

# *'Mere guesswork': Cross-Lingual Voice Comparisons and the Jury*

**Gary Edmond,<sup>\*</sup> Kristy Martire<sup>†</sup> and Mehera San Roque<sup>‡</sup>**

---

## *Abstract*

Over the last two decades there has been significant growth in the number of cases involving voice comparison and identification evidence based on audio surveillance technologies. Courts have generally been open to this evidence, as well as allowing juries to undertake their own comparisons. Overall, decisions on the admissibility of this evidence feature a strong resistance to attempts to require some assessment of the reliability of voice comparisons and an over-reliance on traditional features of the adversarial trial, including directions and warnings, to expose and address weaknesses in this evidence. This article focuses on two cases where the jury was invited to compare voices speaking across two different languages. Drawing on empirical research concerned with the validity and reliability of lay voice comparison, this article outlines the dangers associated with juries engaging in cross-lingual voice identification and comparison. It argues that if courts are to take seriously their obligation to ensure a fair trial, they should be willing to exclude voice comparison evidence of unknown probative value and reliability, rather than rely on a 'properly instructed jury' as the best mechanism for managing the frailties of this evidence.

---

<sup>\*</sup> Professor, School of Law and Director, Expertise, Evidence & Law Program, University of New South Wales. This research was supported by the Australian Research Council (DP0771770 and FT0992041). An earlier version of this essay was presented at the Evidence Law in the 21<sup>st</sup> Century Conference, University of Sydney, September 2010. The authors would like to thank the participants at that Conference, as well as David Hamer, Martina Pasqualino and the anonymous referees, for their comments on earlier versions of this paper.

<sup>†</sup> Lecturer, School of Psychology, University of New South Wales.

<sup>‡</sup> Senior Lecturer, School of Law, University of New South Wales.

## I Introduction

Over the last two decades there has been significant growth in the number of cases involving the reception of voice comparison and identification evidence based on audio surveillance technologies. Where questions have arisen about the admissibility or reliability of such evidence, courts in New South Wales and elsewhere have generally been open to allowing both direct witnesses (for example someone present at a crime scene) and displaced witnesses (for example, an interpreter listening to surveillance material) to give voice identification and comparison evidence, and, further, where tape recordings (or voices) are available, to allow juries to undertake their own comparisons.<sup>1</sup> In this article we focus on two cases where the jury was invited to compare voices speaking across two different languages, one in which an apparent *expert* also gave evidence, and one in which the court rejected the defence argument that such jury comparisons should be accompanied by expert testimony.

Overall, decisions on the admissibility of voice evidence based on audio technologies feature a strong resistance to attempts to require some assessment of the reliability of displaced comparisons, an over-reliance on the capacity of the adversarial trial to expose and address weaknesses and an exaggerated confidence in juror abilities. We argue that these features are troubling in all cases, but are especially so in cases where courts are relying on juries to make their own assessments. Further, absent information about reliable methods of voice comparison, there is an additional risk that the *experts* that are currently allowed to give evidence are not, in fact, able to make reliable comparisons either. Drawing on empirical research concerned with the validity and reliability of lay voice comparison, this article outlines the dangers associated with juries engaging in voice identification and comparison, and, in particular, addresses the shortcomings of reliance on judicial directions as the preferred method for managing the dangers associated with the reception of such evidence.<sup>2</sup>

---

<sup>1</sup> Displaced listeners and witnesses are not present at the scene, but instead listen to recordings (and the accused's voice) at some later time, thus displaced listeners include jurors making their own comparisons. The distinction is significant because 'identification evidence' under the Uniform Evidence Law (the UEL) is restricted to assertions of identity or resemblance about the defendant, based wholly or partly on direct witness perception either at the scene of the crime, or at a related location. The UEL comprises the *Evidence Act 1995* (Cth), *Evidence Act 1995* (NSW), *Evidence Act 2001* (Tas), *Evidence Act 2004* (Norfolk Island) and the *Evidence Act 2008* (Vic).

<sup>2</sup> For extensive discussion of current concerns relating to jury directions, see New South Wales Law Reform Commission, *Jury Directions*, Consultation Paper No 4, (2008); Queensland Law Reform Commission, *A Review of Jury Directions*, Report No 66, (2009); Victorian Law Reform Commission, Final Report No 17 (2009). Each discuss directions required in relation to identification evidence but none address the particular problems of voice identification.

## II The Common Law and the UEL Framework

Visual identification evidence is a type of evidence requiring special attention and caution in terms of both admissibility and warnings to the jury.<sup>3</sup> There are detailed statutory provisions governing the use of eyewitness testimony, identification parades, photo arrays and visual and image comparison evidence.<sup>4</sup> The admission of voice evidence in Australia is, by contrast, hardly subjected to any regulation at all.<sup>5</sup> The lack of specific attention to voice identification evidence is compounded in the case of *displaced* voice comparisons; although often characterised as ‘identification’ evidence, displaced comparison evidence is situated awkwardly at common law and, despite some contrary authority, does not fall within the definition of ‘identification evidence’ under the Uniform Evidence Legislation (‘UEL’).<sup>6</sup>

Judicial consideration of voice identification and comparison evidence, and particularly the use of voice recordings, is relatively recent.<sup>7</sup> In the decade before the introduction of the UEL, courts in New South Wales began to approach *voice identification evidence* — usually where a sensory (or direct) witness positively identifies a voice associated with a criminal act — in a manner that highlighted concerns about reliability, but generally stopped short of imposing mandatory conditions.<sup>8</sup> However, at common law, the courts of New South Wales

---

<sup>3</sup> UEL Part 3.9, especially ss 116 and 165. These concerns are longstanding at common law: see eg *Davies and Cody v The King* (1937) 57 CLR 170; *Alexander v The Queen* (1981) 145 CLR 395; *Domican v The Queen* (1992) 173 CLR 555.

