

WestminsterResearch

<http://www.westminster.ac.uk/westminsterresearch>

**Quis custodiet ipsos custodes? Should Justice Beware: A
Review of Voice Identification Evidence in Light of Advances in
Biometric Voice Identification Technology**

Singh, C.

This is a copy of the final version of an article published in the International Commentary on Evidence, 11 (1), p. 1–28.

It is available from the publisher at:

<https://dx.doi.org/10.1515/ice-2014-0009>

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (<http://westminsterresearch.wmin.ac.uk/>).

In case of abuse or copyright appearing without permission e-mail repository@westminster.ac.uk

Charanjit Singh*

***Quis custodiet ipsos custodes?* Should Justice Beware: A Review of Voice Identification Evidence in Light of Advances in Biometric Voice Identification Technology**

Abstract: The debate relating to the quality of voice identification evidence in the United Kingdom continues against the backdrop of advances being made in the use of biometric voice identification evidence (BIVE) and the technology (BVIT). Anecdotal evidence shows that BVIE is being adduced in criminal prosecutions across the United Kingdom (UK) predominantly in cases involving terror crimes. This also suggests that the courts are willing to accept BVIE as being reliable even though experts in the fields of phonetics and law disagree as to its veracity. The argument against admission rests on the lack of sophistication in the traditional ear-witness voice identification methods of acoustic and auditory analysis (AAA), and now biometrics because of its infancy. Experts therefore argue that scientific reliability should be demanded from such evidence if it is to be used for criminal prosecutions and this not currently achieved. Therefore, a number of issues arise as a result of this. For example, the potential erosion of civil rights, the legal implications that relate to obtaining and using mixed biometric voice identification evidence (MBVIE) – this is the evidence of an ear-witness verified using BVIT. Related to this is the notion that the jury and lawyers need to be educated on how such evidence should be received and used. Presently, there insufficient guidance on where the UKs courts should draw the line in admitting potentially hazardous evidence such as this. Exactly when BVIE becomes unreliable in a legal and scientific sense remains unclear. This significantly contributes to the debate surrounding the codification of evidence law and the introduction of a reliability test, along the lines of that used in other jurisdictions including the United States of America, to mitigate the risks that lie in admitting unreliable evidence. The purpose of this article is to contrast ear-witness and BVIE by exploring the contemporary debates that surround their admission and the notional extent to which BVIT is being used to police the UK. Furthermore, to review whether the advances made in BVIT can contribute to the reliability of the evidence by reducing error rates and false-positive identification.

*Corresponding author: Dr. Charanjit Singh, Head of Subject and Research, School of Law – University of West London, United Kingdom, E-mail: Doctor.CSingh@gmail.com

Keywords: voice identification, risk, biometrics, reliability, evidence

DOI 10.1515/ice-2014-0009

1 Introduction

Crime control is an information hungry activity that is preoccupied with prevention measures; the result is to rewrite the identity of the population. The level of surveillance that was once reserved exclusively for major criminality has arguably been extended to cover the vast majority of societies that are being reinscribed as problematized.¹ There is no doubt that technology has had a profound impact on the creation, prevention and sophistication of criminal activity across the globe. In 2013, the UK's Chief Inspector of Constabulary, Tom Winsor, stated that the primary function of the police is the prevention of crime and not catching those responsible for its commission after its perpetration.² It is difficult to imagine a complete strategy that could achieve this without the use of covert surveillance. It is fair to acknowledge that increases in the perception of risk or threat for instance through terrorism, when coupled with risk-based crime control strategies has significantly contributed towards the development of surveillance techniques, security-society,³ political policing⁴

1 Lyon, D. (2001). *Surveillance Society: Monitoring Everyday Life*. Philadelphia, PA: Open University Press.

2 Travis, A. (2013). *Police Should Focus more on Preventing Crime than Catching Criminals*. Website: Theguardian.com. Available at: <http://www.theguardian.com/uk/2013/apr/29/police-focus-crime-catching-criminals> (Accessed 22 March 2014).

3 Zedner, L. (2009). *Security*. London: Routledge. See also: Deflem, M. (Ed.). (2004). *Terrorism and Counter-Terrorism: Criminological Perspectives*. Amsterdam: Elsevier. See also: Sharpe, C. (2002). Covert Surveillance and the Use of Informants. In: McConville, M. and Wilson, G. (Eds.). *The Handbook of Criminal Justice Process*. Oxford: OUP.

4 Brodeur, J.P. (2002). Undercover Policing in Canada: Wanting What is Wrong. *Crime, Law and Social Change* 18 (10–12): 105–306. In: Loftus, B. and Goold, B. (2012). Covert Surveillance and the invisibilities of Policing. *Criminology and Criminal Justice* 12: 275. Note: Knowledge of crime aids political decision making, measurement of activity and choices made – the Foucauldian notion of discursive regimes can be applied here. In relation to truth and knowledge see: Foucault, M. (2003). Power/Knowledge. In Delanty, G. and Strydom, P. (Eds.). *Philosophies of Social Science: The Classic and Contemporary Readings*. UK: Open University Press.

and the regulatory frameworks within which they can operate (see Police Act 1997 and the Regulation of Investigatory Powers Act 2000 (RIPA)).⁵

The English law of evidence has not escaped unscathed by all of this with new forms of evidence have been consistently presented to the courts. Evidence from covert surveillance is often obtained from sophisticated and technologically advanced identification techniques. This raises important questions relating to admissibility, reliability and risk. One such question focuses on the fact that research is required on how to mitigate perception emphasis; this is undue emphasis that juries often place on expert evidence that they perceive to be scientific and are therefore unchallengeable. The risk from perception-emphasis is exaggerated by the fact that the “scientific” nature of particular evidence or analysis and therefore its reliability continues to evolve.

All being said this is a common problem in research methods and it relates to the notion that scientific objectivity cannot be questioned.⁶ This is exacerbated by the role that evidence ultimately plays in the course of any litigation. The ideal would be to ensure that the jury understands the definition of “science” and “scientific” when they are presented with expert evidence. This would be the reference to a reasoned conclusion that is reached through the systematic study of something using organized principles that are rigorously tested. Indeed there are a number of consequences to this for instance what extent of testing would amount to being “rigorous” and should the testing itself be interdisciplinary so as to facilitate triangulation. It would be fair to state that both AAA and BVIE are far from settled sciences.

BVIE is often perceived as a novel form of evidence, a brief perusal of evidence admitted in the courts over the last ten years highlights how the courts respond to such advents. For example there has been a steady rise in communications such as email, posts and tweets from individual public Facebook and Twitter social media or micro-blogging accounts being presented as admissible evidence that relates to a matter in issue and/or being adduced for the purposes of credibility in the UK courts. *R v Agyei and Others*⁷ provides a good example. The case involved an alleged sexual assault and the evidence included tweets from the 14-year-old victim’s personal account on Twitter. The defendants were

5 For a contrasting view see: Bhatt, H. (2006). RIPA 2000: A Human Rights Examination. *International Journal of Human Rights* 10(3): 285–314.

6 Singh Landa, C. (2010). Capturing Humanity: The Implicated Biography and Validity in a Scientific World. 1:35–45. See also: Einstein, A. (1934). *Mein Weltbild*, p. 168. English translation by Harris, A. (1935). *The World as I see It*. pp. 124–125. Originally found in Popper, K. (2010). *The Logic of Scientific Discovery*. Oxon: Routledge Publishing, p. 9 note six. Detail: this latter text was originally published as *Logik der Forschung* (1935).

7 12 September 2013, Unreported.

on remand in custody; during the trial the defence discovered a number of tweets made by the victim that undermined the allegation. These were adduced as evidence by the defence and resulted in the acquittal of a number of the defendants.

There is a distinct lack of statistical data⁸ on the use of such evidence in the prosecution of criminal offences across the UK. However, some meaningful conclusions can still be drawn from sources such as case reports. Although it seems that cases where such evidence was used have had good success rates, they also highlight issues including reliability that are faced by the courts when dealing with this evidence (see *R v Blackshaw and Others*⁹ also known as the social network or Facebook cases).

