

Can we have faith jurors listen without prejudice?: Likely sources of inaccuracy in voice-comparison exercises

By Jeremy Robson and Dr Harriet Smith

Introduction

This paper addresses the question of how and when fact-finders in criminal trials (primarily, but not exclusively, jurors) can be permitted to listen to audio recordings with a view to identifying the speaker. This is usually done with a view to establishing that the speaker on the tape is the defendant. In conducting such an exercise, jurors are being asked to conduct a voice-comparison exercise. We argue that insufficient attention has been paid by the legal system to the variables which might impact upon the accuracy of that voice-comparison exercise and, as a result of this, there are insufficient safeguards in the law as it currently stands. For the purposes of discussion, we primarily focus our attention on the law as it stands in England and Wales but as what we say in relation to the factors which impact upon accuracy in such exercises is of universal application, it is relevant to the Anglo-American evidence system as a whole.

The Anglo-American tradition of evidence scholarship assumes a rationalist approach to the admissibility of evidence. The rationalist tradition strives for rectitude of decision making achieved through 'rational' methods for the assessment of evidence. An assessment of the validity of an institution's rules of evidence is based upon an assessment of whether it maximises the chances that an accurate decision will be made.¹ The legal system seeks to maximise accuracy through a triangulation of rules of admissibility (which serve to address the risk that fact-finders may make irrational decisions when presented with evidence of persuasive but not objective value), judicial directions to the jury (which act as a framework to promote rational analysis of the evidence) and procedural rules for the collection and presentation of evidence (which serve to facilitate the presentation of evidence in a fair and efficient way).

¹ See for example Twining W. *Theories of Evidence: Bentham & Wigmore*. 1st Edition. London: Weidenfeld & Nicolson, c1985, Chapter 3

In this paper we draw upon the current disciplinary knowledge within psychology to evaluate whether the legal position in relation to voice-comparison exercises does all it can to maximise accuracy. The issue of whether or not jurors can be permitted to listen to a recording of an incident in order to assist them in determining whether the voice on the recording is that of the defendant has long been controversial. Although the inherent danger in permitting such an exercise is well-recognised,² there remain no definitive guidelines for judges or advocates on the circumstances in which it should be permitted, the conditions in which the comparison should be made and how juries should be directed to approach a voice-comparison exercise. Although in the majority of cases, the primary comparison will be presented by means of an expert opinion, the jury's assessment of that opinion will inevitably be informed by their own assessment of the similarity between the two voices they hear. In cases where there is no expert analysis there may be evidence from a 'lay-listener' (someone familiar with the defendant's voice) who purports to recognise the speaker and whose claim the jury will have to assess. Where these issues have been considered by the appellate courts, the primary focus has been to the quality of the recording the jury are presented with rather than other variables which might impact upon the accuracy of the comparison exercise. In this paper we discuss some of the issues that are relevant in assessing the accuracy of voice-comparisons. We suggest that clear guidance is urgently needed to ensure that, as far as possible, if jurors are required to undertake a voice-comparison exercise it is conducted in such a way that the prospects of an erroneous conclusion being reached are minimised. We begin by outlining the current legal position in England and Wales and other jurisdictions. We go on to review research relevant to understanding likely sources of inaccuracy in voice-comparison exercises. Our aim is to identify some of the variables that may impact upon the accuracy of voice-comparison exercise and suggest how the justice system might seek to prevent misidentifications occurring.

The position in Law in England and Wales

² Ormerod D. "Sounding out expert voice identification. Expert Evidence and Scientific Proof in Criminal Trials." *Crim. Law. Rev.*2002. 770 – 790.

Recordings are prima facie admissible as best evidence as a matter of law.³ The purpose for which they can be used varies. Often they are presented by the prosecution as a means of evidencing the contents of a conversation, where the speakers are identified but where there is a dispute as to the context in which the words are spoken or what is being said. In some cases however, the issue the recordings address is the identity of the speaker, the resolution of which may determine whether a 'guilty' or 'not guilty' verdict is reached. In such cases a comparison needs to be made between the recording and an uncontested sample of the defendant's voice (either recorded on a prior occasion or as heard testifying in court). In many cases, this comparison will have been conducted by an expert and the jury will use the recordings to assess whether the expert is correct in their conclusion. In other cases juries have been entrusted to conduct the exercise without an expert.

Voice-comparisons are most frequently deployed by the prosecution in inculcating an accused although they are equally valuable as an exculpatory device for the defence. For example, in *Flynn and St John* the defence sought to invite the jury to compare the defendants' voices with the recording produced by the prosecution to establish a difference.⁴ In *Kapikanya*, a co-accused (Robinson) was permitted to introduce recordings of telephone calls to establish that Kapikanya was the main operator in a telephone fraud rather than him. The jury were permitted to make their own comparisons.

