The 'West Yorkshire Regional English Database': Investigations into the generalizability of reference populations for forensic speaker comparison casework

Erica Gold, Sula Ross, Kate Earnshaw

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

e.gold|s.m.ross|k.earnshaw@hud.ac.uk

Abstract

The West Yorkshire Regional English Database (WYRED) consists of approximately 196 hours of high-quality audio recordings of 180 West Yorkshire (British English) speakers. All participants are male between the ages of 18-30, and are divided evenly (60 per region) across three boroughs within West Yorkshire (Northern England): Bradford, Kirklees, and Wakefield. Speakers participated in four spontaneous speaking tasks. The first two tasks relate to a mock crime where the participant speaks to a police officer (Research Assistant 1) followed by an accomplice (Research Assistant 2). Speakers returned a minimum of a week later at which point they were paired with someone from their borough and recorded having a conversation on any topics they wish. The final task is an experimental task in which speakers are asked to leave a voicemail message related to the fictitious crime from the first recording session. In total, each speaker participated in approximately 1 hour of spontaneous speech recordings. This paper details the design of WYRED, in order to introduce forensic speech science research utilizing this data, and to promote WYRED's potential application in related research and in forensic speech science casework.

Index Terms: forensic speech science, reference population, regional variability, speech database

1. Introduction

The primary motivation for the construction of the West Yorkshire Regional English Database (WYRED) was to provide a collection of regionally stratified speech recordings (by boroughs) from within a single, politically defined region (a county). The corpus aims to facilitate research on methodological issues surrounding the delimitation of the reference population when considering the typicality of a speech sample for a given forensic speaker comparison case. The high quality and large volume of audio data (and accompanying transcriptions) collected as part of the WYRED project has been facilitated by funding from the Economic and Social Research Council in the United Kingdom (ES/N003268/1). The project investigates the empirical implications of defining regional accents too narrowly/broadly for forensic speaker comparison (FSC) casework.

FSC typically involves the comparison of a criminal recording (e.g. threatening voicemail message) and a known suspect sample (e.g. police interview). The expectation of the expert is to conduct an assessment and comparison of the similarities and differences in the speech parameters present (or absent) in the recordings, regardless of whether that is carried out by auditory-acoustic analysis or an automatic speaker recognition system. In the United Kingdom, the suspect sample is usually a recording of a police interview with the suspect [1,2]. The objective of the expert is to provide the trier(s) of fact with an educated opinion regarding the probability of obtaining the evidence (the similarities/differences between the criminal and suspect samples) under the hypothesis that the samples came from the same person, versus the probability of obtaining the evidence (the typicality of the analysed speech parameters) under the hypothesis that two different speakers produced the criminal and suspect samples. However, in order to more precisely assess the typicality of the analysed speech parameters, relevant population/reference data must be consulted. Unfortunately, there are two main obstacles impeding the ease with which forensic experts can consult population data. The first obstacle is fundamental insofar as the high degree of heterogeneity in speech presents a challenge for experts in selecting the appropriate population data with which to compare their criminal sample [3,4,5,6]. The lack of population data is the second obstacle, and is a practical issue as population data is a vital resource in conducting FSC casework. Arguably, the lack of population data is the biggest problem currently facing the field [7,8].

Identifying the appropriate population data for a forensic speaker comparison case (i.e. delimiting the population) is important in accurately representing the strength of evidence [4,5,6]. However, there is no research to inform experts on the level at which population data need to be defined. Consider two hypothetical examples, where both cases share the same prosecution hypothesis - for a specific speech parameter the criminal and suspect sample are very similar. In the first example, the speech parameter under investigation is found to be unique in the relevant population, which would result in a strong strength of evidence in favour of the prosecution. In the second example, the speech parameter under investigation is found to be extremely common in the population, which would result in a much weaker strength of evidence than the first example (however, it would still be in favour of the prosecution). In practice, population data plays a vital role in estimating the strength of evidence, and serious consequences for under- or over-estimating the strength of evidence can occur when the inappropriate population data is selected for consultation.