Similarly voice identification evidence (VIE) presents itself in a number of notable forms. The most basic of which is the evidence of an ear-witness who makes a voice identification of a person they heard speak or utter something. This can be something that they heard within the vicinity but outside of their direct sight or over a device for instance a telephone, mobile, radio, tannoy or as an audio recording for instance a recorded message released like those that are frequently released by terrorist organisations. The identification may be made because the witness is familiar with the person whose voice they have identified. In the majority of cases, a recording will have been obtained by the police in a procedure undertaken in accordance with the Police and Criminal Evidence Act 1984 Codes of Practice (see *R v Deenik*¹⁰) or as part of a covert investigation authorised under the Regulation of Investigatory Powers Act 2000 (RIPA).

Where the voice appears as digital data perhaps in the form of audio recordings of telephone calls or messages in voicemail then potentially a biometric identification can be made, issues in relation to this are discussed later. In such cases, those charged with investigating the offence have two options; use AAA or BVIT to identify the speaker. AAA is mostly undertaken by phoneticians and is heavily criticized where it is used for the prosecution of a criminal offence, which is discussed later. BVIE is still a relatively new form of evidence for the English courts. This raises a number of salient issues relating to

8 A request for statistical data relating to the frequency in use of such evidence as part of a criminal prosecution was made to the Crown Prosecution Service on the 25th of September 2013. As at that date it was confirmed that the service does not centrally hold this information. However, the policy on its use can be downloaded from its website. Available at: http://www.cps.gov.uk/legal/a_to_c/communications_sent_via_social_media/ For the purposes of research, a number of meaningful conclusions can still be drawn from anecdotal evidence such as case reports and academic comment.

9 (2011) EWCA Criminal 2312.

10 (1992) *Criminal Law Review* 578 CA.

admissibility, reliability¹¹ and risk. It has been over a decade since the last comprehensive review of voice identification and the technological advances that have been made provide impetus for further research. Therefore, the aim of this article is to provide an overview of the rules on voice identification evidence in the UK and highlight the need for contemporary review.

2 Past research on voice identification by ear-witnesses

In 1974 the media furor relating to the number of criminal cases that involved unjust convictions based upon mistaken witness evidence resulted in the UK government setting up an official inquiry that aimed to shed light on the possible reasons for mistakes being made by even the most honest of witnesses. Lord Devlin chaired the committee that was charged with the task of producing a report on the matter. The findings were published in the “Report to the Secretary of State for the Home Department of the Departmental Committee on Evidence of Identification in Criminal Cases”, now widely referred to as the Devlin Report (1976).¹² The committee found that there was a distinct lack of research on voice identification, something that is true today in relation to BVIE. The committee called for an urgent response (research) in to establishing a procedure for testing a witness’s capacity to recognize a voice, the practicality of voice identification parades and other “appropriate” methods, generally seeking to promote the credibility and reliability of VIE (see generally pp. 67–69 and specifically p. 68 at paragraph 4.5).¹³

Before the Devlin Report there were a number of authorities in English law that show the pivotal role often played by VIE in prosecutions. In *R v James Hanratty*¹⁴ VIE highlighted the distinctive manner in which an accused spoke, this was used to prove it was he who committed the crime (see also *R v Knight*¹⁵; *R v Lewis*¹⁶).

11 Clifford, B.R. (1980). Voice Identification by Human Listeners: On Earwitness Reliability. *Law and Human Behaviour* 4:373–394.

12 Great B., Devlin, P. and Great B. (1976). *Report to the Secretary of State for the Home Department of the Departmental Committee on Evidence of Identification in Criminal Cases*. London: Her Majesty’s Stationery Office.

13 See: Philippon, A.C., Cherryman, J., Vrij, A. and Bull, R. (2008). Why is My Voice so Easily Recognized in Identity Parades? Influence of First Impressions on Voice Identification. *Psychiatry, Psychology and Law* 15:70–77.

14 (1962) *Criminal Law Review* 409.

15 CA, 14 March 1969, unreported.

16 CA, 14 March 1972, unreported.

In *Ear-witness Voice Recognition Accuracy*¹⁷ Professors Ray Bull and Brian Clifford recommended that prosecutions based solely on VIE, where the accused was unknown to the witness, should not proceed. Bull and Clifford argued that both ear-witness and eyewitness evidence is prone to errors made by even an honest witness. Therefore, such evidence should only act to support the primary evidence against an accused and that it should not be the sole or primary evidence upon which the prosecution is based. A basic perusal of the contemporary debate surrounding the reliability of witness evidence only serves to highlight that recognition of the voice is clearly an activity that is beyond the capability of the majority of jurors and witnesses.

David Ormerod argues that voice identification evidence has demonstrated to pose a greater risk when made by a non-expert witness¹⁸ and he identified the following five forms of analysis/comparison:

- aural comparison of voice data by an expert (auditory analysis)
- acoustic or spectrographic analysis by computer software which is then interpreted by an expert
- witness listens to the recordings and makes an identification
- the jury compare the voice data (recording) with that of the accused who is testifying and
- the jury compares voice data (recordings of both the offender and the accused).¹⁹

Although it is discussed later it should be noted that by reason of the court's decision in *O'Doherty* the first two of these categories are amalgamated. In addition, it is submitted that the following category should now be added to this list:

- automated biometric voice identification evidence (BVIE).²⁰

Note: This process involves the complete autonomous production and comparison of voice data by specialist computer software.

17 Bull, R. and Clifford, B.R. (1984). *Ear-Witness Voice Recognition Accuracy*. In Wells, G.L. and Loftus, E.F. (Eds.). *Eyewitness Testimony: Psychological Perspectives*. Cambridge: Cambridge University Press, pp. 92–123.

18 Ormerod, D. (2001). Sounds Familiar. *Criminal Law Review* 565 and Redmayne, M. (2001). *Expert Evidence and Criminal Justice*. Oxford: Oxford University Press, p. 110.

19 Ormerod, D. (2002). Sounding out Expert Voice Identification. *Criminal Law Review* 771: 772.

20 This process involves the complete autonomous production and comparison of voice data by specialist computer software.

3 Voice identification by experts

I doubt few would disagree that where expert evidence is concerned, there still exist a number of issues relating admissibility, reliability and risk. Each of these is exacerbated where the evidence concerned is voice identification.²¹ Whilst the English courts have cultivated important but liberal rules that relate to “experts” and the admission of their evidence²² they did so to avoid limitation to the categories of evidence that may be presented²³ to them. These minimum requirements do seek to provide key safeguards in terms of admission, reliability and risk. However, decisions such as that in *O’Doherty* continue to highlight the risk for miscarriages of justice where the prosecution case rests on voice identification. These minima clearly fall short of the gate-keeping role that the Law Commission suggests is required in its proposal for reforming the use of expert evidence in criminal proceedings.²⁴

Further still few would disagree with the sentiment that the courts should continue to remain open and receptive to new forms of evidence being adduced even where the risk associated with that is grave.²⁵ Instead, we should seek to mitigate these risks. Currently, UK law does not provide for the satisfaction of a reliability test when it comes to the admission of expert evidence. Lord Chief Justice Auld has proposed the introduction of such a test in his report on the codification of UK evidence law *Expert Evidence in Criminal Proceedings in England and Wales*.²⁶ The proposed test would result in the introduction of a

21 Roberts, A. (2008). Drawing on Expertise: Legal Decision Making and the Reception of Expert Evidence. *Criminal Law Review* 443.

22 See Singh, C. (2014). *Beginning Evidence*. Oxford: Routledge, especially chapter 10 and *R v Turner* (1975) QB 834. Note, there is a lack of clarity in the law that surrounds the reliability of the evidence given by an *ad hoc expert* – perhaps a police officer who has been monitoring a particular person and comparing voice samples for a prolonged period of time. This matter is noted for the purposes of this paper. See also: *R v Chenia* (2002) EWCA Crim 2345.

23 Note the comments of L.J. Steyn in *R v Clarke* (1995) 2 Cr App Rep 425.

24 Law Commission. (2011). Expert Evidence in Criminal Proceedings in England and Wales. Law Commission. No 325 (Accessed 20 September 2013). See also: Law Commission. (2009). The Admissibility of Expert Evidence in Criminal Proceedings in England and Wales: A New Approach to the Determination of Evidentiary Reliability. Law Commission. No 190 (Accessed 20 September 2013).