Much of the discussion surrounding the playing of recordings has arisen in the context of controversies surrounding the two techniques by which expert comparison of voice recordings takes place; auditory analysis and acoustic analysis. Auditory analysis involves the expert listening to a portion of recorded speech and a portion of known speech and then using their phonetic expertise to identify points of dissimilarity. Acoustic analysis is based upon an electronic representation of the sound produced, with the expert being called upon to interpret the spectrograph produced and compare contested and known samples. A much fuller discussion of the merits of these approaches is conducted by Ormerod.⁵ Where auditory analysis is conducted the expert conducting the analysis will usually be working with two or more samples of speech taken from the recording. Where the results

³ *R v Howells* [1965] 2 All ER 464

⁴ *R v Flynn and St John* [2008] 2 Cr. App. R. 20

⁵ Op Cit n 2.

are contested and the expert is required to testify before the court, the materials on which the analysis is based should be presented to the jury.⁶

In 2002, Ormerod cautioned that:

‘If the jury have not heard expert evidence on the voice identification, it is submitted that there are serious risks involved in allowing them to conduct ad hoc voice identification in the courtroom, whether by comparing a recording of the voice of the offender with a recording of the accused or with the voice of the accused heard live at trial. The dangers inherent in any stranger voice identification, exacerbated by the delay between hearing the voices, the stress of the exercise in the courtroom, the danger of bias, the risk of over-confidence from the jurors, all point strongly against this as a worthwhile exercise in terms of the likely accuracy of outcome.’⁷

In discussing the role jurors should play in assessing the evidence of an expert he noted that:

‘It remains arguable that the jury might be usefully performing the qualitatively different exercise of monitoring the experts' activity if they are provided with an opportunity to compare the recordings for themselves. Provided the limitations of voice recognition and comparison are made clear to them, they will be better able to evaluate the expert evidence.’⁸

The Court of Appeal endorsed this approach in *Chenia*.⁹ The case in *Chenia* was based in part on the assertion that the defendant was the speaker captured on a considerable number of surveillance tapes. The jury were directed that they could use their own judgment about ‘who was talking about what, why and with whom.’ In allowing the appeal the Court of Appeal commented that;

“[W]e do not think that a jury should, as it were, be asked to be their own voice expert. We have reached the conclusion that, on the particular facts of this case, where the jury were

⁶ *R v Bentum* (1989) 153 JP 538

⁷ *Op. Cit.* no 2 at page 787

⁸ *Ibid*

⁹ *R v Chenia* [2002] EWCA Crim 2345

unassisted by expert evidence, they should have been warned that they should not compare one voice with another by comparing the characteristics of each because of the dangers of doing so.”

The notion that *Chenia* was a bar to jurors ever conducting their own comparison was rejected in *Flynn and St John*, another case involving surveillance recordings. These were played to the jury with an expert and police officer confirming the speakers were the defendants. The trial judge directed the jury not to attempt any comparison themselves. It was conceded on appeal that this direction was incorrect.

The nature of the direction given by a judge in cases where an expert witness makes a comparison was revisited in *Suleman*.¹⁰ Suleman faced multiple counts of arson and making hoax 999 calls. In relation to these charges the Crown sought to adduce expert evidence comparing the recordings of the 999 calls with the defendant’s voice as recorded in interview. The judge informed the jury that they were not to engage in their own assessment of whether the voices matched but to treat them as a means of showing how the expert reached his assessment. In the summing up, the judge directed the jury that in considering the evidence “you are entitled to consider the voices using your own common sense.” It was argued on behalf of the appellant that this undermined the judge’s previous warning. The Court of Appeal rejected this stating that the judge was doing no more than “reassuring the jury that in judging the weight of Professor French's opinion they were entitled to recollect the voices they heard.”

Flynn now forms the basis for the general guidance given by the Bench Book to judges on how to direct juries when dealing with cases of voice identification/recognition.

Recordings are mentioned briefly where it is stated that;

¹⁰ *R v Suleman* [2012] EWCA Crim 1569

“If juries are permitted to listen to recordings to try to identify speakers, they should be reminded to bear in mind the evidence of the voice recognition witnesses and warned of the dangers of relying on their own untrained ears: *Flynn and St John*. ”¹¹

The authorities and guidance above deal solely with cases where the evidence which is sought to be adduced comes from the prosecution and is substantiated by an expert. The case of *Kapikanya* suggests the Court of Appeal have not prohibited identifications being made solely by juries without the support of an expert.¹² *Kapikanya* was charged, alongside others, with a series of mortgage frauds. His defence was that he was an innocent party to a series of transactions in which he believed the others involved were acting honestly. One of the co-defendants, Robertson, claimed that he too was an innocent party who had been duped by *Kapikanya* into participating in the frauds under false pretences. *Kapikanya* gave evidence over several days. Robertson did not testify but his counsel cross-examined *Kapikanya* at length. In the course of cross-examination, Robertson’s counsel put to *Kapikanya*, a recording of a conversation between one of the building societies who advanced the monies and a man who claimed to be one of the victims. It was accepted that the latter was an imposter. It was put to *Kapikanya* by Robertson’s counsel that his was the voice on the recording. *Kapikanya* denied that he was the person speaker. The prosecution did not seek to rely upon the recording or make the assertion that the appellant was the caller. Counsel for *Kapikanya* requested that the jury be directed that they should not conduct their own identification and to disregard the recording. The trial judge declined to give such a direction. He instead directed the jury that although expert evidence was often called in this type of case, that this should not preclude them making an identification. He gave the jury a ‘Turnbull’ direction on the risks inherent with identification and drew their attention to some of the areas which might affect their assessment. On appeal following conviction, the Court of Appeal concluded that nothing in either *Chenia* or *Flynn* required the judge to prevent the jury conducting their own identification. The Court noted that the judge was under a duty to allow Robertson to advance his defence fairly and therefore had to undergo a difficult exercise in

¹¹ Judicial College. *The Crown Court Compendium – Part 1 Jury and Trial Management and Summing Up* February 2017 at 15.7 para 3

¹² *R v Kapikanya* [2015] EWCA Crim 1507

striking a balance. They also accepted that the clarity of the recordings was better than in *Chenia* or *Flynn* and that the jury were able to hear the contested sample juxtaposed against the defendant's voice in cross-examination. In the circumstances the Court concluded that the directions given by the judge were adequate to meet any risk of injustice.