There are currently a limited number of forensically-relevant, English databases available that are used in forensic speaker comparison research and casework [9,10,11,12]. However, there are no forensically-relevant, English databases that allow for the testing of accent generalizability. WYRED fills this void, as it is the first database of its kind, to include a large volume of high quality audio from a carefully stratified population, to support the investigation of strength of evidence effects in generalizing the reference population in FSCs.

2. Corpus design

The following section details the corpus design of WYRED, including: speakers, metadata, speaking tasks, recording sessions, and recording set-up. All participant recordings are available and accompanied by orthographic transcriptions carried out (manually) in textgrids in Praat.

2.1. Speakers

WYRED consists of recordings from 180 male speakers, aged between 18 and 30 at the time of recording. All participants are British English speakers from Northern England in the county of West Yorkshire (see Figure 1). The 180 speakers are divided between three of the five boroughs within West Yorkshire (Bradford, Kirklees, Wakefield), such that there are 60 speakers from each of the boroughs. Participants were assigned to a borough based on the postcode (zip code) where they grew up and went to primary and secondary school. All participants are native English speakers who grew up in English-only speaking households and did not speak any other languages. None of the participants reported any speech or hearing impairments. Speakers, however, were not included in the database if they were deemed to have spent a significant period (more than a few years) outside the area, had missing/broken front teeth or facial piercings that affected their speech.

Participants were largely recruited from the University of Huddersfield, but also came from the surrounding communities within the boroughs of interest. Recruitment largely took place through email advertisements, but also via flyers, in class presentations, Facebook Ads, and referrals. All interested participants registered their interest in participating through an online survey that allowed us to screen for eligible participants. Speakers were then invited to participate via email. All participants were compensated for their participation.

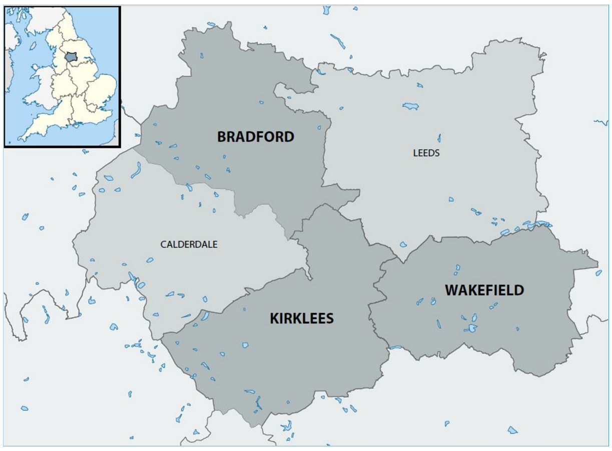

Figure 1: *Map of Great Britain and West Yorkshire*

2.1.1. *Metadata collected from speakers*

In addition to each participant's age and postcode, WYRED also contains metadata that may be of interest to other researchers. The following metadata has been collected for each participant:

- Relationship status and where their partner was from
- Where the participants' parents were from
- Employment status and type of work
- Highest level of education
- Smoker/Non-smoker
- Left or right handed
- Height and weight

2.2. Speaking tasks

All participants were recorded over four different spontaneous speaking tasks. The type of task, recording channel, and approximate length of recording are presented in Table 1.

Table 1: *Speaking tasks*

Task	**Channel**	**Length**
1. Mock Police Interview	Studio	~ 20 mins
2. Accomplice Call	Studio; Phone	~ 15 mins
3. Paired Conversation	Studio	~ 20 mins
4. Voicemail Message	Studio; Phone	~ 2 mins

The first two tasks replicate the methodology used in the Dynamic Variability in Speech (DyViS) project [9] and contain spontaneous speech generated by using a map as a visual stimulus in order to encourage the elicitation of specific tokens. Task 3 is a spontaneous conversation with a paired participant (similar age, same gender and region). The final task is an experimental short recording where the participant is asked to leave a voicemail message, with a rough guide as to the information they have to leave, in a time-pressured situation. Further participant instructions for each task are provided in the subsequent sub-sections.

2.2.1. *Task 1: Mock Police Interview*

On entering the sound booth participants were advised that they were about to take part in a mock police interview. The participants were provided with a brief background to the investigation before being presented with written information displayed on an iPad. The information provided them with an opportunity to familiarize themselves with the role of the suspect prior to the recording.