25 See *R v O’Doherty (Anthony)* (2002) N.I. 263.

26 Law Commission. (2011). Expert Evidence in Criminal Proceedings in England and Wales. Law Commission Report No. 325 (Accessed 20 September 2013). See also: Law Commission. (2009). The Admissibility of Expert Evidence in Criminal Proceedings in England and Wales: A New Approach to the Determination of Evidentiary Reliability. Law Commission Consultation Paper No. 190 (Accessed 20 September 2013).

framework designed to allow expert evidence to be challenged through the proper evaluation²⁷ of evidentiary reliability as best practice that relates to the admissibility of impartial and objective evidence.

Both Ormerod²⁸ and Edmond²⁹ make a valid point when they conclude that the introduction of a reliability test is likely to result in the exclusion of a substantial amount of expert evidence that is currently admitted. In this respect, Auld's report also concedes that at present far too much unreliable expert opinion evidence is being readily admitted in the English Courts. The National Research Council (NRC) of the United States of America (USA) also commented on the issue of testing expert evidence prior to its admission. In *The Age of Expert Testimony: Science in the Courtroom* the NRC highlighted that "... legal disputes [concerning evidence] are resolved by [testing the evidence through] cross-examination. Courts are not designed [nor in the business of determining] whether scientific conclusions [or the expert's opinions] are correct [that is a task for the discipline and its regulation]".³⁰ This approach relies on trial lawyers highlighting the strengths and deficiencies of the evidence; perhaps this supports the notion that trial lawyers still require further education in this respect.

It has long been a tradition that the English courts acknowledge those matters they can judicially note because they are firmly within their competence and experience, and those that are clearly outside of their competence and experience for example foreign law. It is in terms of the later that the court would seek the assistance of experts to inform itself.³¹ The rationale for this is straightforward: an expert is someone whose competence, experience and expertise are beyond those of a normal witness. However, the courts have not set out a requirement for experts to be formally qualified for example the expert need not have a PhD in the particular field within which they are held out to be an expert. To do so would inevitably limit the pool of experts from which the court may source the relevant evidence. Thus,

27 An evaluation which seeks to determine whether the expert can objectively justify his or her conclusions would include consideration of, amongst other things, the quality and extent of the data, the rationale that underpins the drawing of a particular inference, precision and degree of uncertainty, peer review, the completeness of the information that led to the inference being drawn and whether the methods followed establish practices within the field. See *ibid* note 25: provision 4, Schedule part 1.

28 Ormerod, D. (2002). Sounding out Expert Voice Identification. *Criminal Law Review* 771: 774.

29 Edmond, G. (2012). Is Reliability Sufficient? The Law Commission and Expert Evidence in International and Interdisciplinary Perspective (Part 1). *Evidence and Proof* 16: 30–65.

30 Kennedy, D. (Ed.). (2002). *The Age of Expert Testimony: Science in the Courtroom*, Report of a Workshop. In: *Science, Technology and Law Panel*. Washington, DC: National Research Council, p. 23 (Accessed 15 October 2013).

31 The rules that must be followed in so doing were summarized in *Folkes v Chadd* (1782) 3 Doug KB 157 and see also *R v Bonython* (1984) 38 SASR 45.

someone who has worked within a particular field repeatedly or over a period of time may be deemed to be “qualified” enough to be classed as an expert. In criminal cases expert evidence is limited by the ultimate issue rule for that is something for the jury to decide. The issue is the extent to which the court should scrutinize expert evidence to ascertain its reliability.

Issues of reliability with VIE are confounded by the fact that expert evidence by its nature, like the assumptions generated by deoxyribonucleic acid (DNA) evidence,³² is perceived as having an almost impermeable scientific seal in terms of its quality. The risk in miscarriages of injustice occurring is increased when such a fraught position is adopted. This is especially true in relation to VIE of ear-witnesses, which has a greater chance of being incorrect. Examples of this include cases where a witness makes a false-positive identification, false-negative identification or in terms of experts where the techniques that are used to analyse the data are either inconsistent or unsettled.

Traditionally, phoneticians or other voice identification experts typically use auditory or acoustic analysis when identifying a voice. The problems faced with both of these are well documented in a number of authorities including the seminal judgments in *R v O’Doherty (Anthony)*³³ and *R v Robb (Robert McCheyne)*.³⁴ Auditory analysis comprises of the comparative analysis of two audio data sets or voice recordings, this is usually the voice of the alleged offender perhaps gathered through covert surveillance and the person later apprehended for the commission of the crime.³⁵ The intention is to identify the similarities or differences that exist between them.

In *R v Robb*³⁶ the English court accepted this method as producing reliable voice identification. The decision sat in stark contrast to the approach adopted in other jurisdictions. The method was criticized for its lack of the objectivity that was imbued within other scientific approaches.³⁷ Analysis of this sort offers

32 For an example see: *R v Cooke* (1995) *Criminal Law Review* 39 where expert evidence was adduced to prove that DNA evidence obtained from a strand of the defendant’s hair matched the semen specimen taken from the victim and that the chance of the semen specimen having come from another individual was 1:73 million. See also: *R v Adams* (1996) 2, Cr App R 467

33 (2002). N.I. 263.

34 (1991). 93 Cr App R 161.

35 The Islamic State (IS) terrorist “Jihadi John” is a good example.

36 See note 33.

37 *R v Robb* (1991) 93 Cr App R 161 and *R v Turner* ([1975] QB 834 at page 841. See also: Law Commission (2011). Expert Evidence in Criminal Proceedings in England and Wales. Law Commission No 325, p. 63 and Rule 33.2 specifically 33.2(1) of the Criminal Procedure Rules 2013. In *Robb* Lord Justice Bingham made the point that auditory analysis alone was unreliable where it was not supplemented and verified by acoustic analysis (at page 165).

little more than a subjective expert assessment of the voice presented – issues relating to the quality of the voice record itself are discussed later. Although, this method has been criticized it is fair to state that the admission of a voice identification made using expert auditory analysis carried less risk than a comparable voice identification made by a non-expert ear-witness relying on memory function alone.

The decision in *R v O'Doherty*³⁸ (an Irish case) has resulted in a change in the approach by the courts. The court acknowledged some of the criticisms levied at auditory analysis by other jurisdictions and held that the method was inadequate where it was used without acoustic analysis save in three circumstances. Therefore, auditory analysis will continue to be accepted as good evidence where it relates the accent or dialect of the speaker or where the voice has rare characteristics and for the purposes of deciphering the author of the comments (who said what) in recordings that contain multiple speakers.³⁹ In all other instances auditory analysis must now be accompanied by acoustic analysis. This comprises of an expert computing the same two audio recordings through spectrograph software. Once done visual images of the two voices are created in the form of a map or linear graphic chart. The expert then analyses these images to see if they belong to the same person. What we have here is in effect part automation of voice analysis and you will note that the expert still plays a pivotal role. Neither of the two methods is equivalent to biometric analysis because it would be impossible to conduct them without the instruction of an expert.

The current approach to expert VIE can be described as triangulated analysis. Here researchers, usually social scientists in fields such as criminology, test their findings by using more than one method of analysis, this helps establish reliability. The assumption generated is that where more than one method results in the same outcome the finding is more reliable. Often inconsistencies will be revealed but even these can be used to arrive at alternative explanations (Patton, 2002).⁴⁰ The approach is criticized because researchers can fall into the trap of seeking to reinforce their original findings (bias) or use two equally unreliable methods of analysis to produce a result that is untrustworthy but that will be perceived to have greater scientific quality. Therefore, I would

³⁸ *Supra* note 33

³⁹ See: Orchard, T.L. and Yarmey, A.D. (1995). The Effects of Whispers, Voice-Sample Duration, and Voice Distinctiveness on Criminal Speaker Identification. *Applied Cognitive Psychology* 9: 249–260.