<sup>4</sup> See eg UEL ss 114–15. Contrast the approach to images in Gary Edmond et al, ‘Law’s Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images’ (2009) 20 *Current Issues in Criminal Justice* 337.

<sup>5</sup> This is in contrast with other jurisdictions such as the UK. On the UK, see David Ormerod, ‘Sounding Out Expert Voice Identification’ [2002] *Criminal Law Review* 771.

<sup>6</sup> Because of the restrictive definition of ‘identification evidence’ under the UEL, displaced comparisons based on audio or visual surveillance recordings, made by a person who was not present at the time of the offence (or related act), are not within the definition and thus do not trigger the warning provisions in s 116. Strictly, even identifications within the definition are comparison exercises leading to the identification of the person of interest. The terminology used in the cases includes voice identification, voice comparison and voice recognition. However, the boundaries between the different categories are not fixed, and often seem to depend on how the evidence is presented rather than the kind of evidence, and one case can involve a witness engaging in a combination of voice comparison, identification and recognition.

<sup>7</sup> For a detailed discussion of the development of both the common law and UEL jurisprudence on the admission of voice identification and comparison evidence, see Gary Edmond, Kristy Martire and Mehera San Roque, ‘Unsound Law: Issues with (Expert) Voice Comparison Evidence’ [2011] *Melbourne University Law Review* (forthcoming).

<sup>8</sup> *R v Smith* [1984] 1 NSWLR 462; *R v EJ Smith* (1986) 7 NSWLR 444 (Court of Criminal Appeal); *R v Corke* (1989) 41 A Crim R 292, 296; *R v Brotherton* (1993) 29 NSWLR 95; *Bulejck v The Queen* (1996) 185 CLR 375 (‘*Bulejck*’). See also *R v*

did consider the quantity and quality of material available to the witness, the distinctiveness of the voice in question, the level of the listener's familiarity, and whether the voices were compared under similar conditions. In practice, such considerations infrequently led to the exclusion of positive identifications by strangers, though there were cases where the lack of perceived distinctiveness of the voice *identified* resulted in exclusion.<sup>9</sup> Even in the cases associated with a more restrictive approach, rather than exclude voice evidence, appellate judges generally required that limitations and problems with voice identification evidence or voice comparison exercises should be brought to the attention of the jury by the trial judge through specific directions and warnings.<sup>10</sup>

With the introduction of the UEL in 1995, the courts of New South Wales effectively embraced the common law approach, preferred in the remaining common law jurisdictions, that admission is based on judicial discretion.<sup>11</sup> The trend has been to reject the imposition of conditions on admissibility, and even to avoid treating voice identification and comparison evidence as opinion, treating it instead as direct evidence of fact, sometimes characterised as 'recognition'.<sup>12</sup> Voice evidence tends to be treated as admissible if relevant. That is, it is considered sufficient that, *if accepted*, it could rationally affect the assessment of the probability of a fact in issue, and where recorded evidence is available, the tribunal of fact is frequently encouraged to

---

*Colebrook* [1999] NSWCCA 262. The approach in *EJ Smith* was seen to represent a more restrictive approach to the admissibility of voice identification, and is described as imposing a threshold criterion of prior familiarity, see the discussion of *EJ Smith* in *Bulejcik* and later in *Korgbara* (2007) 71 NSWLR 187. The NSW approach in *EJ Smith* was recognised, though not formally endorsed, by the High Court in *Bulejcik* (a jury comparison case where the accused was described as having an accent) and there were different opinions in *Bulejcik* as to the need for threshold criteria.

<sup>9</sup> See eg *R v Brownlowe* (1986) 7 NSWLR 461.

<sup>10</sup> Eg *Smith* [1984] 1 NSWLR 462; *EJ Smith* (1986) 7 NSWLR 444; and *R v Brotherton* (1993) 29 NSWLR 95. See also *R v Colebrook* [1999] NSWCCA 262.

<sup>11</sup> The common law position from the other states, endorsed in NSW in *R v Adler* [2000] NSWCCA 357 (23 August 2000) is summarised in *R v Callaghan* (2001) 4 VR 79, 92[25], 94 [27].

<sup>12</sup> Where characterised as opinion evidence, the courts are similarly inclined to facilitate admission (sometimes invoking UEL s 78). See for example *R v Leung* (1999) 47 NSWLR 405 where the evidence of the interpreter was initially admitted at trial via s 78 but then held to be admissible (*ad hoc*) opinion evidence on appeal. Courts have been inconsistent in their responses to the nature of interpretations of voice evidence. Sometimes this issue is discussed, as in *Neville v The Queen* (2004) 145 A Crim R 108 [44]–[46] (Miller J); *R v Harris (No 3)* [1990] VR 310, 318 (Ormiston J); *Li v The Queen* (2003) 139 A Crim R 281, 286–7 [39]–[42] (Ipp JA), and in relation to images, in *R v Smith* (1999) 47 NSWLR 419, 422–3 [16]–[22] (Sheller JA). On other occasions the question appears to have been overlooked, or simply assumed, as in *R v El-Kheir* [2004] NSWCCA 461 [96] (Tobias JA) and *R v Watson* [1999] NSWCCA 417, [39] (Newman J).

undertake its own comparison.<sup>13</sup> Directions and warnings are the preferred way to manage the problematic dimensions of evidence derived from voices and comparisons of voices, and the mandatory and discretionary exclusions are rarely successfully invoked.<sup>14</sup> With these developments, all Australian jurisdictions have either abandoned, or elected not to follow, the more restrictive approach associated with earlier decisions in New South Wales. Throughout Australia, voice identification and comparison evidence — whether as evidence of opinion or fact — is routinely admitted and questions about probative value and reliability are almost always left to the tribunal of fact as a matter of weight.