2.2.2. *Task 2: Accomplice Call*

The participants were informed that, having being interviewed by the police, their next task was to phone their friend and accomplice. The purpose of this call was to ensure that the accomplice did not implicate the suspect any further in the crime by contradicting any of the details that the suspect provided to the police. Access to the information from the interview was provided in the form of a storyboard poster attached to a wall in the sound booth. Participants were instructed to be thorough and provide as much of the information on the poster as possible, but advised that the researcher receiving the phone call would ask questions and prompt them.

2.2.3. *Task 3: Paired Conversation*

The participants were told that they would be left alone to talk to each other, without a researcher present, for 20 minutes. They were provided with topic cards (e.g. work, hobbies, education, hometown), adapted from [11]. They were requested to avoid mentioning personal details and names of individuals, but advised that any identifiable information recorded would be edited out to ensure the recordings were anonymous.

Participants were advised to act naturally and speak as if they were having an ordinary conversation with a friend.

2.2.4. *Task 4: Voicemail Message*

Participants were reminded of the mock police interview from Task 1. They were then advised that they were about to be arrested and this was their only opportunity to make a phone call. They were instructed to ring their brother, John, and leave a voicemail message. In this voicemail message they had to ask their brother to hide or destroy any incriminating evidence and request that their brother made contact with their accomplice immediately. The participants were provided with four bullet-point examples of evidence that may need to be hidden or destroyed, taken from Task 1, but encouraged to provide further unprompted information in addition to this. Finally, the participants were advised that the recording should ideally be approximately 2 minutes long and a timer was provided so they could monitor how long they had been speaking for.

2.3. Recording sessions

Recordings were carried out over two separate sessions that were separated by a minimum of one week. Participants recorded the first two tasks on their initial visit, and recorded the final two tasks in their second visit. Non-contemporaneous speech was collected as it is a significant concern in forensic research [13,14,15] given the inherent variability present even within a speech recording produced in a single session. Session 1 and 2 were recorded a minimum of a week apart for all participants, but due to limitations in recruitment and participant availability some participants attended their second session up to a couple of months later. However, the dates for all recordings are included in the naming convention for all files in the database.

Participants had the option to be paired with another participant for Task 3 by the research team or to nominate another participant they were acquainted with to be their partner. The large majority of participants were in fact paired by the research team. Pairings were made of course within boroughs, but the research team also aimed to match speakers who were from areas that were geographically close to each other. This resulted in some pairings having even grown up on the same street. All pairs' familiarity level (in terms of previous acquaintance) is marked on all Task 3 file naming conventions in terms of NF (non-friend) or F (friend).

2.4. Recording set-up

The database was recorded in a professional 2.3 by 1.6 meter, purpose-built sound booth in the Forensic Speech Science Lab at the University of Huddersfield. The sound booth is a stand-alone recording studio that has been secured into the floor and ceiling of the lab. The ceiling and interior walls are covered in acoustically transparent fabric and the booth contains laminate flooring. All tasks for each participant were recorded inside the sound booth. Participants sat at a desk inside the sound booth and wore a Sennheiser HSP 4 omnidirectional headband microphone that was situated approximately 2 cm from their mouth. Recordings were made on a Marantz PMD661 MKII Handheld Solid State Recorder in PCM WAV format (44.1kHz, 16 bit). Figure 2 provides a schematic of the recording set up for inside the sound booth.