⁴⁰ Patton, M.Q. (2002). *Qualitative Research and Evaluation Methods*. Thousand Oaks, CA: Sage.

suggest that the reliability in expert VIE is still a salient issue and the courts and juries need to remain vigilant when presented with expert VIE.⁴¹ The idea that expert VIE is somehow of greater scientific reliability, because two methods of analysis produced the same result, needs to be challenged.

This change in approach does not affect the safety of convictions where expert VIE used the auditory analysis alone – of course appeals are a different matter altogether. There are two points that should be noted: s.23(1)(c) of the Criminal Appeals Act 1968 (a) allows the Court of Appeal to hear evidence not previously adduced at trial and (b) a different response to be put forward in relation to evidence that was originally presented. Therefore, where expert VIE based on auditory analysis forms part of the case then it is open for the lawyers to make an *O'Doherty* argument to undermine its reliability.

The fact that it has taken the courts over almost twenty-five years post-*Robb* to adopt a fresh approach can be attributed to the fact that it has taken voice identification some time to develop as a practical discipline. It still has some way to go before it is readily accepted as being a “scientific” discipline. Biometrics adds another interesting and contemporary dimension to this debate. The commercial sector has made some major advances in the development and adoption of biometric voice and speaker identification systems. Research and development in this area is now the biggest it has ever been given the importance of security in governance and society. Criminal justice systems are notoriously slow at harnessing the benefits that technology can bring, something that can be explained by the limitations on governmental budgets, the time and cost related to development of procedures and the technology, and cautious courts.

4 Cases and research

Professor Daniel Yarmey published the last substantive review on the research relating to VIE.⁴² In *Ear-witness Speaker Identification*⁴³ he summarized the findings of thirty-four publications. Yarmey came to the conclusion that a misnomer existed; eyewitness memory and voice recognition was simply common knowledge. He also found that the courts were most concerned with ear-witness voice identification of complete strangers. Thus, it was recommended

⁴¹ It should be noted that bogus experts exaggerate this problem see: *R v Director of Serious Fraud Office, ex parte Smith* (1993) AC 1, as per Lord Mustil.

⁴² Note: this did not concern BVIE.

⁴³ Yarmey, A.D. (1995). *Earwitness Speaker Identification. Psychology, Public Policy, and Law* 1: 792–816.

that a positive voice identification of a stranger made by an ear-witness should be treated with great care.

Issues regarding voice identification arose again in *R v Roberts*,⁴⁴ this time the court referred to research conducted by Bull and Clifford. They had concluded that the identification of the voice was in fact an arduous task for the untrained ear and recommended that the trial judge warn the jury in quite stringent terms – even more so than they would normally do in relation to eye-witness identification evidence.⁴⁵ In this case the court of appeal acknowledged that the following issues existed in terms of VIE:

- identification of the voice is more difficult than visual identification;
- the identification of a strangers voice is more even if the opportunity to hear their voice is relatively good;
- the identification of a voice is more likely to be incorrect than a visual identification by an eyewitness;
- lay people are willing to rely equally on both ear and eyewitness evidence; and
- ear-witness evidence should not form the primary evidential basis for a prosecution without other supporting evidence.

In *R v O'Doherty (Anthony)*⁴⁶ the court in Northern Ireland set out an interesting rationale that approved the use of acoustic analysis and qualified the use of auditory analysis in cases involving voice identification.⁴⁷ In 2003 the UK Home Office issued an advice on the use of voice identification parades,⁴⁸ the use of which is not compulsory and therefore a failure to undertake this does not necessarily affect the admissibility of VIE.

It should be noted that the courts in *Roberts* and *O'Doherty* were not considering BVIE. The Home Office advice did not include guidance in relation to this form of VIE either. It is therefore submitted that these authorities and this advice are of limited application to cases where the primary basis of the prosecution rests upon BVIE. It would be fair to conclude on the basis of a

44 (2000) *Criminal Law Review* 183.

45 See *R v Turnbull* (1977) QB 224. See also: Van Wallendael, L.R., Surace, A., Parson, D.H., and Brown, M. (1994). Earwitness Voice Recognition: Factors Affecting Accuracy and Impact on Jurors. *Applied Cognitive Psychology* 8: 661–677.

46 (2002) N.I. 263.

47 Note the departure from the original approach in *R v Robb (Robert McCheyne)* (1991) 93 Cr App R 161.

48 Home Office circular 057/2003: *Advice on the use of voice identification parades*. London: Her Majesty's Stationery Office. Available at: <http://www.homeoffice.gov.uk/about-us/corporate-publications-strategy/home-office-circulars/circulars-2003/057-2003/> (Accessed 17 March 2013).

fictive hierarchy that BVIE is by far the most reliable form of voice identification evidence that can be presented in court.

5 Analysis of the voice

Although, BVIE may be considered to be a relatively recent activity the analysis and representation of speech in visual form is not entirely new. Melville Bell, father of the notable Alexander Graham Bell, began analysing speech circa 1867. Bell Telephone Laboratories began creating the earliest known versions of what is now commonly referred to as the spectrograph,⁴⁹ which is used as part of *acoustic* analysis. During these early stages of development the majority of the research focusing on voice analysis related to issues concerning national security (war). It then went on to focus on law enforcement – a notion that had sought impetus through lobbyists and political rhetoric concerning the management of a risk-society.

BVIE, in essence, is the product of the complete automation of voice analysis. This is still constrained by the shifting nature of the variables specific to each and every voice for example the vocal range may change by reason of age.⁵⁰ This risk is mitigated by the fact that the technology has developed at an astounding rate, it is now far more sophisticated in terms of the extent to which variations in the voice become detrimental to biometric voice identification. The result is to produce a greater level of reliability – the question is whether this is good enough to satisfy the purposes of proof.

However, in comparison to the standard of ear-witness evidence currently being adduced in the UK criminal courts this seems to be a vast improvement in terms of credibility, trustworthiness and weight. To give context to this the current process can be stated as, where ear-witnesses or experts are used, human-digital-human. The ear-witness makes identification from a recording or does so in a live setting at the scene, the station or at court, or a phonetician produces a visual image of the voice recording(s) using spectrographic software that they then use in a comparative analysis. The latter and BVIE both require some pre-recorded samples from which a comparison can be made; this need is significantly reduced where ear-witness identification is concerned. If we assume that pre-recorded samples are available perhaps via covert surveillance

49 See: Grey, G. and Kopp, G.A. (1944). Voiceprint Identification. Bell Telephone Laboratories Report: 1–14 and Kopp, G.A. and Green, H.C. (1946). Basic Phonetic Principles of Visible Speech. *Journal of the Acoustical Society of America* 18: 74–89.

50 See: Yarmey, A.D. (1991). Descriptions of Distinctive and Non-Distinctive Voices over Time. *Journal of the Forensic Science Society* 31: 421–428.

or a database⁵¹ then the process could realistically become almost completely automated (digital-digital) in the next 2–5 years. One obvious benefit of this is to further mitigate the potential risk that is created by human input (subjectivity) because the requirement for the phonetician and/or ear-witness is removed and replaced by specialized computer software.

The development of this technology has been driven by private finance. Large multinational organizations' are seeking to harness its security benefits and long-term financial gain from licensing it to governmental defence ministries across the world.⁵² Speaker identification is a result of this.⁵³ This technology has had a revolutionizing effect on consumer behaviour, commerce, education, health, transport and television. For instance current figures suggest that over seventy million cars use an adapted form of the technology to communicate with their owners.⁵⁴ Nuance, a Boston based Technology Company, commissioned Harris International to conduct a study in USA on the extent to which consumers were engaging with voice related systems. It found that approximately 41% of American consumers have used voice *recognition* to bank, shop online or travel.⁵⁵ In terms of law and law enforcement the technology has crept into the administrative side of legal practice. Its reach into criminal justice for the purposes of detection, prevention and prosecution of major criminality has been far slower even though RIPA 2000 has been effective for a number of years in the UK.⁵⁶

51 Note there are a number of legal issues relating to the collection and retention of personal data that would need to be settled, some of these are discussed later.