### III Two Cases of Cross-Lingual Comparison

The preference for managing the potential dangers of incriminating voice comparisons via ‘very careful instructions’ to the jury was expressed in both the judgment of Brennan CJ and the joint judgment of Toohey and Gaudron JJ in *Bulejck*.<sup>15</sup> While acknowledging issues such as distinctiveness and the quantity and quality of the samples available to the witness will be factors relevant to whether the evidence ought to be excluded on discretionary grounds, Toohey and Gaudron JJ made it clear that an appellate court should ordinarily confine itself to assessing the warnings:

This Court should be slow to depart from a trial judge’s assessment that material was of sufficient quality and quantity for the jury to be permitted to make the necessary comparison. The question rather is whether the jury were given sufficient warning of the difficulties involved.<sup>16</sup>

The trial judge is required to provide ‘very careful directions as to those considerations which would make a comparison difficult and... a strong warning as to the dangers involved in making a comparison.’<sup>17</sup> However, Brennan CJ noted that the sufficiency of any warning was ‘not assessed by reference to a formula nor by postulating a hypothetical warning against risks of which a reasonable jury would be

---

<sup>13</sup> This encouragement is replicated in the Judicial Commission of New South Wales, *Criminal Trial Courts Bench Book* (at 12 August 2011) ss 3–110 where the directions in one of the *ad hoc expert* cases are cited with approval: ‘You are permitted, indeed encouraged to make your own observations of the voices in those tapes to see if you would accept what opinion has been proffered’: *Li v The Queen* (2003) 139 A Crim R 281, 294 [99].

<sup>14</sup> Tim Smith and Stephen Odgers, ‘Determining “probative value” for the purposes of section 137 in the uniform evidence law’ (2010) 34 *Criminal Law Journal* 292.

<sup>15</sup> (1996) 185 CLR 375. While formally in dissent in this case, Brennan CJ’s approach has been highly influential in terms of reinforcing the preference for directions and warnings over stricter admissibility constraints.

<sup>16</sup> *Ibid* 397 (emphasis added).

<sup>17</sup> *Ibid* 398–9.

as well aware as the trial judge.’<sup>18</sup> Furthermore, he expressed a reluctance to ‘impose an artificial restraint on the jury’s employment of their common sense’,<sup>19</sup> concluding that the recognition of voices was ‘a commonplace of human experience’.<sup>20</sup>

In the two cases discussed below, where this approach is applied to voice comparisons conducted by displaced listeners, it is clear that this reluctance to interfere in what is perceived to be the jury’s fact-finding role, coupled with the reliance on warnings and directions, fails to address adequately the shortcomings of both the evidence allowed to go to the jury and the comparison exercise that the jury was permitted (and indeed expected) to undertake.

### *A Nguyen v The Queen*

*Nguyen*<sup>21</sup> was an appeal in the Western Australian Court of Criminal Appeal from Huy Duc Nguyen’s conviction for involvement (that is, ‘knowing concern’) in the importation of heroin. The circumstantial case against Nguyen was objectively strong. It included the use of his mobile phone to make arrangements for the shipment and collection of a package containing the heroin. Nguyen alleged that his phone had been used by others, thereby putting the speaker’s identity in issue.<sup>22</sup>

At trial, an interpreter, Colin Nguyen (‘CN’), was allowed to express his opinion about the identity of two voices on incriminating voice intercepts he had been asked to translate (from Vietnamese into English) during two months of surveillance of the accused’s phone.<sup>23</sup> Previously, CN had been provided with ‘voice reference material’. This consisted of three verified samples of Nguyen speaking on his mobile phone — two calls to a real estate agent in English and one to his wife in Vietnamese. He was asked to compare these known samples with the voices he had heard while monitoring telephone intercepts to and from the accused’s mobile phone. Without nominating, or being able to describe any particular characteristics of the voice(s), at trial CN was allowed to identify positively one of the ‘unknown’ voices on the incriminating intercepts as belonging to the appellant.

---

<sup>18</sup> Ibid 384. Though practically it may be impossible to prevent the jury making the comparison where such evidence is admitted: see *R v O’Sullivan* (1969) 1 WLR 497.

<sup>19</sup> *Bulejck* (1996) 185 CLR 375, 383.

<sup>20</sup> Ibid, 381.

<sup>21</sup> *Nguyen v The Queen* (2002) 131 A Crim R 341.

<sup>22</sup> However, the person making the incriminating phone calls was often referred to, or referred to himself, as ‘Huy’ and discussed issues relevant to a bakery. The appellant was named Huy and owned a bakery.

<sup>23</sup> *Nguyen* (2002) 131 A Crim R 341, 368 [121]: ‘... that to his ear (that is, in his opinion)’.

In addition to CN's opinion evidence, six incriminating intercepts — in Vietnamese — and the three reference samples were played to the jury during the trial. There appear to have been few objections to this procedure and little was said to the jury about CN's voice identification evidence or the use the jury might make of the nine voice recordings.

The only ground of appeal stated:

The learned judge erred in law in failing to give to the jury a direction as to voice identification and voice comparison with respect to recordings of intercept telephone calls allegedly made by the appellant.<sup>24</sup>

The appellant was concerned about the failure of the trial judge to warn about the dangers of 'human perception and recollection', which, according to submissions, might make even the identification of 'familiar' voices potentially 'unreliable'.<sup>25</sup>

Given the way in which *Nguyen* is relied on in our second case example below, it is illuminating to consider the basis for the admission of CN's opinion about the identity of the *unknown* speaker — which sat alongside, and presumably informed, any comparison undertaken by the jury. Malcolm CJ, whose decision addressed the issue in the most detail, rehearsed the criteria governing the admissibility of expert opinion evidence. There was, for him, 'no challenge to [CN's] expertise or the procedure which he adopted'.<sup>26</sup> CN was an accredited and, according to Anderson J, 'highly qualified interpreter in the English and Vietnamese languages'.<sup>27</sup> CN's opinion evidence about the identity of the speaker was admissible because he 'was qualified to make that comparison both by his familiarity with the Vietnamese language and his study of the 600 intercepted telephone calls'.<sup>28</sup> Malcolm CJ explained the admissibility of CN's opinion on the basis that it 'no doubt required his expertise in and knowledge of the Vietnamese language to be able to express that opinion'.<sup>29</sup> Distinguishing CN's opinion on the identity of the speaker from cases involving 'recollection', Anderson J agreed:

In considering the evidence of [CN], there could be no question about the quantity and quality of the material that was available to him for the purposes of comparison. I repeat that his opinion was not based upon any process of recollection, but was formed in the course of listening to more than 300 intercepts

---

<sup>24</sup> Ibid 368 [123].

