White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/117763/

Article:
French, John Peter (2017) A developmental history of forensic speaker comparison in the UK. English Phonetics. 271 – 286.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

A Developmental History of Forensic Speaker Comparison in the UK

Peter French, University of York Department of Language & Linguistic Science & J P French Associates

peter.french@york.ac.uk

Introduction

The purpose of this article is to provide a review of developments in forensic speaker comparison in the UK from its beginnings in the 1960s up to the present day. The history is documented in respect of milestones and changes in four main areas: methodology, conclusion frameworks, regulation of forensic phoneticians and development of reference databases. The article has been written not only for those directly involved in forensic phonetics, but in a way that I hope will make events in the emergence of the area accessible to phoneticians and linguists more generally.

Forensic Speaker Comparison in the UK

1. Early Days: pre-1990

Forensic speaker comparison (FSC) typically involves the comparison of the voice and speech patterns found in a criminal recording with those found in a recording of a known suspect (Foulkes and French, 2012; French and Stevens, 2013; Jessen, 2008). I am often asked to comment on this as a 'new' and 'minority' area of forensic science. While the frequency with which this type of expert evidence appears in court proceedings is nowhere near that of, say, DNA, fingerprints or tool-mark analysis, it is nevertheless more prevalent than is commonly supposed and also has a much longer pedigree. With regard to frequency, a conservative estimate based on discussions with colleagues is that UK phoneticians are presently consulted in around 500 to 600 criminal investigations and legal cases per year. And the known history of such consultations spans more than 50 years, the first documented occasion of FSC figuring in a criminal trial being in Winchester Magistrates Court in 1965 (Ellis, 1991). The 1970s and early 80s saw no more than a steady trickle of casework, until the introduction of the Police and Criminal Evidence Act (PACE) in 1984. Prior to this, if the police had possession of a criminal recording and they wished to establish the identity of the speaker, a suspect would be asked to provide a voice sample in order for them to have a comparison carried out. Since the provision of the sample was purely voluntary, many suspects being given advice by the legal representatives would refuse to co-operate. PACE, however, required all police interviews with suspects to be audio recorded, and a consequence was that many of those who would otherwise have refused to provide a voice sample found themselves inadvertently doing so as a 'by-product' of the interview procedure. This

resulted in an exponential increase in the volume of FSC through the late 1980s, the time at which I left academic work to establish a practice as a full-time forensic phonetician.

Up to this point, most of the FSC casework had been undertaken on a spare time consultancy basis by a small number of university academics working in phonetics, sociolinguistics and dialectology. Most prominent were Stanley Ellis and Jack Windsor-Lewis at the University of Leeds, John Baldwin at University College London and Peter Wright at the University of Salford. Nan Anthony at Queen Margaret College Edinburgh undertook the work in Scotland. All were willing to act for prosecution or defence and it was not uncommon to find more than one in the same case, each acting for opposing parties. Occasionally, the names of John Laver (University of Edinburgh) and Jim Milroy (University of Sheffield) would surface as experts for the defence, most often providing information about the limitations of the science at the level of principle, rather than an opinion on the identity or otherwise of suspect and criminal voice arising from an analysis of the recordings.

In these early days, those who did undertake examinations of the recorded voices almost always relied exclusively on auditory-phonetic analysis. There was a heavy concentration on vowel and consonant pronunciations, and voice quality was compared purely holistically without recourse to any formal scheme or system of categories to enable its representation in terms of constituent phonatory and vocal tract settings. The depth and detail of the analysis was very low by contemporary standards and the making and keeping of analytic records was scant. The field was without a professional body or organisation, there were no agreed procedures or standards and the work was entirely unregulated. Conclusions concerning speaker identity were generally expressed in binary categorical terms; for example, 'it is my considered professional opinion that the speaker in the criminal recording is (not) Mr Smith'. Indeed, one phonetician (John Baldwin) would state that identity or non-identity was in his opinion 'beyond reasonable doubt'.

2. Developments post-1990

By the end of the 1980s significant changes were afoot; these are described in the Sections below, each of which brings us up to the present day.