52 For example anecdotal evidence suggests that the American National Security Agency (NSA) may via surveillance drones (remotely piloted aircraft) be using biometric voice identification technology at over 33,000 ft. This is in addition to other tactics such as geo-location which tracks mobile telephone devices etcetera. I sought information from both the NSA and the Ministry of Defence in the UK in relation to this; both neither confirmed nor denied the use of this technology as part of their defence armoury.

53 Redpath, B. (2013, 25 October). Telephone Interview.

54 Munford, M. (2013). Voice Recognition Will Smash its Way into the Black Mirror. *The Telegraph* 13 February, 2013 (Accessed 11 September 2013).

55 Schmidt, R. (2003). Identity Confirmed, Access Permitted: The Basics on Voice Authentication Security and Consumer Use of an Emerging Biometric. *BiometriTech* 3 September 2003 (Accessed 3 May 2013).

56 What is disturbing is that covert surveillance under the 2000 Act has been misused, the Investigatory Powers Tribunal confirmed that local authorities were using the statute to detect the commission of misbehavior and minor or petty criminal offences for which this regulatory framework and technology was certainly not meant. This stance limits the use of biometric voice identification to serious criminality; the use of covert surveillance for these other purposes has already seen some challenge, see *Paton and others v Poole Borough Council* (2010) IPT/09/01/C. Also see note: 78.

6 Biometric voice identification

The American Subcommittee on Biometrics, which falls under the auspices of the American Committee on Homeland and National Security, defines voice recognition as a biometric modality in which an individual's voice is used for identification purposes.⁵⁷ This is different from speech recognition, which focuses on the words or commands uttered rather than the biometric data imbued within them. The term biometric is utilized to denote the measure of a physical personal characteristic for instance a fingerprint, handwriting and hand geometry, iris and facial or retinal scan (see Myers 2004⁵⁸ for a general discussion on biometrics). Speech recognition is used in commerce for instance to manage and direct calls, the majority of the time this does not use biometric data.

Speech recognition and biometric voice identification have developed concurrently this is partly due the similarity in the characteristics and challenges associated with both. Unlike speech recognition, biometric data is used as a security measure where confirmation of an individual's unique identity is required for example in e-passports. A biometric trait is primarily unique to each and every one of us. It is this uniqueness of our biometric characteristics that make them so desirable for identification purposes because unlike key-fobs and passwords they are inherently difficult to forge, forget and lose.

Voice is also classified as a biometric characteristic – identification is made by measuring speech rhythm, pitch and tone. Whilst the voice is unique in contrast to other biometrics it often changes and varies according to the circumstances. The premise that underpins voice identification, as a biometric therefore needs to account for such changes, which BVIT has made major advances in relation to. The distinctiveness of the voice is measured in terms of differences in pitch, tone and volume. Myers (2004) highlights that there are a number of factors that contribute to the dissimilarity between voices including the shape and size of the mouth, nose, teeth and throat (articulators), and the shape and tension in the vocal chords. The vocalization process namely the manner in which the jaw, lip and tongue muscles are activated makes the voice further exclusive to the individual. Further, unique speech or sound is created when air passes from the lungs into the throat and through the vocal chords and articulators. Thus, the probability that two people have exactly the same set of

⁵⁷ United States of America. Federal Bureau of Investigation. *Speaker Recognition*. Biometric Centre of Excellence. Available at: http://www.fbi.gov/about-us/cjis/fingerprints_biometrics/biometric-center-of-excellence/files/speaker-recognition.pdf (Accessed 12 March 2014).

⁵⁸ Myers, L. (2004). *An Exploration of Voice Biometrics*. Sans Institute.

characteristics is negligible, a fact that undermines legal challenge by an opposing party.

A spectrograph (an electronic device) is utilized to record a spectrogram; this is a digital image of the dataset entered (voice). In basic terms, the vocal pattern produced by the process of speech is visually represented through a process known as voice printing (spectrographic analysis) the pitch (speed and frequency at which the vocal chords vibrate) and volume.

Biometric voice recognition is made possible once a digital model of an individual's voice, referred to as a voiceprint, is stored. This voiceprint serves as a record from which other voice data can be compared. The voiceprint must be of the highest quality and legitimately verified as belonging to the individual for whom the record relates. Critics of voice printing argue that this data (the voice recording) is inaccurate and unreliable because it is affected by a number of unpredictable variables including age, background noise, disguise, fabrication,⁵⁹ happiness, ill health, mood and stress.⁶⁰ Therefore, there is an obvious need for the continual updating of voiceprints. In addition, my discussion with industry stakeholders (experts and technicians etcetera) revealed that there exists disagreement about the extent of subjectivity in the findings made by the various software packages that are deployed to undertake this task. A lack of research in relation to such variation and absence of regulatory standards only complicates matters.

Voice recognition, whether biometric or otherwise, is also affected by the quality of the recording; a poor quality recording may result from particular equipment being used (analogue microphone/telephone)⁶¹ but modern technology deploying audio enhancement techniques has learnt to cope with this. This stands in contrast to the phonetician who primarily relies on the visual image. Although this point requires some further research the contention here is that specialized software has become sophisticated enough to may make biometric voice identification in circumstances where a phonetician would not be able to do so.

There are a number of advantages and disadvantages in pursuing *automated* voice identification for the purposes of criminal justice. The most obvious

59 Note: authenticity can be determined by analysing discontinuity, editing or tampering. It is postulated that this should be compulsory where the evidence is being tendered at a criminal trial or in civil litigation.

60 See: Rafferty, L. (1999). Anything You Say Can and Will Be Used Against You: Spectrographic Evidence in Criminal Cases. *American Criminal Law Review* 36: 291.

61 Note that there is specialist software that can “clean” an audio data sample improving its audible quality for instance programmes such as Adobe Audition CC or Groove Mechanic.

is the argument that the process *can be* relatively low cost with little investment in new hardware for instance telephony or microphones and that a small digital data storage capacity will suffice. Further, identification can be made from any location in the world and in real time usually in less than one second. Ideally this should translate into major savings in terms of time and costs including human resource. In terms of proof the identification is arguably more reliable than those made by ear-witnesses, the probability of unreliability is clearer which allows the evidence to be more accurately measured.

On the converse, it would be rather naïve to accept that new hardware is never going to be needed for the processing of such data because the standard personal computer should be able to perform this task. The requirement of sophisticated computer software for analysis stands in opposition to this notion. Therefore, the software and eventually the hardware will need to be upgraded probably every two-to-three year's or perhaps even more quickly given the speed by which criminality and technology advances.

There is an obvious benefit to being able to make identification in real time; however this is constrained by issues relating to cross-jurisdictional co-operation which will need to be in place if the perpetrators are to be brought to justice. National interests often present a barrier in terms of the effectiveness of such activities – it is one thing to identify and locate a criminal⁶² for instance by tracking their satellite telephones and quite another to physically apprehend them. The length that the USA went to in *Operation Neptune Spear* or the *Abbottabad Operation* to capture Osama Bin Laden demonstrates. It should be noted that this operation was purported to have been conducted without the Pakistani authorities having any knowledge of it.

The human voice varies for a number of reasons. For instance a voice recording of a sixteen-year-old suspect, putting aside the legal issues that surround the legitimate storage of this data, may be different to their voice when they are aged thirty. Furthermore, the voice may be purposefully disguised so as to conceal the identity of the speaker. Although, current research by the North Carolina State University has made some inroads into this⁶³ some still argue that the software is not yet sufficiently sophisticated enough to allow for either age-related variation or purposeful disguise. Again, further research needs to be undertaken on the frequency and impact of these issues on the reliability of the voice identification.

62 Zabel, R.B. and Benjamin Junior, J.J. (2008). *In Pursuit of Justice: Prosecuting Terrorism Cases in the Federal Courts*. United States of America: Human Rights First.

63 Fränti, P., Kinnunen, T., Mohammadi, H.R.S., Rodman, R. and Saeidi, R. *Joint Frame and Gaussian Selection for Text Independent Speaker Verification*. Paper presented at the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (Accessed 25 September 2013).