<sup>25</sup> Ibid 370 [133].

<sup>26</sup> Ibid 347 [28].

<sup>27</sup> Ibid 366 [112].

<sup>28</sup> Ibid 357 [67].

<sup>29</sup> Ibid 360 [81].

over the life of the warrant relating to the appellant's telephone.<sup>30</sup>

Questions about CN's expertise and the existence of a 'field', raised for the first time on appeal, were simply dismissed.

It was suggested on behalf of the appellant that in this answer [CN] may have conceded that he was not a formal expert in voice comparison. This raises the question whether voice comparison is a recognised field of expertise and, if so, whether there was any evidence about it at the trial. In my opinion, in the absence of any such evidence and the absence of any objection to the evidence itself or to the qualifications or experience of [CN] to give it, it is too late to take this point.<sup>31</sup>

Similarly, the form and basis of CN's opinion seem to have been less important than whether he had listened to enough material to become familiar.<sup>32</sup>

At trial, *in addition* to the reception of the transcripts prepared by CN and of his opinion about the identity of the voice(s) on the incriminating intercepts, the nine recordings were played to the jury. The trial judge said nothing about whether the jury could undertake their own comparison of the voices on the six incriminating intercepts and the three reference calls. Though, according to Anderson J, they had at the very least been impliedly invited to do so. The appellant submitted that 'the learned trial judge should have warned the jury that it should not embark upon such a comparison, or alternatively, if it did embark on such a comparison, to do so with extreme care.'<sup>33</sup> The Court rejected these submissions. Anderson J explained that the jury was entitled to undertake its own voice comparison:

I see no reason why the jury are not entitled to compare voice recordings in order to come to their own conclusions. Voice recognition is not, of itself, an expert process. As Brennan CJ said in *Bulejck*: 'Recognition of a speaker by the sound of the speaker's voice is a commonplace of human experience.' ... It is clear that it is permissible for the jury to make their own comparison.<sup>34</sup>

---

<sup>30</sup> Ibid 373 [142]: This exposure to hundreds of conversations (more than 600 in total) distinguished the circumstances from several cases, relied upon by the defence, which focused on a stranger's recollection from one exposure.

<sup>31</sup> Ibid 355 [60]. The evidence becomes admissible because, rather than in spite of, a lack of agreement as to reliable methods of voice comparison.

<sup>32</sup> Note that this interest in familiarity is not common in other decisions, but is characteristic of the ad hoc expert decisions. See Gary Edmond and Mehera San Roque, 'Quasi-Justice: Ad Hoc Expertise and Identification Evidence' (2009) 33 *Criminal Law Journal* 8.

<sup>33</sup> *Nguyen* (2002) 131 A Crim R 341, 372 [137].

<sup>34</sup> Ibid 371 [138]–[139] (Anderson J), citing *Bulejck* (1996) 185 CLR 375, 381; *Smith* (1990) 50 A Crim R 434, 453–4, and *Barker* [2002] WASCA 127.



Malcolm CJ endorsed this approach: ‘authorities ... make it clear that the members of the jury were entitled to make their own comparison’.<sup>35</sup> For Malcolm CJ and Anderson J, the (assumed) ability to ‘recognise’ a speaker derived from *Bulejckik*, rather than issues relevant to cross-lingual voice identification, is used to ground the jury comparison. Malcolm CJ was, in addition, content for the jury to be *aided* in their comparison by knowledge of the interpreter’s opinion.

The main problem addressed by the Court of Criminal Appeal, reflected in the actual submissions, was not the admission of CN’s opinion or the jury’s comparison, but rather the trial judge’s failure to warn the jury about the dangers with the identification and comparison evidence. All the trial judge had said in relation to the recordings played to the jury and CN’s evidence was:

There are a number of questions for you to decide, but I have listed some which I think will be important for you to consider... First, obviously, was the accused the speaker Mr Huy on the phone in all those conversations; namely, not only the three he admitted being involved in, but the other six.<sup>36</sup>

‘This was not,’ as Anderson J explained, ‘a warning against an uncritical acceptance of voice identification evidence, nor did it give any guidance to the jury as to the matters that they should consider in evaluating [the] evidence’ of CN or in undertaking their own comparison of the voices speaking on the intercepted telephone calls.<sup>37</sup>

In response to the failure to adequately warn the jury about the voice identification evidence and voice comparison exercise, the Court of Criminal Appeal rehearsed the legally-orthodox need for caution in response to evidence of identification. ‘Identification evidence’ should, they explained, be ‘viewed with caution and it is well-settled that in cases of disputed identification evidence there is a need to warn juries about its intrinsic unreliability.’<sup>38</sup> Drawing from cases involving both eyewitness and earwitness identifications, the judgments reiterate the need for caution but also, in relation to voice identification, a reluctance to depart from the trial judge’s assessment, as laid down in *Bulejckik*, of there being material ‘of sufficient quality and quantity for the jury to make the necessary comparison’.<sup>39</sup> Malcolm CJ affirmed ‘the need to warn about the dangers of convicting on identification evidence’,

---

<sup>35</sup> *Nguyen* (2002) 131 A Crim R 341, 355 [57].