Whilst any analysis of *Kapikanya* must reflect the fact that, as the evidence was being introduced by the defence, and therefore the trial judge had a very little by way of an exclusionary discretion it does appear to give foundation to an argument that juries can be permitted to conduct voice comparison exercises without the assistance of an expert. Commentary in recent editions of Archbold have suggested that the approach in the Australian case of *Korgbara* (discussed below) should be followed in England and Wales.¹³ This approach leaves it to the discretion of the trial judge to decide whether the material is sufficient to allow a direct jury comparison, unsupported by evidence.

Some approaches in other jurisdictions

Whilst the issues relating to how recordings are played in court have therefore been touched upon in English and Welsh case law, the issue has inevitably arisen in other jurisdictions. Perhaps the most helpful analysis of some of the problems comes in the Northern Irish case of *O'Doherty*.¹⁴ In this case, in addition to auditory expert evidence called on behalf of the prosecution, the jury were invited to make their own comparison between a recording of a call made to the police and the voice of the appellant which they had heard in court. At the appeal hearing, Dr Francis Nolan, Reader in phonetics (as he then was) at the University of Cambridge, gave evidence that is summarised in the judgment in these terms:

“He expressed surprise that the jury were allowed to listen to the tapes and, in effect, make up their own minds. The members of the jury were, he said in his report, naturally not in a position to carry out instrumental acoustic analysis, nor even to bring to bear the kinds of phonetic and linguistic analysis which the experts used. The jury members were already

¹³ Archbold (2018 edn) at 14-73 (although the section quoted in Archbold appears to come from *Bulecik*)

¹⁴ *R v O'Doherty* [2003] 1 Cr. App. R. 5

inevitably under psychological bias from the very fact that Mr O'Doherty was in the dock and had been confidently identified by a police officer of his acquaintance. In this frame of mind it was not surprising if they heard a faint, noisy, partially unintelligible telephone recording of a broadly similar voice as being that of Mr O'Doherty.”¹⁵

And later on in even more robust terms:

“They had no training as to voice quality, pitch and intonation. The telephone recording was bad. If what they were doing was trying to remember the voice of Mr O'Doherty which they heard under the stressful and distracting context of the court case and later listen to the 999 tape, that would be disgraceful, he thought.”¹⁶

The Court of Appeal (Northern Ireland) rejected the proposition that juries should be prohibited from conducting their own assessment of the identification of a suspect. The Court stated that:

“It seems to us that if evidence of voice recognition is relied on by the prosecution, the jury should be allowed to listen to a tape-recording on which the recognition is based, assuming that the jury have heard the accused giving evidence. It also seems to us that the jury may listen to a tape-recording of the voice of the suspect in order to assist them in evaluating expert evidence and in making up their own minds as to whether the voice on the tapes is the voice of the defendant.”¹⁷

The Court reasoned by analogy with a line of authorities from both England and Wales, and Northern Ireland which permitted tribunals of fact to make their own assessment of whether a person captured in a video is the defendant they can see before them in court. Insofar as visual identifications are concerned this rule remains good law in England and Wales.

A series of authorities from Australia indicate a much less cautious (and as we shall discuss later, potentially unsafe) approach to the question of jury comparisons. In *Bulejck*, the appellant was

¹⁵ Ibid at para 24

¹⁶ Ibid at para 33

¹⁷ Ibid at para 65

charged with drugs offences and part of the case against him came from a recording of the Defendant's voice captured on an undercover recording device, the tape-recording of which was exhibited to the jury.¹⁸ The defendant at the trial made an unsworn statement in front of the jury, implying he was not the person on the tape. The trial judge directed the jury that they could use their own recollection of the defendant's voice at trial in assessing whether he was the person on the tape. Following the summing up, the tape of the defendant's voice made during the court proceedings was replayed to the jury. The appeal was allowed, on the technical ground that the tape of the defendant's unsworn statement had not formally been admitted in evidence, however the High Court of Australia were content to accept that but for that, the conviction would have been safe. Toohey and Gaudron JJ noted the distinction between 'voice identification' and 'voice matching' commenting that:

“Where two voices are being heard side-by-side, as occurred in the present case, the concern is not with familiarity or distinctiveness but with whether the quality and quantity of the material is sufficient to enable a useful comparison to be made... As to the quality and quantity of the material being compared, clearly the greater the amount of material, the greater the similarity in the circumstances in which the voices were spoken or recorded and the greater the number of similar words used, the more useful the comparison. A jury would also benefit from hearing the material more than once so as to enable them to concentrate on both similarities and dissimilarities. Counsel for each side should have the opportunity to point out or emphasise particular similarities or dissimilarities to the jury. The defence may wish to call expert evidence where the jury may have difficulty in drawing a distinction between two voices of a particular nationality or dialect.”