Additionally, recorded samples suffer from rapid digital deterioration when stored over a long period of time; sudden failure can occur when faulty reading or writing devices are used or when they are exposed to heat, humidity and perhaps other airborne contaminants.⁶⁴ Although, this issue can be minimized by the use of durable or persistent media such as gold-CD/DVD the same cannot be said of internal computer storage methods. Regardless of the cost factors involved in the constant back-up and even re-storage as technology becomes obsolete the deterioration of a voice sample still poses a risk, however small, because it may lead to an erroneous identification and this justifies caution. Neither is it viable to store multiple versions of software and hardware to ensure usability.

7 Biometric voice identification as evidence

Over a decade has passed since the decision in *O'Doherty* with speaker and biometric voice identification technology developing at an astounding rate. Using biometric technology in voice identification is a fairly new phenomenon when compared to other criminal justice measures seeking to make the admission of evidence fairer. Expert evidence will remain a valuable source of evidence in litigation but advances in BVIT where coupled with the in-sourcing of the activity of biometric voice identification could lead to a reduction in the role of expert phoneticians.⁶⁵ There is of course an entire debate that surrounds the distortion of existing policing priorities and the privatization of activities such as covert surveillance, and the general perception that the public have of covert surveillance that too without voice capture.⁶⁶ Currently it is only private companies that provide both the hard and soft “ware” for this activity and this is

64 Hedstrom, M. and Montgomery, S. (1998). *Digital Preservation Needs and Requirements in RLG Member Institutions*. Mountain View, CA: RLG (Accessed 6 February 2014).

65 Note, automation of voice identification may still require experts for example as a form of independent verification required to prevent potential abuse where a sample has been “cleaned”.

66 Ericson, R. (1994). The Royal Commission on Criminal Justice Surveillance. In McConville, M. and Bridges, L. (Eds.). *Criminal Justice in Crisis*. UK: Edward Elgar, pp. 113–140. Note the requirements for the conduct of such activity by the Regulation of Investigatory Powers Act 2000. In the United States of America the public seems more willing to accept covert surveillance, perhaps by legitimation through the notion that there is an increased risk, as a justified form of controlling serious crime for example terrorism, perhaps as fail-safe risk reduction technologies relating to such activity or the federal government’s initiatives. See also: Ericsson, R.V. and Haggerty, K.D. (1997). *Policing the Risk Society*. UK: OUP, pp. 52–54, and Barnard-Wills, D. and Wells, H. (2012). Surveillance, Technology and the Everyday. *Criminology and Criminal Justice* 12: 227. Note: the American National Institute of Standards and Technology (NIST) carry out evaluations on biometrics to help promote its development and growth. See also: ANSI

unlikely to change in the near future simply because of the level of investment required. What is clear or perhaps more widely visible is the fact that an increasing number of governments are undertaking covert surveillance as an integral part of their criminal justice strategy.

Advances in the complexity and sophistication of society raise issues that require new and more innovative ways in addressing crime; digital-age cyber-crime is a good example of this. The vast majority of academic debate and research that surrounds the fallibility of witness evidence relates to eyewitnesses. Notable psychologists including Bull, Clifford, Loftus and Yarmey have greatly contributed to our understanding of how age, anxiety, fear, imagery, (mis)leading information and questioning all affect such evidence (see also: Loftus and Palmer, 1974⁶⁷; Yuille and Cutshall, 1986⁶⁸). There is much less research on ear-witness evidence and even less on biometric voice identification.

The use of this technology also raises serious concerns in relation to the type of state power being exercised⁶⁹ and civil liberties; the Latin maxim *quis custodiet ipsos custodes*⁷⁰ highlights one of the most salient. Intelligence-led crime control activity must be proportionate or balanced and justified. For instance policing must be undertaken with civil rights such as the right to privacy⁷¹ where covert surveillance is utilised being protected by the state. This protection should cover the regulation of the retention and storage of biometric voice samples.

Good biometric software that is text-independent⁷² can spectrographically analyse the human voice far more quickly and with ease⁷³ when compared to the

INCITS 398-2005: Information Technology – Common Biometric. Exchange Formats Framework (CBEFF). These are common standards that deal with biometric data element descriptors.

67 Loftus, E.F. and Palmer, J.C. (1974). Reconstruction of Auto-Mobile Destruction: An Example of the Interaction between Language and Memory. *Journal of Verbal Learning and Verbal Behaviour* 13: 585–589.

68 Yuille, J.C. and Cutshall, J.L. (1986). A Case Study of Eyewitness Memory of a Crime. *Journal of Applied Psychology* 71: 291–301.

69 See page 30.

70 The translation of this is as follows: *who is watching the watchman?*

71 Daniels, R., Macklem, P. and Roach, K. (2001). *The Security of Freedom: Essays on Canada's Anti-Terrorism Bill*. Canada: University of Toronto Press.

72 This is where the programme has no previous knowledge about the speaker's phrasing; therefore it is more flexible. This approach is appropriate for covert surveillance or where an individual refuses to cooperate.

73 Redpath, B. (2013, 25 October). Telephone interview. Brian Redpath is the Enterprise Sales Director (Government) at Nuance Communications. Text-dependent systems require passwords and the like to be entered using the telephone keypad, whereas text-independent systems operate autonomously without the need for any physical human interaction. It is the latter that are of most relevance.

phonetician or other appropriate expert. The process is undertaken in real-time⁷⁴ using a Gaussian Mixture Model⁷⁵ (GMM) or Hidden Markov Model (HMM). This means that it can, almost instantaneously, identify the speaker and their contemporaneous location. In contrast, text-dependent systems involve a longer process because of the requirement to already have collected data from which a comparison can be made. Both the GMM and HMM are in layman's terms statistical modeling for time sequenced data. Text-independent recognition systems bring together biometric characteristics such as modulation and intonation with basic semantics i.e. pronunciation and idiolects. Biometric voice identification systems have developed to the level where variables such as gender and language rarely pose an issue. In fact it would be fair to postulate that properties such as dialect are helpful because they lend identifiable characteristics to a voice.

Aside from obvious evidential and methodological issues such as transparency in level of variation that is acceptable for positive voice identification, there are clear benefits in the using this technology. There is potential for a greater degree of objectivity in the automated analysis of data, this should result in a lower false-positive-identification error rate and less chance of general human error (as discussed). There are opportunities for costs reductions for the courts, police and prosecution services. Collaterally, this is also likely to lead increases in the confidence the public have in the criminal justice system. All of which are endeavours worth pursuing but not at an unprecedented cost.

Currently, there are a variety of terms that are used to refer to the process of biometric voice identification including the preferred live or real-time voice printing. Both text-independent and text-dependent systems require the availability of known data with which to analyse a sample whether that is collected before or after some analysis, thus the two systems would need to be used in conjunction as and when appropriate with relevant guidance and legal support frameworks. The recording and storage of some voice data other than the sample collected is a highly contentious point but an unavoidable requirement if the

⁷⁴ The caveat to the time taken in the "entire process" taking into account the fact that some voice data must exist in the first place for this process to be successful is discussed supra at pages 23–24.

⁷⁵ Reynolds, D.A. and Rose, R.C. (1995, January). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE Transactions on Speech Audio Processing* 3 (1):72–83. See also: Saeidi, R., Kinnunen, T., Mohammadi, H.R.S., Rodman, R. and Fränti, P. (2010). Joint Frame and Gaussian Selection for Text Independent Speaker Verification. *IEEE*: 4530–4533.

processes are to be successful. Clear reconciliation must be reached in relation to the collection and the continued retention and storage of such data.⁷⁶

The software works by recording and analysing that which is spoken so as to capture the frequencies associated with that individual's voice. After which a number of statistical models are used to establish the speech patterns associated with that data and therefore the individual concerned. Experts suggest that the result from this is the production of information that can be equally classed as "biometric" because it is as accurate as a fingerprint or iris scan with the added benefit that it surpasses physicality (having the person present in front of you), time and geographical space. Protagonists of the technology argue that suspects can be monitored in the world at large.

The fact that known samples are required at some stage poses what seems at first instance to be more significant a problem than it actually is. The dynamic collection of samples over a period of time is required before text-dependent analysis can be undertaken. This process can prove far lengthier even though the actual identification may only take seconds. The issue is somewhat mitigated where both text-dependent and independent analysis is undertaken. Furthermore, this may seem positive where criminal justice is concerned because supporting evidence for instance a period of surveillance may lend greater credibility or trustworthiness to the results should they incriminate an accused. Boyle⁷⁷ reports that BVIT has already been used to foil a Mumbai-style terror threat against France, Germany and the United Kingdom where it was used to analyse a discussion between Islamic terrorists who spoke of the spree in India that killed 166 people.