<sup>36</sup> *Ibid* 369 [126].

<sup>37</sup> *Ibid* 369 [127].

<sup>38</sup> *Ibid* 368 [124].

<sup>39</sup> *Ibid* 374 [145].

derived from the eyewitness case of *Domican v The Queen*,<sup>40</sup> and then, with reference to the voice comparison case of *Bulejck*,<sup>41</sup> he continued:

Specifically, ‘the jury must be instructed “as to the factors which may affect the consideration of [the identification] evidence in the circumstances of that case”’. Attention must be drawn to any weaknesses in the evidence. Reference to counsel’s arguments is not sufficient. The direction must have the authority of the trial Judge behind it.<sup>42</sup>

Anderson J also cited and recognised a range of authorities, specifically on the need for warnings in relation to eyewitness and voice identification (or earwitness) evidence. While endorsing the various authorities, he distinguished issues in the reported cases from those relevant in the instant appeal.<sup>43</sup> There were not, after all, long delays and comparisons across different media (eg a telephone intercept and live speech).<sup>44</sup> Reviewing the earlier voice identification cases of *R v EJ Smith*,<sup>45</sup> *R v Brownlowe*<sup>46</sup> and *R v Harris (No 3)*,<sup>47</sup> he concluded that they were all ‘distinguishable’ and ‘it does not necessarily follow that the warnings which were held to be necessary in those cases are warnings which would be appropriate in a case such as this.’<sup>48</sup> Notably, the ‘exercise in comparison which was undertaken by [CN] and, it may be inferred, by the jury was an exercise in comparing *like with like*’.<sup>49</sup> Moreover, this was not a case where there was a need to:

alert the jury to the possibility that the witness was insufficiently familiar with the voice.... This is not that kind of case. In making their own comparison, the jury were not engaged in evaluating the evidence of a voice recognition witness, but in forming their own opinion as to the identity of the voices based upon their own sense of hearing.<sup>50</sup>

Lack of familiarity seems to have been tempered by the ability of the jury to listen to the recordings ‘at the same time’; that is, sequentially. Consequently,

the jury were being called on to compare the recorded voice of the appellant in three mobile telephone calls with a recorded voice in six mobile telephone calls, all played to them at the same time. I think it would have been inappropriate for the jury

---

<sup>40</sup> (1992) 173 CLR 555 (*‘Domican’*).

<sup>41</sup> (1996) 185 CLR 375.

<sup>42</sup> *Nguyen* (2002) 131 A Crim R 341, 358 [71], quoting from *Domican* (1992) 173 CLR 555, 561–2 and *Bulejck* (1996) 185 CLR 375, 397–8.

<sup>43</sup> *Nguyen* (2002) 131 A Crim R 341, 369 [127].

<sup>44</sup> *Ibid* 372–373 [141].

<sup>45</sup> (1986) 7 NSWLR 444.

<sup>46</sup> *Ibid* 461.

<sup>47</sup> [1990] VR 310.

<sup>48</sup> *Nguyen* (2002) 131 A Crim R 341, 370 [131].

<sup>49</sup> *Ibid* 373 [141] (emphasis added).

<sup>50</sup> *Ibid* [142].

to be warned of the dangers which arise from weaknesses in ‘human perception and recollection’, to use counsel’s words. In undertaking the comparison for themselves, they would not be relying on the accuracy of someone else’s perception or on their or someone else’s recollection.<sup>51</sup>

The absence of legally notorious dangers did not mean that there was no need to warn the jury about voice comparisons at all, but rather, that many of the risks attending the admission of voice comparison and identification evidence by *witnesses* were (purportedly) not present. Nevertheless, Anderson J was of the opinion that the jury should have been given directions about the voice recordings and any comparison they might undertake.

As to whether the jury should have been given an instruction that they should be careful before concluding that the voices were the same, I do think it would have been better if some such direction had been given. The jury were listening to speakers in a foreign language and they might have been told *to bear that in mind* in attempting a comparison.<sup>52</sup>

For Anderson J, though, the presence of a second Vietnamese voice on the incriminating recordings reduced the danger that the jury would be too easily swayed by the presence of a similar accent. The second voice provided the jury with the ability to compare ‘the putative voice of the appellant’ but also ‘to compare and distinguish (if they could) the admitted and putative voice of the appellant with the voice of “Tran”’ (the alleged co-offender).<sup>53</sup>

In addition, Anderson J explained that it ‘might have been better’ if the trial judge:

had specifically warned the jury against an uncritical acceptance of [CN’s] evidence that the voice of the speaker ‘Huy’ was the same voice as the voice admitted by the appellant to be his, but the failure to give a direction along those lines did not cause the trial to miscarry.<sup>54</sup>

Desirable as some kind of warning might have been, in the end the Court applied the proviso to overcome the *failure* to provide the jury with any indication of the dangers.<sup>55</sup> The Court, like other appellate courts<sup>56</sup> writing on voice comparison, seems to imply that a mono-lingual jury will already be aware of many, and perhaps most, of the problems and risks associated with voice identification and

---

<sup>51</sup> Ibid 371 [134].

<sup>52</sup> Ibid 372 [140] (emphasis added).

<sup>53</sup> Ibid.

<sup>54</sup> Ibid 371 [136].

<sup>55</sup> That is, ‘there was no substantial miscarriage of justice for the purposes of s 689 of the *Criminal Code*’: *Nguyen* 131 A Crim R 341, 361 [84], and see also 375 [149].