In *Korgbara* the trial judge extended this principle in allowing the jury to compare the defendant's voice at trial (spoken in English) with recordings where an individual spoke in Igbo (a Nigerian language). On appeal against conviction, the New South Wales Court of Criminal Appeal rejected the argument that a jury should not be permitted to make comparisons between voices speaking different

¹⁸ *Bulejick v R* [1995] HCA 54; 185 CLR 375

languages without expert evidence. The court concluded that to do otherwise would be to create an arbitrary rule which was inconsistent with the statutory scheme and the principles laid down in *Bulecjik*.¹⁹

In the U.S.A. the practice of allowing jurors to make their own comparisons is permitted. In *U.S. v Williams* a defendant was compelled to read an extract from Time magazine in front of the jury to allow a voice comparison to be made with an undercover recording of a drug deal.²⁰ The primary challenge to this approach on appeal was not to argue that it might promote an inaccurate identification but instead that it was a breach of the defendant's right against self-incrimination. This approach was rejected on appeal.

If it is possible to discern a common thread in these authorities it is of a pragmatic approach, where the greater the clarity of the recording and exposure to the suspect's voice the jury have, the more confidence there is that the jury can make their own assessment. However even in those cases where experts are relied upon, the jury will be the final arbiters of fact and whilst an expert may direct them on those factors of relevance in making the voice-comparison exercise that assessment will be theirs to make.

The problems of voice-comparison exercises: The influence of estimator variables and system variables

We would argue that the current approach of the law to voice matching exercises is too vague and that guidance is needed; primarily to assist courts in determining when and how juries should be permitted to listen recordings but also to assist juries with the factors they should incorporate into their assessment of the evidence. To date, much of the consideration of whether jurors should be allowed to engage in voice-comparison decisions has focused on the quality of the material to be played but with little discussion of other factors which might impact upon the jury's assessment of the evidence, including the way in which the comparison exercise is conducted. Jurors undertaking a voice-

¹⁹*R v Korgbara* [2007] NSWCCA 84 (30 March 2007)

²⁰*U.S. v Williams* 704 F.2d 315 (6th Cir. 1983)

comparison exercise are being asked to undertake a cognitive task: to make a decision, and to assess the certainty they have in this decision. This task is not isolated; certain external factors influence both the accuracy and the confidence in the decision-making process. In order to provide guidance to judges and juries there is a need to identify those factors which maximise the chances of an accurate comparison being made whilst ensuring any factors which might be unduly prejudicial are minimised.

The discipline of psychology is able to inform the development of these guidelines, providing an understanding of the variables which impact upon the decision-making process. It is sensible to take the lead from eyewitness psychology research, and separate the variables likely to influence voice-comparison accuracy into two categories. When assessing the accuracy of witnesses to crime, psychologists often refer to ‘estimator variables’ and ‘system variables.’²¹ The former are variables that the legal system *cannot* control; e.g., the age of the witness, or whether the perpetrator attempted a disguise. The latter are variables which *can* be controlled by the legal system. These include variables such as the way in which questions are asked of witnesses, or the type of identification procedure that is used. These categories are also relevant to voice-comparison exercises. In this context, estimator variables relate to the quality/nature of the recording and the listener, and system variables relate to the way in which recordings are presented for comparison. The criminal justice system must be vigilant to both estimator and system variables if it is to minimise the risk of miscarriages of justice. Roberts uses the analogy of diseases to explain the difference between the two variables both in terms of their cause and the responses the justice system can take to avoid miscarriages of justice. He describes estimator variables as being something the justice system is “powerless to prevent” but to which the criminal justice system should be alert and implement “an effective screening programme.” The system variables on the other hand are “caused by practices and procedures” of the system which can be avoided with appropriate regimens.²² We therefore discuss

²¹ Wells, G. L. (1978). “Applied eyewitness-testimony research: System variables and estimator variables.” *Journal of Personality and Social Psychology*, 36(12), 1546-1557. doi: 10.1037/0022-3514.36.12.1546

²² Roberts A. The problem of mistaken identification: Some observations on process. *International Journal of Evidence and Proof* (2004) 8 E&P 100-119

some examples of each. This list is not exhaustive, but illustrates some areas where variables have the potential to distort the accuracy of the process which deserve consideration by fact-finders.

Estimator variables

Better understanding of estimator variables in the criminal justice system will help to diagnose the likelihood of inaccurate voice-comparisons by lay witnesses. The accuracy of voice-comparisons by non-experts is a relatively under-researched area. In this section we discuss the likely effect of estimator variables on performance based on existing literature, and suggest how the justice system could perform the kind of ‘screening’ suggested by Roberts. There is still much research to be done in this field but the fact there are gaps in knowledge suggest the legal system should be especially cautious of adopting an ‘ad hoc’ approach in this area.

Physical attributes of the listener.