Voice identification evidence has been used in the prosecution of some major known criminals⁷⁸ for many years. In 2007 the United States Drug

76 For an example of the likely issues note that the ECJ struck down the Data Retention Directive 2006/24/EC (this amended Directive 2002/58/EC) in *Digital Rights Ireland Ltd (C-293/12) v Minister for Communications, Marine and Natural Resources, Minister for Justice, Equality and Law Reform, Commissioner of the Garda Síochána, Ireland, The Attorney General, and Kärntner Landesregierung (C-594/12)*, Michael Seitlinger, Christof Tsochold and others. In response to which the UK parliament enacted the Data Retention and Investigatory Powers Act 2014, this will be reviewed in July 2016. In addition, the UK will be creating a Privacy and Civil Liberties Oversight Board. Although note that this does not affect the activities of Government Communications Headquarters (GCHQ).

77 Boyle, R. Biometric Voice-Identification Technology Leads to Arrests of Terrorist Suspects in Europe. Available at: <http://www.popsoci.com/technology/article/2010-10/voiceprinting-technology-helps-thwart-european-terror-attacks> (Accessed 17 March 2013).

78 There is evidence to suggest that the use of covert surveillance is growing – however there is a distinct lack of statistical data that help determine what, if any, percent of this used biometric voice identification. See generally: Loftus, B. and Goold, B. (2012). Covert Surveillance and the Invisibilities of Policing. *Criminology and Criminal Justice* 12: 275.

Enforcement Agency used voice identification technology to identify the voice, from a recording made by Colombian authorities, of the Colombian drug kingpin Juan Carlos Ramirez Abadia. The latter had used plastic surgery and multiple aliases' to prevent being physically detected but it was his voice that gave him away. The USA is also known to have used voice-printing methods in particular activities in Iraq. Putting aside arguments relating to political barriers to cooperation there is a potential opportunity for an international Interpol voice-database to which all 188 of Interpol's member countries could have access. This could result in real-time identification and the contemporaneous capture of criminals, and the prevention of potentially devastating crimes. If it were to succeed this could be one of the greatest information sharing criminal justice initiatives ever taken. This may not be so far-fetched an idea considering the breadth of the UK's own Regulation of Investigatory Powers Act 2000 (RIPA).

Biometric voice identification has steadily proliferated criminal justice. Brian Redpath of Nuance Communications confirmed that over 20 countries friendly with the USA are now using the *Nuance Identifier* product for the purposes of criminal justice related voice identification.⁷⁹ What was interesting is that Redpath stated that the product had become an integral part of those countries defense, intelligence and policing systems. Interestingly, it was confirmed that the technology is advanced enough to accurately identify an individual at a particular place at any given time.⁸⁰ This may be of use to organisations such as the probation service and there may be obvious advantages in terms of penalty enforcement.

RIPA and the Code of Practice issued pursuant to its s.71 regulate covert surveillance. The law relates to “[public authorities] monitoring, observing or listening to persons, their movements, conversations or other activities and communications”.⁸¹ I surveyed over 45 UK police forces and only 1%⁸² admitted to using voice identification technology or any software in their geographical area.⁸³

⁷⁹ As at October 2013.

⁸⁰ Redpath, B. (2013, 25 October). Telephone interview. It should be noted that the United Kingdom has a non-disclosure agreement with Nuance Communications and therefore details as to the extent that this technology and such products are being used in the United Kingdom for purposes of policing is unknown.

⁸¹ *Covert Surveillance and Property Interference: Revised Codes of Practice*. 2010. HMSO, see p. 7. Note: RIPA has been criticized because it is being operationalized for the control of less serious crime rather than the exceptional crimes for which it was intended.

⁸² I collected this data over a period of five months (October 2013–February 2014) and it is therefore current at the time of writing.

⁸³ Specifically the Metropolitan Police Service and Northern Ireland Police both declined to confirm or deny whether it was using any such software in policing. Additionally, those forces

This stands in stark contrast to criminal justice agencies in other jurisdictions for instance the Federal Bureau of Investigation's (FBI) Biometric Centre for Excellence in the United States of America promotes the cultivation of this technology.⁸⁴ There is a continuing debate relating to the balance between preserving democracy and civil liberties in the UK. This may shed some light on why the English police may have been slow to openly adopt BVIT as a tool to deal with contemporary threats and issues in policing. Anecdotal evidence suggests that national governments are more likely to use this technology and/or software with a justified legitimate aim of protecting national security. Recent case studies show surveillance technology is being routinely used for purposes including the protection of national interests; the latter is an issue outside the scope of this current article.

*R v O'Doherty*⁸⁵ fundamentally changed how current forensic speech analysts should compare voice samples for the purposes of tendering them as evidence of identification in the United Kingdom. The court held on successful appeal that "... as a general rule where one of the issues was voice identification, it was necessary for there to be expert evidence of acoustic analysis as well as of auditory analysis. The exceptions to the general rule were: (1) where the voices being listened to were of a known group and the issue was which voice had spoken which words; (2) where the speaker was identifiable because of rare characteristics, and (3) where the issue concerned the accent or dialect of the speaker. Where the prosecution relied on voice recognition and the jury had heard the accused giving evidence, it was appropriate for the jury to be allowed to listen to a tape recording on which the recognition had been based. A specific warning should be given to the jurors relating to the danger of relying on their untrained ears. In the instant case, there was a reasonable possibility that the voice of the speaker who had made the 999 call was not that of O and therefore the convictions were unsafe".⁸⁶ It is postulated that this would be redundant were a biometric voice identification made.

that did admit using "voice recognition software" stated that they did so for purely administrative rather than criminal justice i.e. policing purposes.

84 *NIST Speaker Recognition Evaluations* 2012. 25 April 2005, National Institute of Standards and Technology, Speaker and Language Recognition Projects. Available at: <http://www.nist.gov/itl/iad/mig/spkr-lang.cfm> (Accessed 25 February 2014).

85 (2003) 1 Cr App R 5.

86 Omerod, D. (2002, October). Sounding out Expert Voice Identification. *Criminal Law Review*: 771–790; Barsby, C. and Ormerod, D. (2002, September). Voice Identification: Expert Voice Identification Evidence at Trial Linking Appellant with Crime Based Largely on Auditory Analysis. *Criminal Law Review*: 761–762; O'Brian, W. (2003). Court Scrutiny of Expert Evidence: Recent Decisions Highlight the Tensions. *Evidence & Proof* 7(3): 172–184.

8 Ear-witness voice identification evidence vs. biometric voice identification evidence

It is clear that a holistic cost and benefit analysis in relation to ear-witness evidence is undertaken to prevent a single factor determining its short-term criminal justice future. Protagonists argue that clean voice data samples (clear recording without background or other noise) produce an error rate of under 0.5%⁸⁷ and that issues only arise when the samples are *dirty*⁸⁸ making it difficult to decipher the elements of the voice under scrutiny. The stakeholders in this field have differing views. Experts using spectrographic analysis contend that the current biometric software is not sophisticated enough to clean samples that are dirty. Proponents of the technology that cleans such recordings argue that clean samples, which are then analysed using traditional methods, are far more reliable as evidence. There is a lack of research on this specific point for instance the extent to which BVIT cleans the data sample before analysing it and the effect this has on the recording. This gives some impetus to developing the necessary statutory regulation, which should also determine how issues such as error rates affect reliability and the potential risk associated with that.⁸⁹

It should be noted that the low error rate that is currently accorded to biometric voice identification does suggest that this evidence is more *reliable* than identification made by an ear-witness undertaking an identification procedure. This is a salient point because it begins to address the concerns highlighted by the Court of Appeal, the automated matching of two recordings is more reliable because the likelihood of errors occurring is increased with human input is required. This is another point that desperately requires further research. The cost of using biometrics includes start-up investment, human resource, data storage and currency of the hardware, software and data. This

87 Snyder Sachs, J. *Graphing the Voice of Terror*. Popular Science Magazine. Available at: <http://www.popsci.com/scitech/article/2003-02/graphing-voice-terror?single-page-view=true> (Accessed 17 March 2013). Note there are other software programmes such as Voice Grid Nation being used in the United States with varied error rates, see also: *NIST Speaker Recognition Evaluations* 2012. 25 April 2005, National Institute of Standards and Technology, Speaker and Language Recognition Projects. Available at: <http://www.nist.gov/itl/iad/mig/spkr-lang.cfm>. For details on Voice Grid Nation see: <http://speechpro.com/product/biometric/voicegridnation> (Accessed 25 February 2014).