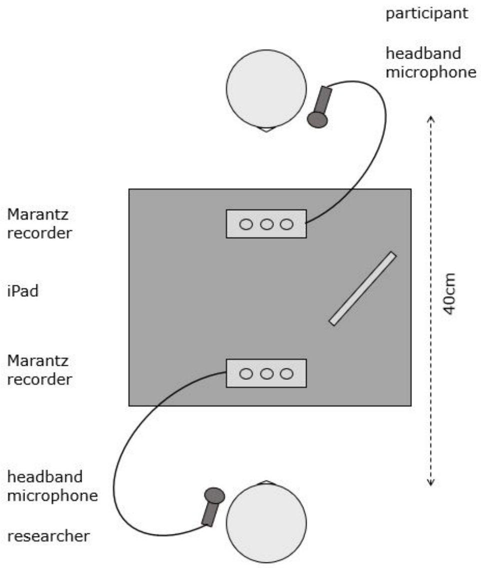

Figure 2: *Participant recording set up*

For Task 1 and Task 3, the Police Interviewer (Research Assistant 1) and the paired partner (respectively) also wore a Sennheiser HSP 4 omnidirectional headband microphone and each were also recorded on a separate Marantz PMD661. For Task 2 and Task 4, participants were recorded in the same format as Task 1 and Task 3, but they were also recorded over a cordless BT Diverse 7410 Plus landline telephone. For Task 2, calls were intercepted and recorded using a Prospect Electronics TC22 telephone balance unit that was connected to both a Mackie micro series 1202 – VLZ line mixer and a Marantz PMD661 MKII Handheld Solid State Recorder. For Task 4, voicemail messages were recorded on a Tiptel 540 answerphone. It is important to note that all equipment was battery operated, aside from the base of the wireless telephone in order to minimize mains interference.

The accomplice's speech (Research Assistant 2) for Task 2 was the only recording made outside the sound booth. The accomplice was recorded from the far end of the telephone line. The accomplice used a Sennheiser MD4ZI – II handheld mic that was placed on a stand on a desk approximately 10 cm from their mouth that was connected to the Mackie line mixer. Figure 3 provides a schematic of the accomplice recording set up.

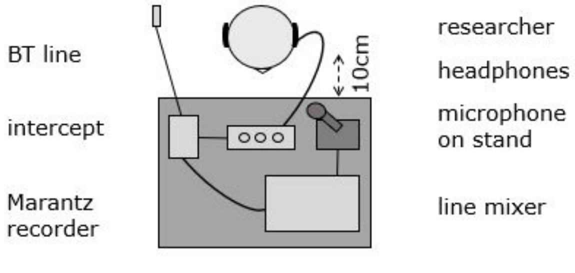

Figure 3: *Accomplice recording set up*

3. Applications of the database

The following sections consider the practical applications of WYRED to the generalizability of populations, regional variation, automatic speaker recognition research, and forensic speech science casework.

3.1. Generalizability of populations

As noted in §1, strength of evidence is sensitive to the delimitation of the background population in FSC cases (e.g. using New Zealand English for a Southern Standard British English speaker; see [9]). However, it is not known to what extent strength of evidence sensitivity exists when narrowly defining reference populations. Given that the database collection process is extremely time-consuming and costly, it would be ideal if reference populations could be generalized and collected at a more macro-level. For this reason, WYRED enables empirical studies into the generalizability of populations, and aids in identifying whether narrowly-defined population groups are advised for the majority of FSC cases, or whether more broadly-defined populations are sufficient.

WYRED is the largest database of its kind and the careful design set up allows for generalizability studies to be carried out across a vast range of phonetic or linguistic variables. In addition to the research that can be carried out within WYRED, further research can be undertaken using similar databases that use some of the same speaking tasks [9,16]. This will allow for additional generalizability studies that are spread across more geographically distant locations.

3.2. Regional variation

In general, there is relatively little sociophonetic research on Bradford, Kirklees, and Wakefield. What does exist is largely outdated [17,18] and does not provide a full representation of variation that might be present within West Yorkshire. Rather, the three areas are often grouped together as an example of a general Yorkshire accent [18,19,20,21], without considering the micro-regional variation that might exist between these three boroughs (see preliminary work on regional variation in the quality of filled pauses, [22]). Participants in WYRED have anecdotally stated that they can hear differences between the three boroughs and are able to determine where a person is from within West Yorkshire. However, these anecdotal statements have not yet been backed up empirically.

Although the investigation of regional variation was not the main motivation for creating WYRED, the corpus readily lends itself to empirical sociolinguistic studies of young male speech across boroughs within West Yorkshire. Studies can be undertaken to establish vowel spaces, phonological processes, and syntactic structure, to name but a few possibilities. We believe that the WYRED corpus will shed light on an underrepresented variety of British English, while also showing that local accents and dialects are perhaps more diverse than currently acknowledged.

