

<p>Institution: University of Cambridge</p>
<p>Unit of Assessment: UoA28</p>
<p>Title of case study: Phonetics applied to Forensic Speaker Comparison</p>
<p>1. Summary of the impact (indicative maximum 100 words) Increasingly in court cases the recorded voice of a perpetrator has to be compared with that of a suspect. Research on speaker characteristics carried out by/under Prof. Nolan has directly contributed to the work of those offering forensic speech services commercially or developing relevant speech processing software. Impact arises from seminal ideas such as LTF (Long Term Formant) analysis, and from the 100-speaker 'DyViS' accent-matched database. The latter has directly enabled: the testing of an automatic speaker recognition system preparatory to its incorporation into forensic casework; the development of speaker recognition and speaker separation software; the adoption of systematic 'voice quality' analysis; and the availability for casework of population statistics on pitch and disfluencies. Public engagement has raised awareness of the possibilities and limitations of speaker identification in legal and general audiences.</p>
<p>2. Underpinning research (indicative maximum 500 words) Research on speaker characteristics has been undertaken at the University of Cambridge in the Department of Linguistics¹ by Prof. Francis Nolan², Dr. Kirsty McDougall³, Dr. Gea de Jong⁴, and Toby Hudson⁵. The general research direction, applying phonetics to speaker identification, was defined in Nolan's (1983, reprinted 2009) <i>The Phonetic Bases of Speaker Recognition</i> and summarised accessibly in Nolan (1997) [3.1]. Subsequent papers, such as Nolan (2005) [3.3] challenging the lack of use in forensic casework of Laver's framework for voice quality analysis, and Nolan and Grigoras (2005) [3.2] demonstrating and advocating the use of long-term average formant analysis, have dealt with specific areas of phonetic description and subsequently influenced practice in forensic casework.</p> <p>Central to forensic speaker comparison is knowing: (a) how the speech of an individual can vary, and (b) how much variation there is among speakers in the larger population. The lack of population statistics relevant to (b) has been often lamented, but is explicable given the multiplicity of quantifiable properties in speech, the mix of linguistic and personal factors determining a person's speech, and the fact that each speaker is a 'moving target', producing quite different speech on different occasions. To help rectify this Nolan undertook the ESRC-funded <i>Dynamic variability in speech: a forensic study of British English [DyViS]</i> [3.6].</p> <p>'DyViS' made tractable the problem that linguistic and personal information are convolved in speech by controlling for linguistic variation, and recording 100 speakers closely matched for accent, and within a narrow age-range (18-25). By controlling in this way, the range of variation attributable solely to personal voice characteristics (resulting from anatomy and individual speech habits) can be studied. This is the 'limiting case' for voice ID, where no difference of accent is apparent. Furthermore, the DyViS database [3.5] includes four different speaking tasks (two involving spontaneous dialogue, one of these being a telephone call recorded in high quality and over the telephone line), and (for 20 of the speakers) a second recording at a later date, so that within-speaker variation can be estimated. The database [created 2006-2011] constitutes a resource of wide utility for forensic (and other) speech research. For instance, at the 2012 Conference of the <i>International Association for Forensic Phonetics and Acoustics</i> 7 out of 25 oral presentations reported research using the DyViS database.</p> <p>As exemplars of the use of the DyViS database, the Cambridge group derived fundamental frequency ('voice pitch') statistics for the 100 speakers as in Hudson et al. (2007) [3.4]; and used the closely matched speakers from the database in an ESRC-funded project, <i>Voice similarity and the effect of the telephone: a study of the implications for earwitness evidence [VoiceSim]</i> [3.7], which explored perceptual similarity of voices and how this, and identification accuracy, are affected by the telephone.</p>

Impact case study (REF3b)

¹ Merged with the Research Centre for English & Applied Linguistics since October 2011 as the Department of Theoretical and Applied Linguistics.

² Professor of Phonetics (10/2004–) and previously Reader, Lecturer, and Assistant Lecturer (since 10/1978).

³ Research Associate (10/2005–12/2009), BA Postdoctoral Fellow (01/2010–).

⁴ Senior Research Associate (01/2006–09/2008).

⁵ Research Assistant (01/2006–09/2007, 01/2008–12/2008).

3. References to the research (indicative maximum of six references)

[3.1] F. Nolan (1997) Speaker recognition and forensic phonetics. In: W. Hardcastle and J. Laver (eds), *The Handbook of Phonetic Sciences*. Oxford: Blackwell, pp 744-67.

[3.2] F. Nolan & C. Grigoras (2005) A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law* 12(2), 143-173.

[3.3] F. Nolan (2005) Forensic speaker identification and the phonetic description of voice quality. In: W. J. Hardcastle & J. Beck (eds), *A Figure of Speech: a Festschrift for John Laver*. Mahwah, New Jersey: Erlbaum, pp 385-411.

[3.4] T. Hudson, G. de Jong, K. McDougall, P. Harrison & F. Nolan (2007) F0 statistics for 100 young male speakers of Standard Southern British English. *Proceedings of the 16th International Congress of Phonetic Sciences*, 6-10 August 2007, Saarbrücken. 1809-1812.

[P. Harrison works for JP French Associates and provided enhanced F0 software used in the work]

[3.5] F. Nolan, K. McDougall, G. de Jong & T. Hudson (2009) The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law* 16(1), 31-57.