2.1 Methodology

In 1989 I convened the first Annual Conference on Forensic Phonetics. This was an international meeting held in York, and was followed by two further such meetings before the formation of the International Association for Forensic Phonetics (IAFP), later - in 2004 - to be renamed the International Association for Forensic Phonetics and Acoustics (IAFPA). This development served to forge and consolidate links between UK practitioners and those from other countries, in particular Germany and the Netherlands where forensic speaker identification was undertaken by phoneticians working in police and government laboratories. At around this time

most UK practitioners began to incorporate acoustic examinations into their hitherto exclusively auditory FSC procedures. Collaboration with European colleagues, in particular with Hermann Künzel, then Head of the Speaker Identification Laboratory at the Bundeskriminalamt in Wiesbaden, who had recently published a full-length book espousing the joint auditory- cum acoustic-phonetic approach (Künzel, 1987), was undoubtedly influential in this regard. Equally important, however, was the emerging availability of computer software packages for acoustic analysis and the publication of an article in *Journal of Linguistics* by Francis Nolan at the University of Cambridge (Nolan, 1991). The article was based around the analysis of recordings from a forensic case and provided not only a theoretical exposition but also a practical substantiation of the shortcomings of auditory analysis alone and the need for complementary acoustic testing.

However, despite its growing use, acoustic testing was not universally adopted, and the England and Wales Court of Criminal Appeal was asked to decide on the safeness of a conviction secured partly on the basis of FSC evidence given by John Baldwin. The evidence was based on auditory analysis only (*R v Robb [1991]*) and the court heard arguments from lawyers representing the appellant concerning the unreliability of the approach. It was not convinced by these and ruled that the conviction was safe. It also affirmed the autonomy of the expert and his right to exercise discretion over the nature of the tests carried out:

> 'Dr Baldwin … was entitled to be regarded as a phonetician well qualified by academic training and practical experience to express an opinion on voice identification. … [His] reliance on the auditory technique must … be regarded as representing a minority view in his profession but he had reasons for his preference and on the facts of this case at least was not shown to be wrong … We accordingly dismiss the appeal …'

In the UK common law system appeal court *dicta* are binding on all lower courts with respect to future cases. In the jurisdiction of England and Wales therefore no court of the first instance could from that point take a decision to reject speaker comparison evidence on the grounds that it lacked an acoustic component.

The principle was again addressed at the appeal court level in 2002, this time in the separate jurisdiction of Northern Ireland[1] in the case of *R v O'Doherty [2003]*. The grounds for the appeal were the same as in R v- Robb, but this time it was successful. The court overturned the conviction weighing the evidence of the experts arguing in favour of acoustic testing against the view of the original prosecution expert who had not used it in her testing and who continued to hold out against it:

[1] For those unfamiliar with legal arrangements, the UK is divided into three separate jurisdictions: England and Wales together, Northern Ireland and Scotland.

> '… we are satisfied, having heard Dr Nolan and Dr French and read the report of Dr Künzel, that in the present state of scientific knowledge no prosecution should be brought in Northern Ireland in which one of the planks is voice identification given by an expert which is solely confined to auditory analysis. There should also be expert evidence of acoustic analysis such as is used by Dr Nolan and Dr French and all but a small percentage of experts in the United Kingdom and by all experts in the rest of Europe, which includes formant analysis.'

From 2002 then, the law relating to the admissibility of FSC evidence in the jurisdiction of Northern Ireland became more stringent than that in England and Wales[2].

This disparity persists to this day, the principle having been re-addressed in the England and Wales Court of Criminal Appeal in the case of *R v Flynn & St John [2008]*. In respect of FSC evidence the judges commented:

> '… we think it neither possible nor desirable to go as far as the Northern Ireland Court of Criminal Appeal in O'Doherty which ruled that auditory analysis evidence given by experts in this field was inadmissible unless supported by expert evidence of acoustic analysis.'

