

Variation in the FACE vowel across West Yorkshire: Implications for forensic speaker comparisons

Kate Earnshaw, Erica Gold

Department of Linguistics and Modern Languages, University of Huddersfield, United Kingdom

k.earnshaw|e.gold@hud.ac.uk

Abstract

In the field of sociophonetics, research is largely focused on the documentation of regional variability. However, the majority of literature in the United Kingdom often reports on variation at a macro-level (e.g. Northern, Yorkshire, West Yorkshire) rather than at a more local level (e.g. West Yorkshire: Bradford, Calderdale, Kirklees, Leeds, Wakefield). Traditionally, for sociophoneticians, examining regional variation at a broader level is adequate for answering research questions related to language change or more general variation. For practical applications (e.g. forensic, speech technology), however, more fine-grained regional analysis is necessary. This paper analyses over 2000 FACE tokens from three metropolitan boroughs (Bradford, Kirklees and Wakefield) within West Yorkshire, in order to determine the extent to which F1~F3 vary across the region. Results suggest that for FACE, these three boroughs within West Yorkshire are more regionally stratified than previously acknowledged. These findings are of particular importance to the forensic speech science community, as experts rely on these regional nuances in order to make important judgments related to strength of the speech evidence in a case. Should decisions be made without the greater understanding of local-level variation, the strength of evidence risks being over- or under-estimated.

Index Terms: forensic speech science, sociophonetics, regional variation, vowels, FACE, West Yorkshire

1. Introduction

It is widely acknowledged that the United Kingdom has relatively high levels of regional variability when considering its size (for example, in comparison to the United States). Numerous textbooks provide detailed descriptions of the ways in which accent and dialect vary between geographical regions throughout the UK [1,2,3,4]. However, at present, the extent to which speech production varies on a local-level is not fully understood. While there are of course studies which focus on single speech communities in very small and specific areas, most of these do not tend to consider how the speech patterns of speakers from neighbouring communities might vary. This paper defines “local” to mean a narrower view on the typical regional stratification. For example, considering Westminster, Camden and Hackney as separate entities rather than taking a holistic view of London. This study does just that by taking a detailed look at three relatively small boroughs within West Yorkshire, in order to examine how much linguistic diversity exists in a seemingly homogeneous population. It is hypothesised that the UK is perhaps even more linguistically diverse than is currently recognised, in contrast to the mainstream media’s popular narrative that regional accents are dying out.

The following subsections look at the current state of regional variation in West Yorkshire, while also demonstrating the importance of this research for forensic speech science.

1.1. FACE in West Yorkshire English

West Yorkshire is situated in the North of England and has received relatively little attention from the sociophonetic community in recent years. Most research which has involved speech communities from within the region has also included other areas of Yorkshire and findings have been reported more generally as examples of “Yorkshire English” [1,2,3,4]. However, there are a few studies which have focussed on some areas within West Yorkshire, although these are largely outdated and auditory only analyses [1,5,6]. After consulting these studies, the vowel in words of the FACE lexical set, as defined by [1], became of particular interest as it is hypothesised to have high levels of variation in West Yorkshire English speech.

Existing phonetic literature has reported [ɛɪ] and [e:] as common variants of FACE in Bradford and Kirklees accents [6]. Although, Bradford English has also been said to typically contain an open-mid monophthong [ɛ:] [4]. Short monophthongs have been reported for the words *make* and *take*, with [ɛ] in Bradford and [a] in Kirklees [4,5]. In Wakefield, FACE has been found to be most commonly realised as [e:] [7]. All of the above studies have presented auditory representations of FACE. However, there is no current literature substantiating these representations with acoustic measurements or even considering the traditionally diphthongal nature of this vowel in General British English. This study investigates the present situation in the region and examines whether production of FACE varies across metropolitan boroughs and if so, to what extent.

1.2. Forensic relevance

In addition to documenting regional variation in Bradford, Kirklees and Wakefield, this study has practical implications for forensic speaker comparison (FSC) casework. FSC involves the analysis of two or more speech samples, including a sample of a known suspect and an unknown criminal’s speech, to assess the probability of obtaining the evidence under the hypothesis that the samples came from the same speaker versus the hypothesis that they came from different speakers. It is the role of the forensic expert to consider how similar (or different) the samples are and also to consider how typical any similar features are within a given population. When considering typicality, it is extremely important that a relevant reference population is selected. For example, if a particular feature found to be similar in the suspect and criminal recordings is extremely unusual in the selected reference population (typically the population that the criminal is thought to be from) this would provide strong support for the same speaker hypothesis.