All outputs can be supplied by the University of Cambridge on request.

Grants

[3.6] *Dynamic variability in speech: a forensic study of British English*. 01/10/2005—31/12/2009, PI Francis Nolan. ESRC Award RES-000-23-1248, £402,942 [pre-fEC].

[3.7] *Voice similarity and the effect of the telephone: a study of the implications for earwitness evidence*. 01/01/2008—31/12/2008, PI Francis Nolan. ESRC Award RES-000-22-2582, £69,253 [fEC grant].

4. Details of the impact (indicative maximum 750 words)

The impact of the research strand has consisted in (a) seeding an overall conceptualisation and specific ideas which have been incorporated in the work of forensic phonetic practitioners, (b) provision of reference data which practitioners can make use of, (c) availability of the DyViS database which can be used by them for test and development, and (d) public engagement.

(a) ideas incorporated in the work of forensic practitioners

J P French Associates (JPFA) at York have taken up the challenge, thrown down in Nolan (2005), of using auditory profiling of voice quality in speaker comparison casework. After testing Laver's voice quality framework on a beta version of the DyViS database (see section (b), below), profiling was routinely incorporated in casework (where appropriate) from 2009 to provide an additional set of parameters potentially separating incriminating and suspect speech samples. This brings a degree of systematic analysis to a facet of speaker comparison ('voice quality') which was

previously the domain of rather vague observations [5.1].

The German Bundeskriminalamt (Federal Forensic Laboratory) have introduced, starting in 2008, the use of Long-Term Formant analysis (Nolan & Grigoras 2005) as an additional tool for characterising speakers [5.2]; and JPFA are testing the technique with a view to introducing it into casework [5.1].

(b) provision of reference data which forensic practitioners can make use of

The statistical distributions derived from the 100-speaker DyViS sample population, reported in Hudson et al. (2007), have been widely consulted by forensic practitioners. Previously the pitch characteristics of speech samples under comparison had to be referred to normative data on other languages, notably German. From 2008 J P French Associates [5.1], Martin Barry Forensic Voice Services [5.3], and Duckworth Consultancy [5.4] all report using the DyViS reference data when evaluating differences in pitch between incriminating and suspect speech samples.

(c) availability of the DyViS database

The DyViS database, with its 100 accent-matched speakers performing a stylistic range of speaking tasks, has facilitated development and practice across a wide range of activities related to speech and speaker analysis:

Oxford Wave Research Ltd. have used a subset of it in their development since 2012 of two software products, one ('VOCALISE') which performs speaker recognition by combining more traditional automatic parameters with phonetic variables, and the other ('CLEAVER') which extracts the speech of one speaker from multi-speaker recordings. These are now used by UK and European law enforcement organisations and UK government as well as private companies, including the Metropolitan Police, the German Bundeskriminalamt, and several whose identities cannot be disclosed [5.5].

J P French Associates (JPFA) have used the whole database extensively in testing the performance of an automatic speaker verification system (the Agnitio S.L. 'Batvox' system) on forensically quasi-naturalistic speech from the database. Generally favourable results having been obtained, this system has been used in certain types of case since 2011 to corroborate JPFA's forensic opinion from more traditional forensic phonetic techniques. Second, the database was crucial in JPFA's testing of voice quality profiling prior to its introduction to casework (see (a), above). Only with the availability of a large accent-matched, controlled database was it possible to test the method's discriminative power, and collecting such data would have been beyond the scope of a firm engaged in casework on a day-to-day basis [5.1].

Martin Barry Forensic Voice Services have used the DyViS database to test various technical enhancements for their casework, including software for plotting the formants (resonances) of speech on a psychoacoustic ('Bark') scale which MBFVS began to use from 2012 [5.3].

Martin Duckworth of Duckworth Consultancy has explored the occurrence of different kinds of disfluency in collaboration with Kirsty McDougall at Cambridge. All speakers, not only those with stutters, manifest disfluency in their speech, but testing on the DyViS database has shown that disfluency strategies are sufficiently distinctive to be incorporated in forensic speaker comparison casework at Duckworth Consultancy since 2011 [5.4].

(d) public engagement

Public engagement has taken place with the community of legal and forensic practitioners, and more widely with the public at large through broadcasts.

Nolan has addressed: the Criminal Bar Association (29/11/2008), and the Forensic Human Identification course (Metropolitan Police Training Centre, Hendon 03/2008 & 2009; Academy of Forensic Medical Sciences 06/2011, 03/2012 and 2013); and both McDougall and Nolan were featured with others in a 'BBC Frontiers' edition (12/12/2012) dealing with the possibilities and limitations of speaker recognition [5.6].

Impact case study (REF3b)

5. Sources to corroborate the impact (indicative maximum of 10 references)

[5.1] Letter from Person 1 (Director, JP French Associates)

[5.2] Letter from Person 2 (Director, Bundeskriminalamt (Germany))

[5.3] Letter from Person 3 (Director, Martin Barry Forensic Voice Services)

[5.4] Letter from Person 4 (Director, Duckworth Consultancy Ltd).

[5.5] Letter from Person 5 (Research Director, Oxford Wave Research Ltd).

[5.6] BBC RADIO 4 'BBC Frontiers':

<http://www.bbc.co.uk/programmes/b01p7bxw>