Notwithstanding the admissibility of auditory-phonetic only analysis in England and Wales, nearly all practising analysts today use the joint auditory-acoustic method. The range of features they assess - some auditorily, some acoustically and some by both methods – potentially includes a very wide range of phonetic variables and may take into account higher order linguistic information and non-linguistic features too:

1. Vocal setting and voice quality. Full analysis (using a version of the Laver VPA scheme; Laver 1980, 1994) distinguishes phonation features, overall muscular tension features and vocal tract features, with up to 38 individual elements to be considered.

2. Intonation, potentially including analysis of tone unit nuclei, heads and tails.

3. General pitch, measured as average and variation in fundamental frequency.

4. Articulation rate.

5. Rhythmical features.

6. Connected speech processes such as patterns of assimilation and elision.

[2] The court made 'three exceptions' to this: (a) where the recording involves a closed set of known speakers and the issue is simply who said what, (b) where there are 'rare characteristics which render a speaker identifiable', (c) where the point at issue concerns just the accent or dialect of a speaker rather than his/her individual identity.

7. A large set of consonantal features, including energy loci of fricatives and plosive bursts, durations of nasals, liquids, and fricatives in specific phonological environments, voice onset time of plosives, presence/absence of (pre-)voicing in lenis plosives.

8. A large set of vowel features, including acoustic patterns such as formant configurations, centre frequencies, densities, and bandwidths, and auditory qualities.

9. Higher-level linguistic information including use and patterning of discourse markers, lexical choices, morphological and syntactic variants, pragmatic behaviour such as turn-taking and telephone call opening habits, aspects of multilingual behaviour such as code-switching.

10. Evidence of speech impediment, voice and language pathology.

11. Non-linguistic features characteristic of the speaker, for example patterns of audible breathing, throat-clearing, tongue clicking, and both filled and silent hesitation phenomena.

(Adapted from French, Nolan, Foulkes, Harrison, McDougall, 2010, 146-147)

The most recent deliberation concerning examination techniques from the England and Wales Court of Criminal Appeal relates to the use of automatic speaker recognition (ASR) software. Briefly, and as background for those outside the forensic wing of phonetics, ASR software works by taking a known (suspect) recording, performing mathematical transformations on it and reducing it to a statistical model based on mel frequency cepstrum coefficients (MFCCs). The voice in the criminal recording is also processed in this way and the software compares the two models and produces a measure of the distance (similarity/difference) between them. In evaluating whether the measure indicates that they are likely to have come from the same or different speakers, the criminal speaker model is also compared with a set of models from a reference population of other speakers held within the system. There are obvious advantages to ASR systems including their relatively high level of objectivity and low degree of dependence on the human analyst and the speed at which they operate - large numbers of samples can be processed in a matter of minutes (French and Stevens, 2013).

In many jurisdictions across the world, the incorporation of ASR analysis into the wider battery of auditory and acoustic tests is quite common. In a recent case (R -v- Slade & Ors, [2015] EWCA Crim 71) in which expert speaker comparison evidence had been used at the original trial, the lawyers representing the appellants sought to introduce new and potentially exculpatory ASR evidence into the appeal against their convictions. Philip Harrison and I provided the evidence. The Court stated that while 'it is neither necessary nor appropriate for us to make any definitive ruling in this

case as to whether … [ASR] evidence can ever be admissible' the evidence should not be admitted into the particular proceedings on grounds including the following:

(a) that there were no guarantees that the small set of speakers which the ASR system automatically selected from its larger reference set to assess the typicality or otherwise of the voices was representative of the variation found in the wider population;
(b) that those presenting the ASR evidence had said it should only be used in conjunction with evidence arising from other, auditory- and acoustic-phonetic testing, thereby indicating a lack of confidence in the robustness of the technology;
(c) that a jury faced with statistical probability estimates such as those produced by the system would have no means of interpreting their significance;
(d) that differently calibrated ASR systems can produce more or less conservative estimates of likelihood.