However, if an inappropriate population is selected it may under- or overestimate the strength of evidence.

For this reason, an important area of research in this field is to understand how narrowly/broadly reference populations need to be defined. For instance, if a reference population is made up of speakers from Kirklees, but the criminal is in fact from Wakefield, it would be useful to know what impact this would have on the typicality estimation and the strength of evidence. This study looks to see how much speech varies between the boroughs of Bradford, Kirklees and Wakefield in order to better understand how to delimit the reference population.

2. Data

This study analyses data from the West Yorkshire Regional English Database (WYRED) [8]. WYRED is the largest forensically-relevant database of British English speech, containing recordings from 60 speakers from each of three metropolitan boroughs of West Yorkshire: Bradford, Kirklees and Wakefield. In total, 180 participants were recorded undertaking four style-controlled tasks.

2.1. Participants

The first 30 participants from WYRED were selected (10 each from Bradford, Kirklees and Wakefield) for this study. Participants are all male, aged 18-30 (mean=21.8, range=19-29), have English as their first and only language and were raised in an English-only speaking household. Participants were classified as being from one of the three boroughs based on the postcode of where they grew up and went to school. All of the participants were enrolled on undergraduate or postgraduate degrees at university or had already completed a university qualification at the time of recording.

2.2. Tasks

The present study considers three of the four tasks from WYRED. All tasks elicited spontaneous speech and were designed to mirror the contexts that may occur in evidential recordings. The studio recordings used in this study are presented in Table 1 below.

Table 1: *WYRED tasks used for this analysis.*

Task	Avg. Length	Speech Style
Task 1: Mock Police Interview	25 minutes	Formal, Spontaneous
Task 3: Paired Conversation	21 minutes	Relaxed, Spontaneous
Task 4: Answer Message	2 minutes	Time-constrained, Spontaneous

The experimental design for Task 1 involved a mock police interview in which participants had to provide certain information whilst denying anything incriminating. Using a map on an iPad as visual stimuli, certain target words were elicited which included a range of phonetic variables. Task 3 consisted of a casual conversation between pairs of participants from the same borough. Participants were provided with topic cards as prompts, however they were instructed that they could discuss any topics they like. Finally, Task 4 related to the crime scenario from Task 1 and involved participants leaving an answerphone message in a time-pressured situation. Participants were instructed to contact their fictional brother

and ask him to hide or dispose of any incriminating evidence. They were provided with some brief examples of evidence to mention but were encouraged to talk about additional unprompted information.

3. Methodology

This section outlines the methods that were used to conduct this study. It describes how the FACE tokens were selected, segmented and measured, as well as how statistical analysis was performed.

3.1. Token selection

For each participant, a maximum of 35 tokens were manually segmented from the sound files of Tasks 1, 3 and 4. Due to the length of the Task 4 recordings, significantly less tokens were available than in Tasks 1 and 3. The average number of tokens selected from Tasks 1, 3 and 4 were 33, 29 and 9 respectively. In total, 2113 tokens were segmented and analysed acoustically.

Tokens of FACE were selected from clearly articulated speech where there was no uncertainty as to what the intended target was. Any tokens produced in overlap or when the participant was laughing were disregarded. Tokens were only selected from mono- or bi-syllabic content words that contained FACE in the stressed syllable position. As a result of the experimental design of Tasks 1 and 4, a range of specific keywords occurred frequently in almost all of the participants' recordings, such as *steak*, *make* and *Rachel*.

As tokens of FACE occurred in a range of phonetic environments, care was taken to group tokens accordingly in case the phonetic context influenced the acoustic properties under examination. Tokens which occurred before or after a liquid were treated separately as liquids often cause lowering of F2 for front vowels [9,10]. It was also recorded when tokens occurred next to either a glide or a nasal as these segments are acoustically similar to vowels. The distribution of tokens across the aforementioned phonetic environments is presented in Table 2.

Table 2: *Phonetic environments of FACE tokens*

Environment	Number of tokens
Preceding a nasal	277
Preceding a liquid	12
Following a liquid	471
Following a glide	133
Between a liquid and a nasal	49
None of the above	1171

3.2. FACE segmentation and extraction

Tokens were manually segmented in Praat [11] and the word from which the vowel token was extracted was labelled in a TextGrid. For each token, the beginning of the first complete cycle and the end of the last complete cycle of the vowel were marked at zero crossings. All tokens were visually inspected in the spectrogram to determine the most appropriate number of formants required for the Linear Predictive Coding algorithm to take suitable measures. A modified version of a Praat script [12] was used to automatically extract measurements of the first three formants at 25%, 50% and 75% across the vowel. The data was subsequently exported to Microsoft Excel to be organised before statistical analysis was conducted in R [13]. The distances between the vowel onsets and offsets (measured at

25% and 75%, respectively) for F1~F3, were measured to capture how much movement there was across each token.