The possibility that personal characteristics of a witness may impact upon their fallibility as an eye-witness is well known; the short-sighted eyewitness would fall within the criteria of an observation which was ‘impeded’ within the criteria of Turnbull.²³ There has, however, been little discussion in the legal literature of the impact of the personal characteristics of the listener on voice identification accuracy when assessing evidence in court. Perhaps nowhere is this more important than in cases where perceptions of the qualities of two different voices are relevant to the determination a jury has to make. The ability to perceive all the attributes of a sample of speech which may make it similar or dissimilar to another sample is dependent to a large part upon the range of sounds a listener is able to hear.

Age is associated with a decline in hearing; hearing begins to degrade from around 40 years of age,²⁴ which has obvious implications for the accuracy of voice-comparison. The upper age for eligibility for jury service in England and Wales is now 75 years of age which suggests approximately 61% of

²³ R v Turnbull [1977] Q.B. 224.

²⁴ Hoffman, H.J., Dobie, R.A., Losonczy, K.G., Themann, C.L., Flamme, G.A.. (2017).” Declining prevalence of hearing loss in us adults aged 20 to 69 years.” JAMA Otolaryngology Head and Neck Surgery, 143(3), 274–285. doi: 10.1001/jamaoto.2016.3527

jurors are at risk of experiencing age-related hearing degradation.²⁵ Age-related hearing loss is particularly associated with difficulty in extracting a ‘signal’ (i.e. the voice) from ‘noise’ (i.e. any kind of background interference)²⁶. This is particularly relevant in cases where a voice has been recorded against a noisy background; older listeners may struggle to isolate vocal sounds from background noise and make an accurate comparison.

Advocates and judges need therefore to be alert to the fact that members of the jury may vary in terms of their hearing sensitivity, and that this may compromise their ability to accurately compare recordings. Jurors should be directed to be aware of this in their deliberations.

Quality of recording.

The quality of the recording (or recordings) has formed the basis for much of the discussions in the case law. If a voice is recorded on separate occasions using different equipment, the voice will sound different in each sample. Many of the cases likely to be encountered in the criminal justice system involve recordings of telephone calls where there is a dispute as to the speaker. In many cases the comparison will be being made with either a recording of the defendant speaking in interview or in court: this is likely to make accurate comparisons difficult.²⁷ The frequency spectrum of natural speech can fall in a range between 50-10000 Hz, but telephones only capture and transmit that speech which falls within the range of 300-3400 Hz.²⁸ This reduced bandwidth means that certain sounds (for example those characterised by high frequency energy, such as ‘s’ and ‘sh’) may be removed by transmission over a telephone, distorting the sound which is recorded. Where different telephones are

²⁵ Based on figures contained in the 2011 census. Office for National Statistics 2011 “Census: Population Estimates for the United Kingdom, March 2011” (<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationestimatesfortheunitedkingdom/2012-12-17>) [Accessed 25/05/2018]

²⁶ Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2011). Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response. *Ear and Hearing*, 32(4), 498-510. doi: 10.1097/AUD.0b013e31820512bb

²⁷ Hirson, A., French, P., & Howard, D. (1995). “Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics.” *Studies in General and English Phonetics: Essays in Honour of Professor JD O’Connor*, 230-240.

²⁸ Kerstholt, J. H., Jansen, N. J., Van Amelsvoort, A. G., & Broeders, A. P. A. (2006). “Earwitnesses: Effects of accent, retention and telephone.” *Applied Cognitive Psychology*, 20(2), 187-197. doi: 10.1002/acp.1175

involved the situation is further complicated as voices can sound different when recorded through landline and mobile phones,²⁹ making comparison less accurate.³⁰

Voice recordings obtained through covert recording are likely to feature background noise, which may disrupt the perception of voice identity. Background noise impairs speech perception³¹ and masks cues such as pitch,³² which is one of the most important features used by listeners in attaching an identity to a voice.³³ There is some evidence that background noise negatively affects performance in voice matching exercises.³⁴ This is particularly so when background noise features in one recording but not the other.³⁵

Although the quality of the recording is recognised in the case law as causing difficulties in voice-comparison exercises, the reasons and nature of the problems are not fully articulated. Whilst recording-quality may be acknowledged as an obstacle to be overcome it does not appear to be fully understood as a factor which may distort the outcome and result in erroneous decisions being made.

Differences in voice (especially voice used in court which may adjusted to meet conditions).

The human voice is not produced consistently and there is a wide a range of variation in the voice qualities of an individual speaker. Even if the recording equipment is identical, ‘within-speaker variability’ means that the same voice can sound very different in different situations. Although some voice features are relatively stable (e.g. pitch), others (e.g. tempo variation) are more likely to vary,

²⁹ Gonzalez-Rodriguez, J., Rose, P., Ramos, D., Toledano, D. T., & Ortega-Garcia, J. (2007). “Emulating DNA: Rigorous quantification of evidential weight in transparent and testable forensic speaker recognition.” *IEEE Transactions on Audio, Speech, and Language Processing*, *15*(7), 2104-2115. doi: 10.1109/TASL.2007.902747

³⁰ Alexander, A., Botti, F., Dessimoz, D., & Drygajlo, A. (2004). “The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications.” *Forensic Science International*, *146*, S95-S99. doi: 10.1016/j.forsciint.2004.09.078

³¹ Mattys, S.L., Davis, M.H., Bradlow, A.R., & Scott, S.K. (2012). “Speech recognition in adverse conditions: A review.” *Language and Cognitive Processes*, *27*, 953-978. doi: 10.1080/01690965.2012.705006

³² Qin, M. K., & Oxenham, A. J. (2003). “Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America*, *114*(1), 446-454. doi: 10.1121/1.1579009

³³ Nolan, F., McDougall, K., & Hudson, T. (2011, August). “Some acoustic correlates of perceived (dis) similarity between same-accent voices.” In *International Congress of Phonetic Sciences (ICPhS)* (pp. 1506-1509).