88 Dirty samples are those recordings where background noise interference is present for instance a voice recording with music or traffic playing in the back or foreground, which obscures the voice.

89 See notes 25 and 29.

should not deter the authorities from using advanced technology in everyday criminal justice namely the detection, prevention and prosecution of crime rather than reserving it exclusively for more serious and organised criminality. In any regard, biometric evidence is open to be challenged in court through cross-examination but this task is made more difficult in the absence of procedure, process and rules. The risk is the manifestation of evidence gathered inconsistently with clear rules on fair process being adduced in court with possible exclusion in the interests of justice not being enough to prevent this. There are of course many other issues that any guidance should cover but which are beyond the scope of this article for instance the presumption of scientific reliability that is often attributed to expert evidence.

9 Voice identification, ear-witnesses and the right to remain silent, privilege and the right to privacy (Article 8)

Before concluding this review there are a number of collateral legal issues that also require mention; the right to silence, the privilege against self-incrimination and the right to privacy. Ear-witness voice identification presents some novel problems in relation to an accused's common law right to remain silent, which is based on the premise of fairness and the privilege that is afforded to an accused against incriminating his or herself. It is only possible for an ear-witness to attempt identification where they can compare the voice of the offender with that of the accused. This requires a recorded voice sample to be available and therefore the accused must unequivocally consent to participate in any identification procedure as set out in Code D of the Police and Criminal Evidence Act 1984. The code states that whilst it '... concentrates on visual identification procedures, it does not preclude the police making use of aural identification procedures such as a "voice identification parade", where they judge that appropriate⁹⁰ (at p. 142). Although the wording of the code (as at 2014) does not prevent the police from conducting "aural" identification parades it provides very little guidance on their conduct. The fact is that an accused need not answer any questions nor give evidence making voice identification without covert surveillance less likely. The negative effect of this is immense where voice identification forms a substantial part of the prosecution case.

⁹⁰ Police and Criminal Evidence Act 1984 (PACE): *Code D: Code of Practice for the Identification of Persons by Police Officers*. London: Her Majesty's Stationery Office, p. 142.

In the normal course of affairs the police cannot use an audio recording of an interview as material with which to carry out the procedure. There is an obvious risk that a witness may make a positive identification after listening to the content of a recording. Of course the accused can consent to this prior to the interview but the likely impact on his or her responses requires further research. In such circumstances an accused may be more likely to remain silent which stands in opposition to the principle that the accused should be encouraged to admit guilt as early in the proceedings as is possible.⁹¹ Alternatively, *RIPA* could be extended to cover voice recordings obtained at interview so that they can be sampled for analysis whether that is AAA or through BVIT. There are issues relating to police conduct of investigation and for the collection and use, retention and destruction of evidence. Furthermore, this would affect the public perception of the police and its functions. One thing is certain; the sophistication of BVIT has made its use more acceptable and routine in some jurisdictions. The real-time completion of procedures raises questions similar to those posed regarding DNA evidence.⁹²

The acceptance of BVIE as good evidence by the UK courts and increases in the use of covert surveillance have implications for the right to privacy under article 8 of the ECHR⁹³ which provides everyone⁹⁴ with a qualified right to respect for their private and family life, their home and correspondence. It is provided on the basis that no public authority should interfere with this right unless it is in accordance with the law and it is necessary in a democratic society⁹⁵ in the

91 This was one of the premises upon which the common law right to silence was eroded by the adverse inferences that are permitted to be drawn from silence by ss.34–37 of the Criminal Justice and Public Order Act 1994. See also the privilege against self-incrimination; *Blunt v Park Lane Hotel Ltd* (1942) 2 KB 253, *Rio Tinto Zinc Corporation v Westinghouse Electric Corporation* (1978) AC 547 and Article 6 of the European Convention on Human Rights and Fundamental Freedoms. Note: the privilege against self-incrimination is not an absolute right for instance the pursuance of clear and proper objectives in a proportionate manner by a government, see *Saunders v United Kingdom* (1996) 23 EHRR 313 and s.13 of the Fraud Act 2006.

92 See note 32.

93 See the European Court of Human Rights decision in *S and Marper v United Kingdom* 30562/04 (2008) ECHR 1581 particularly whether the biometric voice profile of those that are arrested for any recordable offence, even trivial or minor offences should be retained and if so for how long, and the prescribed time limit within which the biometric voice profile of those that are acquitted should be removed from the database. The latter problem is particularly acute as automated biometric voice identification is reliant on pre-recorded voice data.

94 The wording of this provision is taking directly from the ECHR however it is restated in a slightly altered form for the purposes of this article.

95 See Marx, G. (1988). *Undercover: Police Surveillance in America*. Oakland, CA: California University Press.

interest of national security, public safety or economic well-being of the country to prevent crime, disorder or for the protection of the rights and freedom of others.⁹⁶ BVIE gathered using covert surveillance authorised under RIPA 2000 has been used in the prosecution of offences in cases predominantly involving terrorism but also in cases involving trivial criminality. The Act allows material collected in relation to one investigation to be used as part of another that is unrelated.

10 Conclusion

Modern crime control is an information hungry activity that is preoccupied with prevention and as a result has rewritten the identity of entire populations. There is an urgent need for a debate on the unrelenting admission of unreliable evidence, this should not impede the long and admirable tradition of the UK court's being receptive to new forms of evidence regardless of how much a guarded approach they may take when enforcing minimum standards. This stance accepts the fact that knowledge evolves and that an accepted body of knowledge including particular methodological approaches may be proven unscientific or unreliable in the future. The latter outcomes are relevant considerations when setting standards, ensuring expediency and fairness, and preventing abuse.

Experts will continue to disagree how reliable AAA is but there is little doubt that the decision in *O'Doherty* changes the quality of voice identification evidence so as to provide a greater degree of certainty in the identification itself by means of comparative analysis. The court implicates the validity of the current techniques that are used for identifying the voice. This lends some support to the proponents of biometric voice identification and to the introduction of a reliability test designed to militate against the admission of unreliable evidence.

Biometric voice identification was not included in Yarmey's research nor has it seen mention in case law for example *Roberts* and *Robb*. There is a distinct lack of cohesive research on the adequacy and risks posed by the various types of voice identification evidence. This is surprising given the major technological advances and changes to the law on covert surveillance. It is submitted that

⁹⁶ The full text of the convention can be downloaded in English from the Council of Europe. Available at: www.echr.coe.int/Documents/Convention_ENG.pdf (Accessed 11 October 2013). See also: Gragl, P. (2013). *The Accession of the European Union to the European Convention on Human Rights*. UK: Hart Publishing. See also: *Uzan v Germany* (2011) 53 EHRR 24 and *El Haski v Belgium* (2013) 56 EHRR 31.

research is required to bring together the different types of voice identification evidence and to produce a coherent body of knowledge and framework for its admission and exclusion. This would help mitigate the risks posed by erroneous identifications and methodologically unreliable evidence, and of course to raise constructive discourse on BVIT/BVIE. In addition, an engaging and transparent debate must be had on the extent to which this technology has proliferated criminal justice in the UK.

Bionote

Charanjit Singh

Dr. Charanjit Singh is a Barrister and completed his PhD at the University of Southampton. He is currently the Head of Subject and Research at the School of Law – University of West London. Dr. Singh has written a number of textbooks and conducted research in evidence law and his current research, of which this article is part, is on biometrics, identification, security, surveillance and technology.