3.3. Statistical analysis

R [13] and lme4 [14] were used to perform a linear mixed effects analysis of the relationship between the quality of the vowel (midpoint formant values) and location. A further analysis was also conducted which considered the relationship between the movement of the vowel (offset formant values minus onset formant values) and location. In both cases, *borough* was entered into the model as a fixed effect and as random effects, there were intercepts for *subjects*, *phonetic environments* and *tasks*. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained by likelihood ratio tests of the full model with the fixed effect of *borough* against the model without the effect in question.

4. Results

The acoustic analysis of the formant data shows that there is relatively little movement from vowel onset to vowel offset, particularly in terms of F1. Figure 1 presents the average trajectories from vowel onset to offset, for all 30 participants. Each participant's trajectory is colour coded according to the borough that they are from. In terms of the direction of the trajectory, F2 consistently increases across the vowel and, for the majority of participants, F1 decreases across the vowel. Slight separation can be seen across boroughs both in terms of F1 and F2; however, the regional differences on the F2 dimension appear to be strongest, indicating that vowel front/backness is perhaps most regionally marked.

FACE in West Yorkshire is generally monophthongal. Impressionistically, FACE in Kirklees is close to [e] but slightly backed, Bradford is between [ɛ] and [e], while Wakefield is closer to [ɛ] and slightly fronted. The boxplot in Figure 2 visualises the midpoint formant data for each of the three boroughs. It can be seen that the distribution of FACE formant values varies across Bradford, Kirklees and Wakefield.

4.1. Linear mixed effects analysis results

The results of the linear mixed effects analysis described in §3.3 reveal that there are significant differences in terms of F1 and

F2 midpoint values; corresponding to vowel height as well as vowel front/backness, respectively. Taking Bradford as the intercept in the model, the ANOVA results are as follows:

- Borough affected F1 ($\chi^2(2)=8.60$, $p=0.0135$), lowering it by about 38.78 Hz \pm 13 (standard errors) for Kirklees and lowering it by about 13.99 Hz \pm 12.98 (standard errors) for Wakefield.
- Borough affected F2 ($\chi^2(2)=9.0765$, $p=0.0107$), lowering it by about 31.18 Hz \pm 35.87 (standard errors) for Kirklees and increasing it by about 77.92 Hz \pm 35.83 (standard errors) for Wakefield.

Significant differences were not found across boroughs in terms of F3 ($\chi^2(2)=3.0102$, $p=0.222$). In terms of distance measures for F1~F3, the results of the linear mixed effects analysis showed that there were no significant differences across boroughs. This means that, when taking into account all tokens rather than comparing averages (as in Figure 1), the movement in the FACE trajectories, did not significantly vary as a result of which borough the participant was from.

5. Discussion

This section discusses the results of this study in more detail and highlights the implications of the findings for researchers in the fields of forensic speech science, automatic speech recognition and sociophonetics more generally.

5.1. General findings

The existing literature on West Yorkshire English had previously indicated that the production of FACE varied across the region. This study illustrates that this is still the case by using both auditory and acoustic analyses to substantiate these findings. Although the auditory analysis did not reveal any clear distinctions in the way FACE was pronounced between Bradford, Kirklees and Wakefield, the acoustic data shows that there is variation across boroughs in terms of F1~F3. The linear mixed effects and ANOVA results illustrated that F1 and F2 are more regionally influenced than F3. One explanation for why F3 may vary the least across boroughs is that this parameter is generally considered to be more idiosyncratic [15,16,17] and therefore less likely to be dependent on external factors, such as the place the speaker is from.