³⁴ Bartle, A., & Dellwo, V. (2015). “Auditory speaker discrimination by forensic phoneticians and naïve listeners in voiced and whispered speech.” *International Journal of Speech, Language & the Law*, *22*(2), 229-248. doi: 10.1558/ijsl.v22i2.23101

³⁵ Smith, H. M. J., Baguley, T. S., Robson, J., Dunn A. K., Stacey, P. C. (under review). “Forensic voice discrimination: The effect of speech type and background noise on performance”

for example, between read speech and conversational speech.³⁶ Witnesses giving evidence from the witness box in a trial are frequently required to speak at a volume which is audible to the judge and jury and to adjust the pace of their speech to ensure that notes can be taken. This may result in substantial changes to the speech from that which is present on the contested recording, and will affect listeners' ability to compare speech samples. Research has shown voice-comparison accuracy is higher when within-speaker variability is limited.³⁷

It is not just the mismatch which is important though; some speaking styles obscure important identity cues. Covert recordings commonly feature whispered speech, but when a person is whispering, cues such as fundamental frequency (which is the lowest and loudest component of the voice and therefore central to recognition) are not present. Bartle and Dellwo compared discrimination accuracy across voiced and whispered speech and, perhaps unsurprisingly, found that accuracy was lower when speech was whispered.³⁸

It is not uncommon for perpetrators committing an offence via the telephone to disguise their speech, particularly if the crime is abduction or blackmail. The most commonly adopted disguises are relatively crude, involving methods such as deliberately lowering pitch or mimicking a foreign accent.³⁹ Nevertheless, even unsophisticated methods lead to higher rates of incorrect voice discrimination when one voice is disguised and the other is not, compared to when neither voice is disguised.⁴⁰ Although this is an area where there is scope for further research, it would appear to be unwise to allow juries to reach their own conclusions in this respect.

Language (s)

³⁶ Remez, R. E., Rubin, P. E., & Nygaard, L. C. (1986). "On spontaneous speech and fluently spoken text: Production differences and perceptual distinctions." *The Journal of the Acoustical Society of America*, 79, S26. doi: 10.1121/1.2023137

³⁷ Alexander, A., Botti, F., Dessimoz, D., & Drygajlo, A. (2004). The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications. *Forensic Science International*, 146, S95-S99. doi: 10.1016/j.forsciint.2004.09.078

³⁸ Op.Cit no 34.

³⁹ Künzel, H. (2000). Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics* 7, 149-179.

⁴⁰ Reich, A. R., & Duke, J. E. (1979). Effects of selected vocal disguises upon speaker identification by listening. *The Journal of the Acoustical Society of America*, 66(4), 1023-1028. doi: 10.1121/1.383321

Although the courts in England and Wales have avoided asking juries to conduct voice-comparisons in languages other than English, cases such as *Korgbara* indicate that it is feasible that juries might be asked to listen to recordings which capture speech in a foreign language. Evidence shows that performing the voice-comparison task under these circumstances has the potential to compromise accuracy. People are more accurate when comparing samples of the same language rather than different languages, even when one of the languages is their own.⁴¹ Based on these results, it would seem that if a sample captures speech in a language other than that of the tribunal then a voice-comparison exercise should never be carried out by the jury. However even where the voice being compared is in English it should not be assumed that this ends the need for care. In the 2011 census, 7.7% of the population reported speaking a language other than English as their main language.⁴² Although statistically jurors who do not speak English as a first language are unlikely to comprise a significant proportion of any given jury panel, the potential for an individual juror to be less able to perform a voice-comparison exercise is something that the court should be alert to and address.

System variables

Much of the discussion in this area (both legal and psychological) has centred on estimator variables, however without knowledge of system variables, the criminal justice system will be unable to improve procedures and maximise the chances that accurate voice-comparison decisions are reached. In this section we identify those areas where an understanding of system variables might help improve the accuracy of decision making.

It is perhaps helpful to consider how voice-comparison exercises might operate within the trial process. Assuming that the prosecution are seeking to rely on a recording of speech where they allege the speaker is the defendant, this is likely to be played at a comparatively early stage of the prosecution case. If there is an expert being called to perform an auditory analysis they may present the results of their analysis alongside a known sample. Otherwise the exhibit may be presented by the

⁴¹ Wester, M. (2012). Talker discrimination across languages. *Speech Communication*, 54(6), 781-790. doi: 10.1016/j.specom.2012.01.006

⁴² Ons.gov.uk. (2018). Language - Office for National Statistics. [online] Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language> [Accessed 1 May 2018].

witness responsible for the recording. A known sample of the voice may be presented later during the prosecution case (for example by playing a recording of the police station interview) or the prosecution may be reliant on the jury hearing the defendant speak in court, usually when testifying as part of the defence case. Upon retirement the jury may be permitted to rehear any exhibits but cannot compel the defendant to speak again.