Whilst the ruling does not exclude the possibility of ASR evidence figuring in future trials in England and Wales, it nevertheless creates a significant obstacle to be overcome through argument by those seeking to adduce it. Further, and this is worth stressing as it may not be fully appreciated by forensic phoneticians outside of the UK, in the absence of a local ruling concerning ASR systems, courts in other Commonwealth countries such as Canada, New Zealand and Australia when attempting to decide the issue of ASR admissibility, may look to appeal court precedents from other commonwealth jurisdictions, including the UK. Such rulings are likely to have persuasive force. In this sense, the ruling has a wide ambit of potential influence.

Bringing together the effects of the various Appeal Court rulings then, the present day situation in the UK with regard to legal requirements is very much as it was pre-1990, and may be summarised as follows:

(i) Acoustic analysis not a legal requirement in England and Wales; auditory analysis alone is sufficient, the use of acoustic analysis being at the discretion of the expert. In Northern Ireland, acoustic analysis is a necessary component of the examinations, and this must include an analysis of vowel formants.

(ii) ASR system evidence, while not inadmissible, would need to have its case argued if it were to be introduced into a trial in England and Wales, and potentially in other UK and Commonwealth jurisdictions also.

2.2 Conclusion Frameworks

As mentioned above, until the late 1980s phoneticians involved in FSC most commonly expressed their conclusions concerning identity/non-identity of speakers

categorically, within a binary 'yes'/'no' framework. By the turn of that decade, this practice had been largely abandoned in favour of classical likelihood scales. The scales contained verbal rather than numerical expressions of likelihood in view of the fact that there are no population statistics available for most of the features analysed by forensic phoneticians in the course of their comparisons. So, for example, one might note that the suspect and criminal speaker share a particular, retracted pronunciation of /s/ or a tense larynx vocal setting, but there is no database one might consult to determine the incidence of these features, either individually or in combination, in the wider population.

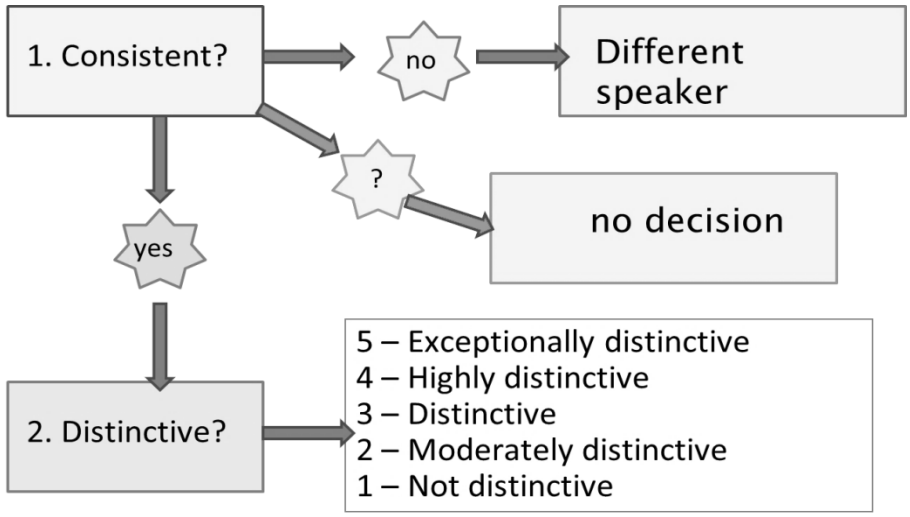
The classical likelihood scales varied somewhat in terms of their detail, but the scale I used was quite typical of those in use more generally:

5	Satisfied X and Y are the same speaker
4	Highly probable X and Y are the same speaker
3	Probable X and Y are the same speaker
2	Fairly probable that X and Y are the same speaker
1	Rather more likely than not that X and Y are the same speaker
0	No balance of probability
1	Rather more likely than not that X and Y are not the same speaker
2	Fairly probable that X and Y are not the same speaker
3	Probable X and Y are not the same speaker
4	Highly probable X and Y are not the same speaker
5	Satisfied X and Y are not the same speaker

