a conditional inference tree was generated with the location of each regional identification as a dependent variable, and *Guise*, *Word*, *Speaker*, *Variable* and *Variant* as predictors. Consistent with the regression models, only *Variant* was selected as a significant predictor. The above tree represents a simplified version created from a subset of the data which only included responses placed in the three largest clusters: London, Manchester or Newcastle. Each branch on the tree represents a split identified as most informative in classifying listeners' placements. The coloured nodes represent the most likely outcome given the factor levels shown on each branch. The numbers express the conditional probability of each class being selected given the predictor levels on that branch. Overall, this classification tree appears to support the claims made previously. When listeners hear a speaker using southern variants [ɑː], [eɪ], [əʊ] or [ʌ], there is a probability of .73 that they will place that speaker in London over Manchester or Newcastle, with a smaller probability that they will choose Manchester (.18), and much smaller probability that they will select Newcastle (.09). The most likely placements of speakers using northern variants [a], [eː], [oː], and [ʊ] are split between Manchester and Newcastle – there is a probability of .65 that [a] or [ʊ] will cue a Manchester placement, and a probability of .25 that a speaker using these vowels will be placed in Newcastle. London placements of speakers using the variants are far less likely (.10). [eː] and [oː] strongly predict a Newcastle placement, with a probability of .64 that a speaker using these sounds will be placed there. The probability that monophthongal FACE and GOAT variants will result in a Manchester placement is lower (.35), and London placements of speakers these sounds are particularly unlikely (.01). Overall, the conditional inference tree analysis confirms the main finding of this study – listeners interpret variation in BATH, STRUT, FACE and GOAT as indexing a speaker's region, with northern BATH and STRUT variants associated with central-northern locations, and northern FACE and GOAT variants associated with locations further north.

4. Discussion

This study aimed to test the claim that variation in the BATH, STRUT, FACE and GOAT vowels can be used by British listeners as cues to the regional origin of a speaker. The status of these vowels as north-south 'shibboleths' is widely referred to in the literature on phonological variation in the UK (Wells, 1982; Chambers & Trudgill, 1998; Watt & Milroy, 1999; Haddican et al., 2013). However, these claims have been based primarily on production/interview data, and the presence of sociolinguistic stereotypes in popular discourse. This experiment presented listeners with a set of recorded words and asked them to identify the regional origin of the speakers. Among the recordings were a set of instances of the target variables which had been digitally manipulated to represent either the northern or southern realisation of the vowels under study.

An analysis of listeners' regional placements confirms that the digital manipulation of the target variants alone was sufficient to cue differences in listeners' intuitions about a speakers' regional origin. The statistical analyses indicate that even when individual-level variation is taken into account, participants had consistent intuitions about the regional information carried in each variant – while southern variants [ɑː], [eɪ], [əʊ] or [ʌ] generally predicted a placement in the south-east of England, northern variants of BATH and STRUT were more likely to be placed in central-northern locations, and monophthongal FACE and GOAT are strongly associated with the North-East. A brief exploratory analysis using recursive classification trees confirmed the patterns identified, demonstrating that each variant contributes to the probability of a placement in a particular location to different degrees.

Overall, the results demonstrate that listeners have intuitions about the regional indexicality of phonological variation which can be captured through perceptual experiments. Presenting listeners with phonetically variable stimuli and asking them to make social judgments allows for a calculation of the probability that a particular social judgment will be assigned given a particular set of acoustic characteristics. This presumably reflects the kind of probabilistic reasoning that speakers do in real life when they interpret the social significance of phonetic variation, with the important difference that real-life sociolinguistic inference presumably involves the integration of a range of socially-meaningful cues (linguistic and non-linguistic) in forming social impressions. Intriguingly, a brief look at listeners' commentary on the task suggests that, although they are explicitly aware that some variants are typical of northern and southern varieties, they do not articulate the fact that monophthongal FACE and GOAT carry a different regional association to northern variants of BATH and STRUT. Typical comments are shown below:

- *The task appeared to be relatively easy in roughly discerning between northern and southern accents. I tried to be more precise about where in the North or South I thought the accents were from but it was mostly guesswork based on my (limited) knowledge of regional dialects.*
- *I didn't feel sure about many - less than half. A single word isn't much to go on. I could pick out things like short and long a sounds. But mostly I just went with a hunch.*
- *Easy to tell whether the speaker was northern or southern but harder to define more precisely.*
- *Some were easier to place than others. Some seemed more generally northern or southern and I couldn't place them by cities.*

While listeners are able to express explicit awareness of some aspects of regional variation, it appears that some of their intuitions are less available to introspection – consistent with the findings of Plichta and Preston (2005). The strength of the modified matched-guise method applied in this study is that it appears to allow researchers to capture listeners implicit awareness of regional (or other) variation, apparently allowing the quantification of the 'hunches' expressed by the participant in the second comment above.

5. Conclusion

The present study adds to a growing body of work exploring the perceptual dimension of sociolinguistic variation. Using a fairly simple experimental design whereby listeners are asked to intuit social information from acoustically-manipulated stimuli, it is possible to obtain information about listeners' social interpretation of linguistic features that would be hard to elicit through direct questioning. However, there are a few issues to be considered if the method is to be employed in future work. Firstly the usefulness of structuring the task around a clickable map is debatable – in the present study, responses tended to be tightly clustered around the regional labels provided, effectively rendering the map superfluous. The need to subsequently convert the data into a set of categories added an unnecessary extra layer of complexity to the analysis, and means that there is no guarantee that responses treated as 'London' or 'Newcastle' in analyses were actually intended as such by listeners. A more useful approach might be to provide listeners with a set of pre-specified locations. Alternatively, a menu-type interface could be constructed with varying levels of granularity – for example, listeners could start by selecting the country, then broad region, then individual city, stopping at the point where they felt they had no further intuitions.

A second issue with this experiment is the extent to which the experimental design might be contributing to the observed patterns. When listeners form intuitions about a speakers' origin in real life, they do so based on a range of possible cues, both linguistic and non-linguistic. The situation presented by the experiment, where a series of regional intuitions based on only a small amount of linguistic material are elicited one after the other, is highly unnatural. It is possible that the observed patterns emerge as an artifact of the task – perhaps listeners notice that BATH, STRUT, FACE and GOAT re-occur in the stimuli and attempt to distinguish them by placing them in arbitrarily different locations. The fact that the observed patterns appear to be consistent across listeners suggests that this may not be the case; however, based on this experiment alone there is no way of knowing if the responses captured reflect the same kind of intuitions that might influence the way listeners perceive and produce variable forms in their day-to-day lives. One avenue presently being explored to address this is testing the experiment among American listeners with limited exposure to British English, who would be expected to posses less consistent intuitions about the regional indexicality of the vowels under study.

A third point to be considered in future developments of this work is the importance of participants' backgrounds and their linguistic experience. While the main patterns in the data appear to hold despite considerable diversity in the regional backgrounds of the listeners, the present analyses are seriously limited by the fact that they overlook the role that this diversity might play. A further area left unexplored in the present work is the question of exactly *why* these patterns of perceptual responses emerge as they do, and the extent to which they represent listeners' exposure regional varieties versus more complex motivations (such as the perception of difference from one's own speech, or (non)-standardness). The next step for this project will

be to address these gaps, exploring systematic variation in listeners' social perceptions of these variable forms, and the relationship between those perceptions and production patterns.

Acknowledgements

Thanks to my supervision team: Dr. Lauren Hall-Lew, Prof. Alice Turk and Dr. Josef Fruehwald for their comments and encouragement. Thanks also to the audience at PARLAY 2014 for their helpful comments, especially Dr. Paul Foulkes, Ania Kubisz and Ella Jeffries. This research was supported by an ESRC studentship attached to grant ES/J500136/1.

References

BOERSMA, P. & WEENINK, D. (2009). Praat: Doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2014.

CAMPBELL-KIBLER, K. (2007). Accent, (ING), and the social logic of listener perceptions. American Speech, 82(1):32–64.

