

**Position Statement Concerning Use of Impressionistic Likelihood Terms
in Forensic Speaker Comparison Cases^{1, 2}**

(Published 2nd February 2007)

1. Introduction

We begin from the observation that speech scientists are in agreement that there may be people who cannot be distinguished from another speaker or speakers by the analysis of their voice and speech patterns. This is so even in the ideal circumstance, i.e. when one has available large amounts of speech in high quality recordings. The possibility of discriminating between like-sounding individuals may be vastly diminished with smaller samples and poorer quality recordings of the kinds often encountered in forensic casework.

It is at least partly in recognition of this that forensic phoneticians, in carrying out speaker comparison work, have avoided making categorical statements of identification and have instead used likelihood or probability statements. These are often drawn from formalised impressionistic scales and include such terms as 'very likely/fairly probably/more likely than not to be the same speaker'.

Until recently, many of the signatories to this document, as practising forensic phoneticians, used this way of expressing our conclusions. More recently, however, we have become aware of a serious logical flaw in these kinds of formulations. They give a false weighting to the expert evidence.

2. The Prosecutor's Fallacy

The awareness initially came from a consideration of a Court of Criminal Appeal ruling in the DNA case R -v- Doheny & Adams ([1996] EWCA Crim 728). At the court of first instance the DNA evidence had been presented using a form of reasoning which has been described as the 'prosecutor's fallacy'.³ An illustration of this type of reasoning is as follows:

- (i) A stain containing DNA is found at a crime scene. It is estimated that only one person in a million will have a DNA profile that matches that of the stain.
- (ii) The defendant has a DNA profile that matches the stain.
- (iii) Ergo, there is a million to one probability in favour of the defendant having left the stain.

The reasoning at (iii) is incorrect. The correct interpretation of the facts is, of course, that in the UK population of 60 million, where it is estimated that one person in a million will have a matching profile, there will be, potentially, 60 people whose DNA profile matches the stain. On the basis of the DNA evidence alone there is a 1 in 60 chance that the DNA came from the defendant, not a million to one chance in favour. It is equally possible that the stain came from any one of the 59 others.

The Doheny & Adams ruling determined that DNA evidence should avoid the prosecutor's fallacy and be presented within a logically correct framework. When so presented, it falls upon the jury to assess its significance in the light of the other evidence in the case. If, for example, there were no evidence other than that the defendant was in the UK at the time of the crime, then little or no significance could be accorded to it. However, if there were, say, evidence of motive, evidence of a connection between the defendant and the victim, and the defendant was seen somewhere near the scene at the relevant time, then the fact that only 59 others in the UK are likely to share the profile becomes highly significant.

3. Applicability to Forensic Speaker Comparisons

The estimation that 1 person in a million will share the DNA profile is known as its 'random occurrence ratio'. Phoneticians can calculate the random occurrence ratio for very few features of speech. Exceptions are fundamental frequency (a measure of voice pitch), articulation rate (speed of speaking) and

stammering. However, for the overwhelming majority of voice and speech features examined in casework, it is simply not known how widely they are distributed in the population. It is because forensic phoneticians are unable to provide numerical statements of probability that they have used impressionistic probability terms of the types described under 1 above. However, by using these, the speech expert gives the same false weighting to the evidence that DNA scientists did prior to Doheny & Adams. Thus, the speech scientist falls foul of the prosecutor's fallacy.

Let us say, purely arbitrarily and for the sake of explanation, that a combination of phonetic features found in a criminal recording and in a known recording of the suspect were to be shared by a further 10 people, then one could not justifiably conclude that the voice in the criminal recording was 'likely' to be that of the suspect. On the phonetic evidence alone, it would be no more likely to be that of the suspect than any one of the other 10.

Arguably, in the light of the Doheny & Adams authority, to say that is 'likely' to be the suspect is inadmissible in law.

4. New Framework For Expressing Conclusions

Certain of the signatories to this document have developed an alternative conceptual framework for expressing conclusions, which is as follows:

4.1 Consistency

First, a decision is made concerning whether the known and questioned samples are compatible, or consistent, with having been produced by the same speaker. In considering consistency one would assess the degree to which observable features were similar or different. In judging the samples one would regard differences as weighing against compatibility unless they can be explained by models of acoustic, phonetic or linguistic variation (e.g. by reference to differential channel characteristics, sociolinguistic, psychological and/or physical factors). The presence of substantive non-explained differences between the samples would be taken as grounds for

judging them to be inconsistent with having been produced by the same speaker. In some cases it may not be possible to reach a decision on consistency. The outcome of this first decision process would then be:

Consistency Outcomes: Consistent
 Not consistent
 No decision

4.2 Consistent Outcome – Distinctiveness

The decision that two samples are consistent with having been spoken by the same person does not necessarily entail that they were. Unless the cluster of features leading to the consistency decision is distinctive or unusual, they will be shared by a substantial number of other people in the population. Therefore it is necessary to proceed to evaluate the distinctiveness of the features common to both sets of samples. A convenient way of assigning an overall rating to the distinctiveness of the common features is to locate them on a five-point scale, as follows:

5. Exceptionally distinctive – the possibility of this combination of features being shared by other speakers is considered to be remote

4. Highly distinctive

3. Distinctive

2. Moderately distinctive

1. Not distinctive

4.3 Not Consistent Outcome

Where the samples are not consistent we see no logical flaw in making the statement that the samples are spoken by different speakers. This may be stated with a degree of confidence appropriate to the exigencies of the data.

5. An Exception: Closed Set Comparisons

In most speaker comparison cases the work is carried out on assumption that the questioned speaker could be anyone within a potentially large and open-ended population of speakers. In a minority of cases, however, there is independent evidence (e.g. video surveillance) to show that a closed set of known speakers was present and participating in the conversation. In such cases the comparison task becomes an issue of who said what. In these circumstances, if the voices are sufficiently distinct from one another, we consider it justified to make categorical statements of identification.

6. Likelihood Ratio Approaches

Finally, we accept in principle the desirability of considering the task of speaker comparison in a likelihood ratio (including Bayesian) conceptual framework. However, we consider the lack of demographic data along with the problems of defining relevant reference populations as grounds for precluding the quantitative application of this type of approach in the present context.

In view of these difficulties, the framework we endorse is the one set out under 4 above.

Working Group Authors

Prof Peter French

JP French Associates and Dept of
Language & Linguistic Science, University
of York

Louisa Cawley

JP French Associates and Dept of
Language & Linguistic Science, University
of York

Dr Gea de Jong	Department of Linguistics, University of Cambridge
Martin Duckworth	Dept of Speech & Language Therapy College of St Mark & St John, Plymouth
Dr Paul Foulkes	Dept of Language & Linguistic Science, University of York and JP French Associates
Philip Harrison	JP French Associates and Dept of Language & Linguistic Science, University of York
Toby Hudson	Department of Linguistics, University of Cambridge
Dr Kirsty McDougall	Department of Linguistics, University of Cambridge
Prof Francis Nolan	Department of Linguistics, University of Cambridge

Co-Signatories

Martin Barry	School of Languages, Linguistics and Cultures, University of Manchester
Dr Zoe Butterfint	Department of Speech and Language Therapy, University of East Anglia
Dr Frantz Clermont	JP French Associates

Prof John Harris	University College London & Forensic Research Associates
Allen Hirson	Dept of Language & Communication Science, City University, London
Dr Frederika Holmes	Independent forensic phonetician and City University, London
Prof David Howard	Dept of Electronics, University of York
Dr Ghada Khattab	Department of Speech & Language Sciences, University of Newcastle
Dr Geoff Lindsey	Forensic Research Associates
Prof Michael K C MacMahon	Department of English Language, University of Glasgow
Elizabeth McClelland	Independent forensic phonetician, Forensic Analysis of Speech and Language Samples
Dr Jane Setter	Department of Applied Linguistics, University of Reading
Marion Shirt	Formerly Department of Linguistics & Phonetics, University of Leeds
Dr Linda Shockey	School of Psychology and Clinical Language Studies, University of Reading
Dr Kate Storey-Whyte	AudioLex

Notes

1. The main points concerning the applicability of the prosecutor's fallacy argument to expert evidence on speaker comparisons were set out in a presentation given to the International Association for Forensic Phonetics & Acoustics Annual Conference, Marrakech, Morocco, 3rd – 6th August 2005, JP French '*Forensic Speaker Identification Evidence and the Prosecutor's Fallacy*'. The points are also summarised in JP French & P Harrison (2006) 'Investigative and evidential applications of forensic speech science' Ch 14 of A. Heaton-Armstrong, E. Shepherd, G. Gudjonsson and D. Wolchover (eds.) *Witness Testimony: Psychological, Investigative and Evidential Perspectives*. Oxford: Oxford University Press.

2. It will be apparent from the arguments developed here that the term 'forensic speaker identification' should be replaced by 'forensic speaker comparison'.

3. 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.