This kind of framework had a great advantage over the binary framework, in that it allowed the phonetician to express his or her degree of certainty in the identification or elimination of the suspect. However, it also embodied problems. Specifically, it contained a logical flaw and overstated the weight of evidence in favour of the prosecution. In order to demonstrate this, one might envisage a case where the combination of features common to the suspect and criminal recordings was considered to be very rare in the wider population, perhaps shared by only a handful of other people. However, it could not logically be concluded from this that it was 'highly probable' that the criminal speaker was the suspect. This is because, on the voice evidence alone, the criminal is no more likely to be the suspect than any of the small number of other speakers who also share the characteristics. The statement that identity of criminal and suspect is 'highly probable' is an example of what has come to be termed the 'prosecutor's fallacy' (Thompson & Schumann, 1987).

In 2005 a small group of UK phoneticians based at the Universities of York and Cambridge began discussions about ways of expressing conclusions that avoided this problem. In 2007 an alternative framework was written up and published in the *International Journal of Speech, Language and the Law* (French & Harrison, 2007). This involved the phonetician making a two-part decision on the basis of the outcomes of his/her analysis. The first part entailed deciding whether the speech

samples under consideration were compatible, or *consistent*, with having come from the same speaker: ('yes', 'no' or 'no opinion'). If that decision was positive, then the analyst progressed to a second decision concerning how unusual, or *distinctive*, he/she considered the cluster of features found in the samples to be. Distinctiveness was rated on a 5-point scale, as shown below[3]:



An example of a conclusion produced by the application of this framework is *'the samples are in my view consistent with having been produced by the same speaker and the features common to them are considered highly distinctive'*. This avoids the prosecutor's fallacy, and leaves the decision concerning whether the suspect and criminal are the same person to the triers of fact (jury, judge, magistrate).

That approach to expressing conclusions was widely adopted by UK practitioners. However, over time, a problem in its application began to emerge. In asking as the first part of the decision simply whether the speech samples were consistent with having come from the same person, it precluded any expression of the *degree* of similarity or difference between them[4]. In was largely in view of this that from January 2015 it was supplanted by a verbal likelihood ratio[5] framework in my laboratory and then by UK forensic phoneticians more generally.

This latest framework has the advantages of allowing an assessment of the degree of similarity *and* typicality of the features found in the recordings, and avoiding the prosecutor's fallacy. The latter is avoided by not considering how likely it is that the suspect is the voice in the criminal recording, but instead turning the question around and asking: 'how likely it is that one would find the features present in the criminal recording if the speaker were to have been the suspect as opposed to another user of a similar variety of the language?' (Rose, 2003). The scale presently in use is that recommended across forensic science disciplines by the UK-based Association of Forensic Science Providers (2009). It has 6 levels of verbal

[3] Diagram reproduced by kind permission of Paul Foulkes.
[4] See Rose and Morrison (2009) on the 'cliff edge' effect.
[5] See Rose (2003) for an extended explanation and discussion of likelihood ratios.

probability – positive and negative – each of which is accompanied by a 'support statement' (Champod & Evett, 2000), i.e. a statement of the level of support the conclusion provides for the same or different speaker view.

	Support statement	**Relative probabilities**
↑	extremely strong support	*the possibility that these results could be found under a different speaker hypothesis can effectively be ruled out*
↑	very strong support	*the probability of obtaining these results is very much greater under a same-speaker hypothesis than under a different-speaker hypothesis*
↑	strong support	*the probability of obtaining these results is much greater under a same-speaker hypothesis than under a different-speaker hypothesis*
↑	moderately strong support	*the probability of obtaining these results is greater under a same-speaker hypothesis than under a different-speaker hypothesis*
↑	moderate support	*the probability of obtaining these results is somewhat greater under a same-speaker hypothesis than under a different-speaker hypothesis*
↑	limited support	*the probability of obtaining these results is only slightly greater under a same-speaker hypothesis than under a different-speaker hypothesis*
	inconclusive	*the results do not provide support for either hypothesis*
↓	limited support	*the probability of obtaining these results is only slightly greater under a different-speaker hypothesis than under a same-speaker hypothesis*
↓	moderate support	*the probability of obtaining these results is somewhat greater under a different-speaker hypothesis than under a same-speaker hypothesis*
↓	moderately strong support	*the probability of obtaining these results is greater under a different-speaker hypothesis than under a same-speaker hypothesis*
↓	strong support	*the probability of obtaining these results is much greater under a different-speaker hypothesis than under a same-speaker hypothesis*
↓	very strong support	*the probability of obtaining these results is very much greater under a different-speaker hypothesis than under a same-speaker hypothesis*
↓	extremely strong support	*the possibility that these results could be found under a same-speaker hypothesis can effectively be ruled out*

