

WYRED: West Yorkshire Regional English Database

Erica Gold, Kate Earnshaw, Sula Ross

Department of Linguistics and Modern Languages, University of Huddersfield

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

Introduction

In the field of forensic speech science, the availability of population statistics and forensically-relevant speaker databases are limited (Gold and Hughes 2014). Without population statistics it is difficult for experts to accurately (and transparently) estimate the typicality of a speech parameter for a given population. The population statistics that are available for British English are largely thanks to analysis that has been carried out on the first forensically-relevant English database (DyViS; Nolan et al., 2009).

Motivated by the inherent lack of population statistics, the West Yorkshire Regional English Database (WYRED) - supported by the Economic and Social Research Council (ES/N003268/1), is the largest forensically-relevant British English database. This paper provides an overview of the database and the recordings and metadata included.

The Database

WYRED includes a total of 180 speakers that come from three metropolitan boroughs (Bradford, Kirklees & Wakefield) within the county of West Yorkshire in the United Kingdom. Participants have English as their first and only language, have gone to school in one of the three areas, and have grown up in an English-only speaking household. Each borough is comprised of 60 male speakers between 18 and 30 years old.

WYRED consists of four studio-quality spontaneous speech tasks recorded over two sessions that are separated by at least one week. The recordings included in WYRED closely follow the collection procedures used by the DyViS database for Task 1 and Task 2 (Nolan *et al.*, 2009). New tasks replace the DyViS Tasks 3 and 4. The WYRED Task 3 consists of non-crime related discussion (e.g. sport, food) in a paired conversation (another participant from the same area). Task 4 is a short voicemail message to the participants' fictional 'brother' relating to the disposal/hiding of evidence from the crime discussed in Task 1 and 2. In total each speaker has about an hour of speech recorded, with the entire database including over 180 hours of recorded speech.

WYRED acknowledges the multifaceted uses that the database could have, and as a result includes additional metadata for each participant, including: postcode, age, education level, parents' home town, smoker/non-smoker, and length of time spent outside West Yorkshire. The database also provides orthographic transcriptions of the recordings in PRAAT and copies of all the materials used to elicit speech.

References

- Gold, E. and Hughes, V. (2014). Issues and opportunities: The application of the numerical likelihood ratio framework to forensic speaker comparison, *Science and Justice*, 54 (4), pp. 292-299.
- Nolan, F., McDougall, K., de Jong, G. and Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogenous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law*, 16(1), pp. 31-57.