As the case law illustrates, the ways in which to-be-compared voices may be presented to a jury vary from case to case. There may be the situations of the type which occurred in *Kapikanya* where the recording is heard for the first time in cross-examination. In most cases the jury will be given no direction about what use they can make of the recordings until after the evidence has finished and they are retiring to consider their decision. It is necessary to fully explore whether such variations in procedure are likely to affect the accuracy of comparisons.

Sequencing of playback

Unlike faces, which can easily be presented alongside each other for comparison, when comparing voices listeners must hear the voices sequentially. The order in which the sequence is played may produce different results. When there is more than one recording of a voice, each recording is likely to have different characteristics; one might be better in quality than the other, or both might feature different styles of speech. In our own research, we asked participants to compare one recording that featured background noise to another that did not. Accuracy was higher when listeners heard the recording with background noise after they had heard the one with no background noise.⁴³

Gap in time between two samples

Sequential presentation of voices requires consideration for another reason. Although there will inevitably be a short gap in between the two voices, the length of this gap is likely to affect performance. To reach an accurate voice-comparison decision, it is necessary for the listener to have a clear mental representation for the voices that are to be compared - even high-quality auditory

⁴³Smith, H. M. J., Baguley, T. S., Robson, J., Dunn A. K., Stacey, P. C. (under review). "Forensic voice discrimination: The effect of speech type and background noise on performance"

representations start to degrade after only a couple of seconds.⁴⁴ It is possible there may be a critical time period within which both samples must be heard for any decision on the comparison to be meaningful although research has not yet established this. On the basis of what is known however, there is a strong argument for juries only being permitted to hear voices when they can hear both voices in immediate succession.

Bias

It could be argued that voice-matching decisions made in court are inherently biased compared to those conducted in experimental conditions. There is a danger that the jury will allow the knowledge they have about a case (i.e. that there is other evidence against the suspect is on trial) to influence their decision. This risk is amplified by the fact that, unlike visual identification procedures conducted on eyewitnesses, the identifying party (the jury) will have been told that a witness (expert or lay) has identified the suspect as the speaker.

The role of bias deserves more attention in relation to voice-comparison decisions. Bias plays an important role in eyewitness identifications decisions.⁴⁵ For example, to reduce the risk of a mistaken identification, it is recommended that eye and ear witnesses completing identity parades be warned that the perpetrator may or may not be present (Annex A para. 11 Code of Practice D, Police and Criminal Evidence Act⁴⁶). This instruction has been shown to reduce false positive identifications.⁴⁷ Whilst it is impossible to eliminate subconscious bias caused by the knowledge that the accused is on trial and, in most cases, the existence of other evidence, consideration should be given to a clear direction, delivered in advance of any evidence being presented that the voice on the recording may or

⁴⁴ Lu, Z. L., Williamson, S. J., & Kaufman, L. (1992). "Behavioral lifetime of human auditory sensory memory predicted by physiological measures." *Science*, 258(5088), 1668-1670. doi: 10.1126/science.1455246

⁴⁵ Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. E. (1998). "Eyewitness identification procedures: Recommendations for lineups and photospreads." *Law and Human Behavior*, 22(6), 603-647. doi: 10.1023/A:1025750605807

⁴⁶ Code of Practice D to the Police and Criminal Evidence Act 1984: Code of practice for the exercise by police of statutory powers to identify persons (most recent edition made pursuant to The Police and Criminal Evidence Act 1984 (Codes of Practice) (Revision of Codes C, D and H) Order 2017/2017, SR & O 2017/103)

⁴⁷ Malpass, R. S., & Devine, P. G. (1981). "Eyewitness identification: Lineup instructions and the absence of the offender." *Journal of Applied Psychology*, 66(4), 482-489. doi:10.1037/0021-9010.66.4.482

may not be the defendant. The jury should also be reminded in summing up of the factors which may cause extraneous influence on their decision making.

Acoustics of the courtroom

The physical location in which the identification takes place is another area which may impact upon the accuracy of decision making. In research investigating voice-comparison accuracy, participants complete experiments in a quiet room, and/or wear headphones to suppress external ambient noise. Courtrooms do not replicate these conditions. Acoustics of a room can impact the sound which is heard.⁴⁸ The court estate in England and Wales comprises a diverse body of premises whose design has been influenced by a number of competing factors.⁴⁹ Although standardised guides for court design which incorporated acoustic standards were in existence between 2004 and 2010⁵⁰ there is still evidence of issues of audibility even within courtrooms which meet these standards.⁵¹

Lay-listeners in court must contend with noise from their immediate environment in addition to any interference with the recording. Furthermore, issues relating to the acoustics of the courtroom such as reverberance, echoes, and sound diffusion can affect intelligibility as well as voice quality,⁵² and may make the voice-comparison task challenging.

There is still research needed on the accuracy of voice-comparison exercises conducted in a courtroom setting and the extent to which the effect of system variables can be minimised. Playing the recordings to jury members through headphones might be advisable in cases where two voice recordings exist. In cases where a recording is compared to a live speaker (i.e. the defendant), it should also be acknowledged that speaker-listener distance has an effect on speech and voice

⁴⁸ Bilsen, F. A. "Thresholds of perception of repetition pitch. Conclusions concerning coloration in room acoustics and correlation in the hearing organ." *Acta Acustica united with Acustica* 19.1 (1967): 27-32.