<sup>56</sup> See, eg *Bulejcik* (1996) 185 CLR 375.

comparison. The trial would have been fair (or fair enough) had the jury been told not to approach CN's evidence 'uncritically' and to 'bear in ... mind' that in undertaking any comparison they were listening to voices speaking in a foreign language.<sup>57</sup>

### ***B R v Korgbara***

In the more recent case of *Korgbara*,<sup>58</sup> the New South Wales Court of Criminal Appeal had the opportunity to address directly the specific question of whether it is acceptable to ask a jury to engage in a cross-lingual voice comparison exercise. Ozone Korgbara was convicted of being knowingly concerned with the importation into Australia of a trafficable quantity of cocaine. During the trial, the Crown tendered, over objection, a series of incriminating telephone intercepts recorded from a mobile phone Korgbara had purchased. The majority of these calls were in Igbo, a Nigerian language, with one call in English to the NRMA. All of these calls were played to the jury, and the jury also had access to the intercepts in the jury room. The defendant conceded he had made the call to the NRMA, but denied that he was a participant in any of the Igbo calls. The appellant gave evidence during the trial in English, though it does not seem to have been in dispute that he could speak Igbo. When the telephone intercepts were admitted, the expectation was that the jury would be asked to compare the voice in the Igbo calls with the voice of the appellant speaking in English in the NRMA call. After Korgbara had given evidence, the jury was also invited to make comparisons between the voice of the accused speaking in court and the recorded Igbo calls.

The primary defence objection to the admissibility of the intercepts, and thus the comparison exercises that the jury was invited to undertake, was that no expert evidence had been tendered supporting the conclusion that the voice in the Igbo calls was the same as the voice in the NRMA call. Further, the defence had not been allowed the opportunity to call their own expert to comment on the comparison task it was anticipated that the jury would undertake.<sup>59</sup> These objections were taken up on appeal, with counsel for the appellant referring to the

---

<sup>57</sup> Both Malcolm CJ and Anderson J expressed confidence in the overall strength of the case against Nguyen, with Anderson J concluding that 'the circumstantial case against the appellant was so strong that, putting the aural identification evidence to one side, it was inevitable that the appellant would be convicted.' *Nguyen* (2002) 131 A Crim R 341, 374 [148]; see also 359 [76], 361 [84], 375 [149].

<sup>58</sup> (2007) 71 NSWLR 187. Special Leave refused: *Korgbara v The Queen* [2007] HCA Trans 485 (31 August 2007).

<sup>59</sup> The defence also requested that the use of the NRMA call be limited, under the *Evidence Act 1995* (NSW) s 136, so that it could only be used to prove that the accused had used the phone on the day in question: *Korgbara* (2007) 71 NSWLR 187, 190 [9].

expert evidence admitted in *Nguyen*,<sup>60</sup> or alternatively the expert acoustic analysis offered in *R v Chenia*<sup>61</sup> and *R v O'Doherty*,<sup>62</sup> as providing the type of foundation required. The appellant also argued that the directions given to the jury were inadequate in relation to both comparison exercises.

In summing up, the trial judge noted that voice similarity exercises are 'notoriously open to mistake' and told the jury that it must exercise 'special caution when undertaking the task that is asked of you'.<sup>63</sup> It seems that the defence did not, at least at trial, request additional or more detailed directions in relation to these general points. More specifically, the trial judge referred to the difficulties of this voice comparison exercise, pointing out that the samples that the jury had were limited in a number of ways, noting that they did not have before them any material which allowed them to make a direct comparison between the accused speaking Igbo and the voice(s) in the Igbo calls. Outlining the competing arguments, the trial judge repeated the Crown assertion that they could make comparisons between the pitch and the tone of the voice, as well the manner of speaking: 'what the Crown Prosecutor described as a "relaxed, laid back voice"'. She then referred to the defence argument that no additional expert evidence had been tendered and to the defence's argument that they had 'been set an impossible task': in effect they were being asked to compare 'apples and oranges'.<sup>64</sup> However, the most revealing aspect of the judge's summing is where she outlined the different ways the jury could use the voice evidence:

Ultimately you may decide that you cannot be satisfied beyond reasonable doubt by voice comparison alone, that you can positively identify that it is the accused speaking on the IGBO [sic] calls. If you come to the view that it sounds like the accused's voice then you can use that evidence, together with other evidence which I will take you to on which the Crown relies, to satisfy you beyond reasonable doubt that it is the accused's voice on the IGBO calls and thus he was knowingly concerned in the importation. On the other hand if you came to the view that the voice on the IGBO calls nominated as the receiver, or usually nominated as the receiver, didn't sound anything like the accused's voice, well then you would put the calls to one side. On the way this case has been run the Crown Case would fail and the accused would be acquitted because the other evidence standing alone would not be sufficient to

---

<sup>60</sup> (2002) 131 A Crim R 341.

<sup>61</sup> [2003] 2 Cr App R 6.

<sup>62</sup> [2003] 1 Cr App R 5 ('*O'Doherty*').

<sup>63</sup> *Korgbara* (2007) 71 NSWLR 187, 192-3 [20].

<sup>64</sup> *Ibid* 193 [20].

satisfy you beyond reasonable doubt that the accused was knowingly concerned in the importation.<sup>65</sup>

The chief difficulty for the appellant was that he was proposing that the Court of Criminal Appeal should impose a threshold requirement; arguing that cross-lingual jury comparisons must be accompanied by expert testimony linking the samples in question. Given the liberalisation in NSW, this was always going to be a difficult argument. In response, McColl JA, with whom James J agreed, explained that since *Bulejck*,<sup>66</sup> Australian courts have avoided imposing mandatory conditions on the admission of voice identification or comparison evidence. In what are presented as analogous cases, McColl JA noted that juries had been permitted to make comparisons where the voice of the accused was said to be accented or speaking in a language other than English, and that interpreters and other investigators had been allowed to make cross-lingual comparisons.<sup>67</sup> To the extent that some of the judges in *Bulejck*<sup>68</sup> do suggest a more restrictive approach, or raise concerns about the difficulties inherent in voice comparisons (and indeed in relation to issues such as investigative bias) these are not taken up in any substantive way. McColl JA preferred Brennan CJ's approach, reducing the test from *Bulejck* to the relatively mechanical question of whether 'the quality and quantity of the material [was] sufficient to enable a useful comparison to be made', with courts of appeal reluctant to depart from the trial judge's assessment.<sup>69</sup>