CHAMBERS, J. K. & TRUDGILL, P. (1998). *Dialectology*. Cambridge: Cambridge University Press.

ECKERT, P. (2000). *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Oxford: Blackwell

FERRAGNE, E. & PELLEGRINO, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, 40(01):1–34.

FRIDLAND, V., BARTLETT, K., & KREUZ, R. (2004). Do you hear what I hear? Experimental measurement of the perceptual salience of acoustically manipulated vowel variants by Southern speakers in Memphis, TN. *Language Variation and Change*, 16(01):1–16.

FRIDLAND, V., BARTLETT, K., & KREUZ, R. (2005). Making sense of variation: Pleasantness and education ratings of southern vowel variants. *American Speech*, 80(4):366–387.

HADDICAN, B., FOULKES, P., HUGHES, V., & RICHARDS, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change*, 25(03):371–403.

HARRISON, T. (1978).Them & [uz]. In HARRISON, T., ed. *From the School of Eloquence and Other Poems*. London: Rex Collings.

HARTIGAN, J. A. & WONG, M. A. (1979). Algorithm AS136: A K-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108.

JOHNSON, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, 3(1):359–383.

LABOV, W. (2001). *Principles of Linguistic Change, Vol. 2: Social Factors*. Malden, MA: Blackwell Publishers

MACFARLANE, A. E. & STUART-SMITH, J. (2012).'One of them sounds sort of Glasgow unish': Social judgements and fine phonetic variation in Glasgow. *Lingua*, 122(7):764–778.

MILROY, L. (1980). *Language and Social Networks*. Oxford: Blackwell.

PEACOCK, J. (1983). Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society*, 202(3):615–627.

PLICHTA, B. & PRESTON, D. R. (2005). The/ay/s have it: The perception of /ay/ as a north-

south stereotype in United States English. *Acta Linguistica Hafniensia*, 37(1):107–130.

R CORE TEAM. (2014). *R: A Language and Environment for Statistical Computing.* [Computer program]. Retrieved Jan 15, 2014.

REIMERS, S. & STEWART, N. (2007). Adobe flash as a medium for online experimentation: A test of reaction time measurement capabilities. *Behavior Research Methods*, 39(3):365–370.

SIMCOX, T. & FIEZ, J. A. (2014). Collecting response times using Amazon Mechanical Turk and Adobe Flash. *Behavior Research Methods*, 46(1):95–111.

TAGLIAMONTE, S. A. & BAAYEN, R. H. (2012). Models, forests, and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, 24(02):135–178.

THERNEAU, T., ATKINSON, B., & RIPLEY, B. (2014). *Rpart: Recursive Partitioning and Regression Trees.* R package version 4.1-8.

TRUDGILL, P.(1974). *The social differentiation of English in Norwich*. Cambridge: Cambridge University Press.

WALES, K. (2000). North and South: An English linguistic divide? *English Today*, 16(01):4–15.

WATT, D. (2000). Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change*, 12(01):69–101.

WATT, D. (2002). 'I don't speak with a Geordie accent, I speak, like, the northern accent': Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics*, 6(1):44–63.

WATT, D. & MILROY, L. (1999). Patterns of variation and change in three Newcastle vowels: Is this dialect levelling? In FOULKES, P. & DOHERTY, G. (eds.). *Urban Voices: Accent Studies in the British Isles*, pp 25–46. London: Arnold.

WEINBERGER, S. H. & KUNATH, S. A. (2011). The Speech Accent Archive: Towards a typology of English accents. *Language and Computers*, 73(1):265–281.

WEINSTEIN, Y. (2012). *Flash Programming for the Social & Behavioral sciences: A Simple Guide to Sophisticated Online Surveys and Experiments*. SAGE Publications.

WELLS, J. C. (1982). *Accents of English: Volume 1*. Cambridge: Cambridge University Press.

Daniel Lawrence
Linguistics & English Language
School of Philosophy, Psychology and Language Sciences
Edinburgh
EH8 9AD
email: dlawrenc@staffmail.ed.ac.uk