⁴⁹ Mulcahy, Linda. "Architects of justice: The politics of courtroom design." *Social & Legal Studies* 16.3 (2007): 383-403.

⁵⁰ Department for Constitutional Affairs (2004) *Court Standards and Design Guide*. London: Department for Constitutional Affairs.

⁵¹ Justice.org.uk. (2018). [online] Available at: <http://justice.org.uk/wp-content/uploads/2015/07/JUSTICE-In-the-Dock.pdf> [Accessed 1 May 2018].

⁵² Kob, M., Behler, G., Kampwolf, A., Goldschmidt, O., & Neuschaefer-Rube, C. (2008). "Experimental investigations of the influence of room acoustics on the teacher's voice." *Acoustical Science and Technology*, 29(1), 86-94. doi: 10.1250/ast.29.86

perception.⁵³ Distance risks weaker voice cues becoming inaudible, particularly for listeners with age-related hearing loss.⁵⁴ This may have a deleterious effect on voice-comparison performance. Careful consideration should be given to whether it is appropriate for juries to be comparing voices being relayed from two different locations in the courtroom. If this is unavoidable they need careful directions on the possibility of distortion and change in sound.

Towards a rational approach

The apparent simplicity with which a voice comparison exercise can be conducted in court belies the complexity of what is being asked of jurors and the, potentially, very significant consequences for an accused. Although this is an area where further research may be of benefit the theme which emerges from the current research is that there a number of different variables which can result in juries misinterpreting the evidence they are presented with. This is not to say important evidence should be excluded from the trial process, but that if it is to be introduced, it must be done with caution, ensuring that where variables cannot be eliminated, steps are taken to minimise their impact.

As a starting point, jurors should not be asked to make a comparison between voices unless there is expert auditory evidence which supports the assertion that they are similar (or dissimilar as the case may be). The role of the juror should be limited to that described by Ormerod, namely scrutinising and evaluating the work of the expert. The difference may initially appear a semantic one as a jury conducting such an exercise will still be required to hear and make an evaluation of the similarity between two voices; there is however an important distinction between this approach and an unaided comparison. Jurors who are scrutinising the auditory analysis of an expert are provided with guidance as to which support the conclusion being reached; their task is to examine whether these features are present. Jurors conducting this task without that assistance are being required to make a simple match with very little guidance on what they should be looking for.

⁵³ Crandell, C. C., & Smaldino, J. J. (2000). "Classroom acoustics for children with normal hearing and with hearing impairment." *Language, Speech, and Hearing Services in Schools*, 31(4), 362-370.

⁵⁴ Erber, N. P., Holland, J., & Osborn, R. R. (1998). "Communicating with elders: Effects of speaker-listener distance." *British Journal of Audiology*, 32(3), 135-138. doi: 10.3109/03005364000000058

If a party seeks to introduce a recording without such expert (either to inculcate or exculpate a defendant) it should be treated as inadmissible for that purpose. This should apply equally, whether it is the prosecution seeking to adduce the evidence to implicate a defendant, or a co-defendant seeking to implicate a co-accused. Whilst the criminal courts are understandably reluctant to limit defence evidence, there is a requirement that such evidence be legally relevant, namely that a jury properly directed 'could place some weight upon it.' Whilst this threshold is low it is arguably not met in cases where there is no expert to perform the matching exercise.

When an expert presents evidence of a voice comparison they should produce the recordings upon which their conclusions are based. This should be in a format where the known voice and the contested voice are played sequentially and in quick succession. Where factors such as background noise are present, the sequencing of the recording should be such that the recording with the least interference is heard first.

Where a recording is to be played to the jury, before hearing the exhibits, they should receive a clear judicial direction as to the nature of the exercise they are being asked to perform and the potential limitations. Ideally jurors should have the opportunity to listen to the tape via headphones. If the jury request that the recording be played again at a later stage in the trial, they should listen to the entire recording in open court and be reminded of those features which the expert has identified as being relevant to the comparison.

The warnings given to the jury about the limitations should be repeated in summing up with jurors receiving a warning akin to a Turnbull direction. This will include case specific warnings on factors such as the length and quality of the recordings being used as well as cautioning the jury on factors which may impede their individual perceptions such as degradation in hearing. Where the language being compared is not English, the jury should be warned in the clearest possible terms about the difficulty of non-native speakers performing comparisons.

Voice identification will always be a particularly challenging forensic exercise. We suggest that the steps outlined above go some way to rationalising the process when part of the evidence against an

accused is a contested voice recording. By adopting an approach informed by inter-disciplinary research the possibility of mistakes being made by jurors can be reduced. As research in this area develops it may well be the case that this process is refined and improved, however unless there is a recognition of the difficulties involved with voice comparison, the risks of wrongful convictions will be substantial.

Jeremy Robson, *Senior Lecturer, Faculty of Business and Law, De Montfort University, Leicester.*

Dr Harriet Smith, *Independent Research Fellow, Department of Psychology, Nottingham Trent University, Nottingham.*