A strong practical advantage of using this framework is that, because it is widely used across different forensic science areas, the courts are familiar with it (Kirchhübel et al, 2016).

2.3 Regulation

It will be clear from Section 2.1 that, as an area of forensic science, FSC is not subject to any stringent regulation or imposition of standards by the courts. Several

other bodies, however, have attempted to introduce regulation into the UK but until recently attempts have failed and been abandoned.

First, almost from the time of its inception the IAFPA placed on its agenda the implementation of standards, and by the time of its second meeting in 1993 had formulated a Code of Practice (COP) that was to be binding upon its members. It was envisaged that complaints against members would be considered by its Ethics Committee (subsequently renamed the 'Professional Conduct Committee') and infractions against the COP would be subject to sanctions ranging from formal admonition to expulsion from the Association.

While formulated in rather general terms[6], the ethical strictures and recommendations of the COP could potentially have been supplemented with standards concerning analysis protocol and speech laboratory procedures. However, owing to the high costs likely to accrue to IAFPA from any legal action taken against it by a member who was found to be in infraction of the COP and formally sanctioned, the Association in 2010 withdrew the powers of enforcement vested in its Professional Conduct Committee and in 2016 dissolved the Committee altogether. Today, the COP stands only as best practice recommendations. As such, a working group has been formed to review and expand its content.

A further attempt at regulation by IAFPA in 1997 consisted of its forming an Accreditation Committee to develop and implement procedures for the accreditation of members. A process was established consisting of - albeit 'light touch' - proficiency testing. However, being entirely voluntary, accreditation proved unpopular to the extent that only one member ever came forward and the idea was subsequently abandoned.

Regulation by an entirely separate body, the UK Council for the Regulation of Forensic Practitioners (CRFP) was attempted in 2008 – 2009. Under a Government initiative, the CRFP had been established in 2001 with the objective of bringing forensic practitioners of all disciplines under regulation. The general proficiency testing procedure was not dissimilar to that attempted by IAFPA. I was approached in 2008 to take the role of Lead Assessor for Forensic Speech and Audio Analysis and by January 2009 was ready with a team of trained Assessors to begin the examination of candidates. However, within two months CRFP had been dissolved and no accreditations ever took place.

The CRFP was, in fact, dissolved by the Forensic Science Regulator. This post was created by the Home Office in 2008 and carries responsibilities for regulation, ensuring maintenance of standards and quality assurance in forensic science. International Standards Organisation standard ISO 17025, *Competence of Testing and Calibration Laboratories,* was identified by the Regulator as the main vehicle for achieving this. This is an extremely comprehensive set of standards applicable

[6] http://www.iafpa.net/code.htm

across disciplines and relating to all aspects of practice and competence of providers of scientific services. A document entitled *Forensic science providers: codes of practice and conduct* (hereon *Codes*), which adapted ISO 17025 to forensic science generally, was produced by the Regulator's office. It was acknowledged that *Codes* would need further adaptation by the writing of supplementary guidelines and appendices to cover the needs of specific fields of forensic practice. In 2013 I convened a meeting of the UK forensic speech and audio community in London to consider the issue of regulation and it was unanimously agreed by the 22 attendees that forensic speech and audio analysis should be brought under the regulatory framework. Two groups were established to write the necessary documentation tailoring the general standards set out in *Codes* to the specific requirements of these disciplines. The ensuing documents were put out for public consultation and then published in June 2016). The timescale for compliance by forensic phonetics and audio practitioners is October 2017.