3.3. Automatic speaker recognition research

WYRED was constructed from the perspective of phoneticians, however, automatic speaker recognition (ASR) research was also considered in the development of the database. Although on the smaller side of typical speech corpora used in ASR research, the innovative design of WYRED allows for a number of possible research projects.

The corpus includes both short and long recordings of participants speaking to different interlocutors, recorded over different channels, and includes non-contemporaneous speech. All of these variables are useful in testing, calibrating, and developing automatic systems. In addition to the technical variables provided in the corpus, there are metadata variables (e.g. age, smoker/non-smoker, level of education) provided that researchers may find useful should they wish to work on accent identification (similar to [23,24]) or speaker classification. Even if researchers or developers are not interested in a specific configuration of variables, they may at least find the corpus useful as an additional set of reference data or testing material.

3.4. Forensic casework

In the United Kingdom there are roughly 500-600 forensic speaker comparison cases each year [25]. Of those 500-600 cases, an appreciable number of cases will include voices from Northern England and more specifically West Yorkshire. However, at present we lack solid empirical data in which to ground estimates regarding the distribution of any phonetic variables for any variety of Northern British English speech. These phonetic variables relating to the West Yorkshire accent(s) are vital in determining accent profiles insofar as they are able to document typically expected phonetic features for a given region. This type of detail will allow experts to make more transparent and robust interpretations of the evidence they are presented with. In addition to general accent or even dialect profiles, future research on WYRED will begin to generate reference population data in respect of a wide range of phonetic or linguistic variables. As noted in §1, population data is vital in determining the strength of evidence in a forensic speaker comparison case. Without such data or literature, experts are left either to use their experience as a substitute for empirical statistics or to estimate distributions of phonetic features based on other regional accents or languages that might not be entirely relevant. For these reasons, for forensic speech science cases that are carried out on Northern British accents, and specifically those from West Yorkshire, WYRED is an invaluable resource.

4. Conclusion

The West Yorkshire Regional English Database, a high-quality collection of audio data from 180 male speakers of West Yorkshire English, enables research on the generalizability of reference populations, regional variation, and automatic speaker recognition. This has paper specified the corpus design and introduced a number of practical applications for the data including forensic speech science casework. It has also provided a brief introduction to problems currently facing the field of forensic speech science and in turn introduced the motivation for the creation of the West Yorkshire Regional English Database.

5. Acknowledgements

This project has been funded by the Economic and Social Research Council in the United Kingdom (ES/N003268/1) under the 2016 Future Research Leaders scheme. The authors would like to thank all of our West Yorkshire participants as well as our team of undergraduate student transcribers at the University of Huddersfield.

6. References

[1] F. Nolan, *The Phonetic Bases of Speaker Recognition*. Cambridge: CUP, 1983.

[2] P, Rose, *Forensic Speaker Identification*. London: Taylor and Francis, 2002.

[3] E. Gold and V. Hughes, "Issues and opportunities: the application of the numerical likelihood ratio framework to forensic speaker comparison," *Science & Justice*, vol. 54, no. 5, pp. 292-299, 2014.

[4] V. Hughes, *The definition of the relevant population and the collection of data for likelihood ratio-based forensic voice comparison.* York: University of York PhD thesis, 2014.

[5] V. Hughes and P. Foulkes, "Regional variation and the definition of the relevant population in likelihood ratio-based forensic voice comparison using cepstral coefficients". *Proceedings of the 15th Australasian International Conference on Speech Science and Technology.* Hay, J. & Parnell, E. (eds.) University of Canterbury, New Zealand, vol. 15, pp. 24-27, 2014

[6] V. Hughes and P. Foulkes, "What is the relevant population? Considerations for the computation of likelihood ratios in forensic voice comparison," in INTERSPEECH 2017 – *97th Annual Conference of the International Speech Communication Association, Stockholm, Sweden, Proceedings,* 2017, pp. 3772-3776.