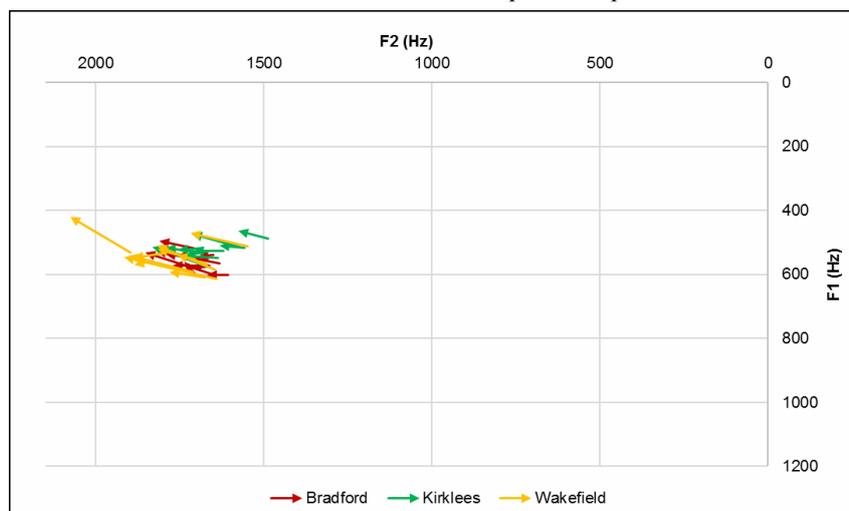


Figure 1. Average F1 and F2 values at 25% and 75% of FACE vowel plotted for all 30 participants

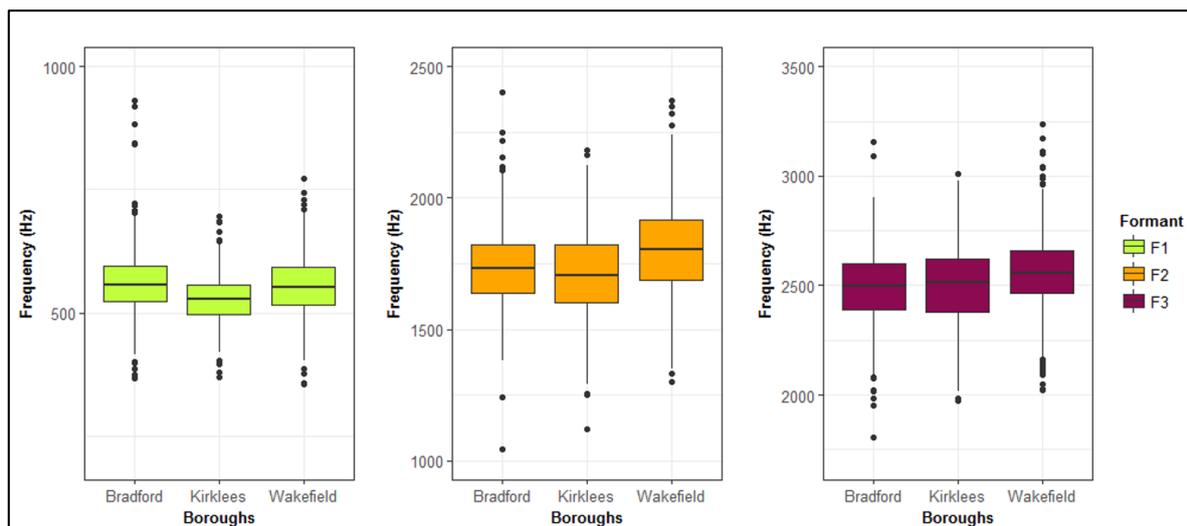


Figure 2: Midpoint formant data across West Yorkshire boroughs

It is important to note that even though the participants formed a fairly homogeneous group insofar as their social characteristics (such as sex, age group and influence of other languages), significant differences in F1 and F2 were observed across boroughs. If we were to take a random sample of the male population of West Yorkshire, it is likely that the extent to which FACE varies across the region would be even greater than this. Furthermore, anecdotal evidence observed during the data collection process suggests that regional variation could be even more fine-grained than at the local-level. Participants have claimed to be able to tell the difference between accents within the same borough (e.g. Pontefract and Hemsworth); however, further analyses on an even more “microscopic” level is required to examine whether this can be corroborated with acoustic information.

5.2 Implications

The findings of this study highlight that there is regional variation present on a more local-level than one might expect. The following sections address the potential implications this could have on practical applications.

5.2.1 Forensic speech science

A vital part of FSC casework involves making an assessment of how typical a particular speech parameter is in a given population. This study informs caseworkers of how FACE is realised across West Yorkshire by providing new reference population data for an area for which no acoustic data was previously available. Furthermore, this study addresses the broader question of how narrowly a reference population needs to be delimited. The results of this study indicate that, for FACE at least, significant acoustic differences exist at a local-level even within a fairly homogeneous community. For this reason, it is advised that attention must be paid to local-level variation when assessing the strength of evidence in order to avoid over- or under-estimations.