This development, following in the wake of failed initiatives by other bodies, is set to have a significant effect on professional practice. In addition to setting standards for practice, it requires that practitioners undergo recurrent testing and that the analysis in every case is checked by a second, appropriately qualified phonetician. While these requirements will not be easy to meet, especially by those working as sole practitioners, the difficulties are not insurmountable and must in any case be considered against the considerable safeguards and quality assurances they bring. Help and guidance for those in sole practice or working in small practices is being provided by the Regulator's office.

2.4 Development of the Science

Forensic disciplines, such as DNA, which have a database and population statistics available, are able to estimate probabilities numerically. Forensic phoneticians are not in this position as, with the exception of a very small number of features such as stammering or average fundamental frequency, no-one has systematically compiled information concerning the incidences of occurrence. Even with such statistics, it would not be an entirely simple matter to progress to numerical probabilities as the extent to which features are correlated with one another is generally unknown and this information would need to be factored in in arriving at numerical likelihoods (Rose, 2006).

To a large extent, the non-availability of information concerning distributions and cross-feature correlations has been brought about by forensic phoneticians themselves. An examination of the web pages of six UK forensic phoneticians provides an indication of the very considerable volume of casework they have undertaken, most of which has involved speaker comparison:

'… worked on around 1100 cases involving evidential speech recordings over a 23 year period'

'… been consulted in an average of approximately 50 cases a years since 1998'

'… over twenty years' experience in criminal and civil cases'

'… worked on over 1800 voice identification cases'

'... undertaken forensic speech analysis work since 1993'

The records of analysis from these cases are potentially an invaluable source of distributional data for the features analysed on all parameters, yet practitioners have not hitherto attempted to extract the data from case records, either ongoingly or retrospectively, and compile databases for use in emerging or future casework. The lack of such information leaves forensic phonetics as a science very much in the position it was in thirty-five years ago when I entered the field.

In order to take a beginning step towards remedying the situation, my laboratory has recently embarked on the extraction exercise. This procedure has only been implemented in 2016, but we are investigating the feasibility of extracting the material from past as well as progressing cases. Features identified for extraction include:

- First and second vowel formant (F1 - F2) data logged vowel phoneme by phoneme and to provide an indication of the use of vowel space overall;
- F1 - F2 values for vocalic elements of hesitation markers;
- Long term average F3 values and distributions calculated from vowel mixtures;
- Consonant data concerning, for example, labialisation and labiodentalisation of /r/, retraction and whistling of /s/, presence of multiple plosive bursts on /p, t, k/;
- Phonatory and supralaryngeal vocal tract settings;
- Articulation rate;
- Fundamental frequency averages, standard deviations and distributions.

The information is being extracted in respect of known speaker (suspect) recordings only and each speaker in the database is tagged with metadata including sex, age, and regional/social/ethnic variety.

My personal view is that in auditory-acoustic FSC cases forensic phoneticians will never be in a position to provide conclusions in the form of fully numerical statements of likelihood. I take this position because the range of features we examine is extremely wide and nearly all features are subject to a great deal of regional, social and ethnic variation, as well as change over time. Even if we had the benefit of unlimited research resources, there would be no possibility of our ever being able to establish distributional information for every analysable feature for every variety one might encounter in FSC casework, and even if this were to be possible, the 'shelf life' of such information would be limited (French, Nolan, Foulkes,

Harrison & McDougall, 2010). This does not, of course, obviate the need for the feature extraction exercise we are engaged in. The resource we currently use to assess whether features found in recordings are unusual or distinctive is casework experience. However, recent research indicates that estimates of the incidence of occurrence of features vary even across experienced experts (Ross, French & Foulkes, 2016). The advantage of having distributional data available for at least some features in some varieties is that it lessens the degree to which the assessment of distinctiveness is dependent upon the individual analyst and moves us further towards the goal of objectivity in interpreting our findings – even if our conclusions have to remain as opinions - in verbal rather than numerical form.

Acknowledgement

I am grateful to Dr Christin Kirchhübel for comments on an earlier draft of this article. Any errors remain my own.