McColl JA's judgment is striking both in terms of the latitude it grants in relation to the admission of such evidence and the conclusions it draws from the existence of competing expert approaches to voice identification and comparison evidence. Reiterating that the UEL does not impose any specific controls on the admission of voice identification evidence, she concluded that 'it is not open to this court to establish a prescriptive rule that voice comparison evidence should only be admitted where supported by expert testimony' and that the UK cases relied on by the defence did not support the restrictive approach proposed.<sup>70</sup> Most problematic, though, is the manner in which the judgment treats the lack of agreement as to the appropriate expert technique for voice analysis not as a reason for exercising caution in relation to such evidence, but instead as a further reason to avoid imposing any restrictions on the admission of voice comparison evidence *at all*. In what seems to be a counter-intuitive move, the

---

<sup>65</sup> Ibid (emphasis altered).

<sup>66</sup> (1996) 185 CLR 375.

<sup>67</sup> Eg *R v Leung* (1999) 47 NSWLR 405.

<sup>68</sup> (1996) 185 CLR 375.

<sup>69</sup> *Korgbara* (2007) 71 NSWLR 187, 208 [79] quoting *Bulejck* (1996) 185 CLR 375, 394-5.

<sup>70</sup> (2007) 71 NSWLR 187, 207 [74].

judgment turns expert disagreement as to the most reliable method for voice identification and comparison into a reason to allow completely untrained lay listeners to undertake such comparisons instead.<sup>71</sup>

While the majority's decision might appear inevitable, given the difficulties associated with a threshold test, *Korgbara* includes a strong, cautionary dissent.<sup>72</sup> Drawing on Nicholson LJ's judgment in *O'Doherty*,<sup>73</sup> Grove J points to the opinion of Dr Nolan in that case: 'Auditory identification of speakers known to untrained listeners, contrary to popular belief, yields high error rates even under ideal listening conditions.'<sup>74</sup> While accepting that the UEL does not place restrictions on comparisons between voices speaking in English, Grove J explained that any such permissiveness should not necessarily extend to cross-lingual comparisons:

In my view, permitting the comparison of one language with a different language without suitable material which I would contemplate as evidence of someone either possessing relevant expertise or familiar with the voice of the accused in the language used where identity is challenged (an 'ad hoc' expert) is not to establish a prescriptive rule but, to the contrary, to extend the scope of what is permissible beyond recognised boundaries.<sup>75</sup>

Grove J emphasises that the UEL should not be interpreted as if it 'establish[ed] a statutory scheme governing the admissibility of voice identification evidence without restriction'. He points out that s 9 of the Act 'expressly preserves the common law'.<sup>76</sup> Consequently, it is open to judges to impose restrictions or conditions if juries (and presumably others) are being asked to engage in cross-lingual comparisons. Noting that '[i]t is self-evidently not a commonplace human experience to recognise a speaker's voice in a language other than that which is otherwise familiar', Grove J was apprehensive that without being properly informed as relevant differences between languages, jurors making cross-lingual comparisons would be relying on 'mere guesswork'.<sup>77</sup>

As discussed further below, we are not suggesting that Grove J's reliance on the use of (*ad hoc*) expert opinion evidence is the best solution to the difficulties of cross-lingual jury comparisons. But he is correct in identifying that the presumptions made in the majority

---

<sup>71</sup> Ibid 208 [78].

<sup>72</sup> Ibid 209 [111].

<sup>73</sup> [2003] 1 Cr App R 5.

<sup>74</sup> *Korgbara* (2007) 71 NSWLR 187, 209 [112] quoting *O'Doherty* [2003] 1 Cr App R 5, 79.

<sup>75</sup> *Korgbara* (2007) 71 NSWLR 187, 210 [113].

<sup>76</sup> Ibid 210 [114].

<sup>77</sup> Ibid 210 [118]–[119].

judgment about the capacities of juries to engage in meaningful or reliable comparisons between voices are simply not borne out and that courts can and should seek to inform themselves about the characteristics of the languages in question, and the possibility of meaningful comparisons, and if warranted, place some restrictions on such exercises. Thus, in the section below, we offer an overview of empirical research that might afford some guidance to assessments of the probative value of voice comparison and identification evidence, both in relation to the two cases described above, and more generally.

#### IV Scientific Research Applied to Forensic Voice Comparisons

While Brennan CJ in *Bulejck* correctly states that recognition of a speaker by the sound of his or her voice is a ‘commonplace of human experience’,<sup>78</sup> this statement as a summary of human voice identification *performance* is far from complete. When considering the frequency with which voices are correctly identified (or ‘hit’) one must also consider the number of times an unfamiliar voice is mistaken for the voice of someone known (a ‘false alarm’) because it is errors of this kind which may contribute to wrongful convictions. As it turns out, based on the empirical evidence, false alarms are commonplace human experiences, which have important implications for the probative value of voice identification evidence in the courts.

The accuracy<sup>79</sup> with which someone can correctly name a person based on hearing their voice and comparing it to memory is largely driven by the concept of ‘familiarity’. Although the precise threshold for familiarity is difficult to isolate (generally exposures lasting many minutes and hours are required, rather than seconds) the evidence suggests that identification of the voices of family, colleagues, famous people and some acquaintances can be reasonably accurate, even in demanding circumstances.<sup>80</sup> Overall, despite substantial variability in the reported experiments, depending on the specific methodology employed, correct identification or ‘hit’ rates higher than

---

<sup>78</sup> (1996) 185 CLR 375, 381.