5.2.2 Speech Technology

A greater understanding of regional variation at a local-level could assist engineers in designing automatic speech recognition (ASR) systems. ASR systems are typically trained

using many hours of speech and the more closely the training data matches the test data, the higher the expected recognition performance. As this study has shown that there are fine-grained acoustic differences across relatively small, neighbouring communities, it may be the case that using more local-level datasets would help to improve recognition performance. For example, if an ASR system were to be trained using speech from across all of West Yorkshire, as opposed to from one specific part of the region, it is likely that the system would more accurately recognise West Yorkshire English speech.

5.2.3 Sociophonetics

As well as providing a description of an area which has seldom been analysed from a sociophonetics perspective, the findings of this study can be compared to previous studies in order to investigate language change in West Yorkshire. The evidence of regional variation in this study suggests that speakers from West Yorkshire may have a particularly strong sense of local identity even at the level of their metropolitan borough.

6. Conclusions

This paper has shown that FACE realisations are significantly different in terms of F1 and F2 across Bradford, Kirklees and Wakefield. These results suggest that West Yorkshire, or Yorkshire generally, is more regionally stratified than previously recognised. This provides motivation for further investigation into other phonetic variables and how they vary at a local-level. This is especially important for the field of forensic speech science when an expert is delimiting the reference population in a criminal investigation.

7. Acknowledgements

We would like to thank Sula Ross, our colleague on the WYRED Project, our team of undergraduate transcribers at the University of Huddersfield and all of the people from West Yorkshire who participated in the study. We would also like to thank the anonymous reviewers for their valuable feedback and suggestions.

8. References

- [1] J. C. Wells, *Accents of English* (Vol. 1). Cambridge: Cambridge University Press, 1982.
- [2] J. C. Wells, "English accents in England," in P. Trudgill (Ed.) *Language in the British Isles*. Cambridge: Cambridge University Press, 1984, pp. 55-69.
- [3] P. Foulkes and G. Docherty (Eds.), *Urban voices: accent studies in the British Isles*. London: Arnold, 1999.
- [4] A. Hughes, P. Trudgill and D. Watt, *English accents and dialects: An introduction to social and regional varieties of English in the British Isles, 5th edition*. London: Arnold, 2012.
- [5] A. Easter, *A glossary of the dialect of Almondbury and Huddersfield*. Vaduz: Kraus, 1883 (1965 reprint of original).
- [6] K. M. Petyt, *Dialect and Accent in Industrial West Yorkshire*. Amsterdam: John Benjamins, 1985.
- [7] K. Burland, "Where the Black Country Meets 'Black Barnsley': Dialect Variation and Identity in an Ex-Mining Community of Barnsley," in C. Montgomery and E. Moore (Eds.) *Language and a Sense of Place: Studies in Language and Region*. Cambridge: Cambridge University Press, 2016, pp. 234-257.
- [8] E. Gold, K. Earnshaw, and S. Ross, "An Introduction to the West Yorkshire Regional English Database (WYRED)," in K. Burland and C. Upton (Eds.), *Transactions of the Yorkshire Dialect Society*, 2016, pp. 1-8.
- [9] P. Ladefoged, and I. Maddieson, *The Sounds of the World's Languages*. Oxford: Blackwell, 1996.
- [10] P. Ladefoged, *A Course in Phonetics, 4th edition*. Fort Worth: Harcourt, 2001.
- [11] P. Boersma, and D. Weenink, *Praat: doing phonetics by computer* [Computer program]. Version 6.0.29, retrieved November 2017 from <http://www.praat.org/>.
- [12] M. Lennes, *SpeCT - Speech Corpus Toolkit for Praat (v1.0.0)*. First release on GitHub (Version 1.0.0). Zenodo, retrieved March 2017 from <http://doi.org/10.5281/zenodo.375923>.
- [13] R Core Team, "R: A language and environment for statistical computing," *R Foundation for Statistical Computing*, Vienna, Austria, 2017, URL <https://www.R-project.org/>.
- [14] D. Bates, M. Maechler, B. Bolker, and S Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1-48, 2015.
- [15] K. McDougall, "Speaker-specific formant dynamics: an experiment on Australian English /a/," *International Journal of Speech, Language and the Law*, vol. 11 no.1, pp. 103-130, 2004.
- [16] V. Hughes, K. McDougall, and P. Foulkes, "Diphthong dynamics in unscripted speech," in *International Association of Forensic Phonetics and Acoustics Conference, August 2-5, Cambridge, UK, 2009*.
- [17] E. Gold, J. P. French, and P. Harrison, "Examining long-term formant distributions as a discriminant in forensic speaker comparisons under a likelihood ratio framework," *Proceedings of Meetings on Acoustics (POMA)*, vol. 19, no.1, pp. 060041-060041, 2013.