References

Association of Forensic Science Providers (2009) Standards for the formulation of evaluative forensic science expert opinion, *Science and Justice 49*; 161–164.

Champod, C., & Evett, I. W. (2000). Commentary on Broeders, A.P.A. (1999) some observations on the use of probability scales in forensic identification. *Forensic Linguistics 6 (2)*, 228-241.

Ellis, S. (1990) It's getting rather serious ... early speaker identification. In H. Kniffka (ed.), *Texte zu Theorie und Praxis forensicher Linguistik.* Tiibingen: Max Niemeyer.

Forensic Science Regulator (2016) *Codes of Practice and Conduct for forensic science providers and practitioners in the Criminal Justice System, Issue 3.* https://www.gov.uk/government/publications/forensic-science-providers-codes-of-practice-and-conduct-2016

Forensic Science Regulator (2016) *Codes of Practice and Conduct Appendix: Speech and Audio Forensic Services, FSR-C-134, Issue 1.* https://www.gov.uk/government/publications/speech-and-audio-forensic-services

Foulkes, P. & French, P. (2012) Forensic speaker comparison: the linguistic-acoustic perspective. In L. Solan & P. Tiersma (eds.) *Oxford Handbook of Language and Law*. Oxford: Oxford University Press.

French, P. & Harrison, P. (2007) Position statement concerning use of impressionistic likelihood terms in forensic speaker comparison cases. *International Journal of Speech, Language and the Law, 14 (1)*, 139-163.

French, P. & Stevens, L. (2013) Forensic speech science. In M. Jones & R. Knight (eds.) *Bloomsbury Companion to Phonetics*. London: Continuum.

French, P., Nolan, F., Foulkes, P., Harrison, P. & McDougall, K. (2010) The UK position statement on forensic speaker comparison: a rejoinder to Rose and Morrison. *International Journal of Speech Language and the Law, 17 (1)*, 143-152.

Jessen, M. (2008). Forensic Phonetics. *Language and Linguistics Compass, 2(4)*, 671-711.

Kirchhübel, C., Bhagdin, A., French, P., Harrison, P., Rhodes, R. & Wormald, J. (2016) Understanding of forensic speaker comparison conclusions in the legal setting. Presented at *International Association for Forensic Phonetics and Acoustics conference*. University of York, 24th – 27th July.

Kiinzel, H. J. (1987) *Sprechererkennung*. Heidelberg: Kriminalistik Verlag.

Neocleous, T., Napier, G., Aitken, C., Foulkes, P., French, P., Gold, E., Harrison, P., Hughes, V. & Stevens, L. (2016) Models for forensic speaker comparison. *Presented at the Royal Statistical Society conference*, University of Manchester. 6th-8th September 2016.

F. Nolan (1991) Forensic phonetics. *Journal of Linguistics 27*, 483-493.

Police and Criminal Evidence Act (1984) (PACE) codes of practice. https://www.gov.uk/guidance/police-and-criminal-evidence-act-1984-pace-codes-of-practice

Rose, P. (2006) Accounting for Correlation in Linguistic-Acoustic Likelihood Ratio-based Forensic Speaker Discrimination, *Speaker and Language Recognition Workshop, 2006. IEEE Odyssey*.1–8, 28th - 30th June.

Rose, P., & Morrison, G. (2009). A response to the UK Position Statement on forensic speaker comparison. *International Journal of Speech Language and the Law, 16 (1)*, 139-163.

Ross, S., French, P. & Foulkes, P. (2016) UK practitioners' estimates of the distribution of speech variants. Presented at *International Association for Forensic Phonetics and Acoustics conference*. University of York, 24th – 27th July.

Thompson, W.C. & Schumann E.L. (1987) Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defence attorney's fallacy. *Law and Human Behaviour 11(3)*, 167-187.

Law Reports

R -v-Flynn & St John [2008] Cr App R 20 (E & W).

R v O'Doherty [2003] Cr App R 5 (NI).

R v Robb [1991] Cr App R 161 (E & W).

R v Slade & Ors [2015] Cr App R 71 (E & W).