[7] P. Rose and G. S. Morrison, "A response to the UK position statement on forensic speaker comparison," *International Journal of Speech, Language and the Law,* vol. 16, no. 1, pp. 139-163, 2009.

[8] P. French, F. Nolan, P. Foulkes, P. Harrison, and K. McDougall (2010). "The UK position statement on forensic speaker comparison: a rejoinder to Rose and Morrison," *International Journal of Speech, Language and the Law,* vol. 17, no. 1, pp. 143-152, 2010.

[9] F. Nolan, K. McDougall, G. de Jong & T. Hudson, "The DyViS database: style-controlled recordings of 100 homogenous speakers for forensic phonetic research," *International Journal of Speech, Language and the Law*, vol. 16, no. 1, pp. 31-57, 2009.

[10] N. Fecher, "The 'Audio-Visual Face Cover Corpus': Investigations into audio-visual speech and speaker recognition when the speaker's face is occluded by facewear," in *INTERSPEECH 2012 - 13th Annual Conference of the International Speech Communication Association*, *September 9-13, Portland, Oregon, USA*, *Proceedings,* 2012, pp. 2250-2253.

[11] J. Wormald, *Regional Variation in Panjabi-English*. York: University of York PhD thesis, 2016.

[12] K. McDougall, "Listeners' perception of voice similarity in Standard Southern British English versus York English," in *IAFPA 2014 - 23rd Annual Conference of the International Association for Forensic Phonetics and Acoustics*, *August 31 - September 3, Zürich, Switzerland*, 2014.

[13] E. Enzinger and G. S. Morrison, "The importance of using between-session test data in evaluating the performance of forensic-voice-comparison systems," in *the 14th Australasian International Conference on Speech Science and Technology, Sydney, Australia, Proceedings,* 2012, pp. 137-140.

[14] G. S. Morrison, F. Ochoa, and T. Thiruvaran, "Database selection for forensic voice comparison," in *Proceedings of Odyssey 2012: The Language and Speaker Recognition Workshop, Singapore*, 2012, pp. 74.77.

[15] G. S. Morrison, P. Rose, and C. Zhang, "Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice". *Australian Journal of Forensic Sciences*, vol. 44, no. X, pp. 155-167, 2012.

[16] C. Llamas, D. Watt, P.French, A. Braun, and D. Robertson, "Routinised mobility and vowel change in the North East of England' Paper presented at iCLaVE, June 6-9, Málaga, Spain, 2017.

[17] K. M. Petyt, *Dialect and Accent in Industrial West Yorkshire*. Amsterdam: John Benjamins, 1985.

[18] J. C. Wells, *Accents of English* (Vol. 1). Cambridge: Cambridge University Press, 1982.

[19] J. Wright, *A grammar of the Dialect of Windhill in the West Riding of Yorkshire: Illustrated by a series of dialect specimens, phonetically rendered; with a glossarial index of the words used in the grammar and the specimens*. London: Kegan Paul, Trench, Trübner and Co., 1892.

[20] J. C. Wells, "English accents in England" in P. Trudgill (Ed.) *Language in the British Isles*. Cambridge: Cambridge University Press. 1984, pp. 55-69.

[21] S. Wilhelm, "Segmental and suprasegmental change in North West Yorkshire – a new case of supralocalisation?", Proceedings of the 16th Conference on Spoken English (ALOES), Villetaneuse, April 1-2, 2016.

[22] E. Gold, K. Earnshaw, and S. Ross, "An introduction to the West Yorkshire Regional Database (WYRED): Examining the variability of hesitation markers". *UKLVC 2017 - 11th UK Language Variation and Change Conference, August 29-31, Cardiff, Wales, UK*, 2017.

[23] G. Brown and J. Wormald, "Automatic Sociophonetics: exploring corpora with a forensic accent recognition system," *Journal of the Acoustical Society of America: Special Issue on Advancing Methods for Analyzing Dialect Variation*, vol. 142, no. 1, pp. 422-433, 2014.

[24] G. Brown, *Modelling pronunciation with distance: an automatic speaker comparison system for forensic applications*. York: University of York PhD thesis, 2017.

[25] Peter French, personal communciation