<sup>79</sup> Unless otherwise specified the term ‘accuracy’ refers to the overall correct performance of participants in a study. This includes the correct identification of a target and the correct rejection of a foil stimuli.

<sup>80</sup> Diana Van Lancker, Jody Kreiman, and Karen Emmorey, ‘Familiar Voice Recognition: Patterns and Parameters – Part I: Recognition of Backward Voices’ (1985) 13 *Journal of Phonetics* 19; Diana Van Lancker, Jody Kreiman and Thomas D Wickens, ‘Familiar Voice Recognition: Patterns and Parameters – Part II: Recognition of Rate-Altered Voices’ (1985) 13 *Journal of Phonetics* 39. In one influential study, friends and family were named with 66 per cent accuracy based on a single sentence: Peter Ladefoged and Jenny Ladefoged, ‘The Ability of Listeners to Identify Voices’ (1980) 49 *University of California, Los Angeles Working Papers in Phonetics* 43.



80 per cent for ‘familiar’ voices are not uncommon. The same is not true, however, for the identification of unfamiliar or ‘stranger’ voices with some experiments reporting error rates (that is, false alarms) of 51 per cent.<sup>81</sup>

The experimental research indicates that familiars tend to be much more accurate than non-familiars, but that even familiars experience a significant rate of error in the identification of known voices, and accuracy can vary markedly as a result of factors such as health, fatigue, intoxication or emotional state of the speaker.<sup>82</sup> Those not familiar with a voice tend to have relatively high levels of error, and the accuracy for all listeners is affected by the circumstances and conditions in which any comparison (or recollection) exercise is undertaken. Where familiarity has not been established through *varied* exposure, a wide range of other factors has been shown to affect the accuracy of voice identifications. Recognition of previously heard voices is less accurate if the tone or pitch of the voice has been altered,<sup>83</sup> the quality of the speech is poor (eg if heard through a telephone, whispered, or part of a low quality recording),<sup>84</sup> if the speech duration<sup>85</sup> or exposure time is short<sup>86</sup> and, to some extent, as the delay between exposure and identification increases.<sup>87</sup> Correct identification rates of incidentally heard voices have at times been

- 
- <sup>81</sup> This was after hearing 30 to 70 seconds of a previously unknown voice. Accuracy (hit) rates were lower, at 42 per cent: José H Kerstholt et al, ‘Earwitnesses: Effects of Speech Duration, Retention Interval and Acoustic Environment’ (2004) 18 *Applied Cognitive Psychology* 327; see also, Brian R Clifford, ‘Voice Identification by Human Listeners: On Earwitness Reliability’ (1980) 4 *Law and Human Behavior* 373; A Daniel Yarmey et al, ‘Commonsense Beliefs and the Identification of Familiar Voices’ (2001) 15 *Applied Cognitive Psychology* 283.
- <sup>82</sup> Francis Nolan, ‘Forensic Speaker Identification and Phonetic Description of Voice Quality’ in William J Hardcastle and Janet Mackenzie Beck (eds), *A Figure of Speech: A Festschrift for John Laver* (Lawrence Erlbaum Associates, 2005) 385.
- <sup>83</sup> Tara L Orchard and A Daniel Yarmey, ‘The Effects of Whispers, Voice-Sample Duration, and Voice Distinctiveness on Criminal Speaker Identification’ (1995) 9 *Applied Cognitive Psychology* 249; Howard Saslove and A Daniel Yarmey, ‘Long-Term Auditory Memory: Speaker Identification’ (1980) 65 *Journal of Applied Psychology* 111.
- <sup>84</sup> Yarmey et al, above n 81; Olaf Köster and Niels O Schiller, ‘Different Influences of the Native Language of a Listener on Speaker Recognition’ (1987) 4 *Forensic Linguistics* 18; Orchard and Yarmey, above n 83.
- <sup>85</sup> Susan Cook and John Wilding, ‘Earwitness Testimony: Effects of Exposure and Attention on the Face Overshadowing Effect’ (2001) 92 *British Journal of Psychology* 617.
- <sup>86</sup> Susan Cook and John Wilding, ‘Earwitness Testimony: Never Mind the Variety, Hear the Length’ (1997) 11 *Applied Cognitive Psychology* 95; Lori R Van Wallendael et al, ‘“Earwitness” Voice Recognition: Factors Affecting Accuracy and Impact on Jurors’ (1994) 8 *Applied Cognitive Psychology* 661; A Daniel Yarmey, ‘Earwitness speaker identification’ (1995) 1 *Psychology, Public Policy, and Law* 792.
- <sup>87</sup> José H Kerstholt et al, ‘Earwitnesses: effects of accent, retention and telephone’ (2006) 20 *Applied Cognitive Psychology* 187.

shown to peak at 49 per cent after a delay of one week, only to decline to approximately eight per cent after three weeks.<sup>88</sup> Conversely, additional speech variety,<sup>89</sup> contextual consistency and distinctiveness have been associated with improved voice identification accuracy.<sup>90</sup>

### *A A Simple Voice Comparison*

#### *Exercise?*

The judges in *Nguyen*<sup>91</sup> and the majority in *Korgbara*<sup>92</sup> devote little attention to the actual issue of cross-lingual comparisons and comparisons in foreign languages, with McColl JA in particular approaching the issue as merely an extension of what had gone on in previous cases. Thus, such comparisons become part of a continuum of ordinary (that is, 'commonplace') human experience. Anderson J in *Nguyen* also implies that because it is a straightforward voice *comparison* exercise, few of the regular dangers (such as those associated with memory, delay or different media) apply.<sup>93</sup> The decisions thus avoid any substantive consideration of the difficulties of cross-lingual comparison, particularly when comparing voices across different media, the potentially corrosive effects of contextual bias, and the synergistic or potentially buttressing effect of the voice evidence when considered in the context of other evidence in the trial (discussed further below).

Voice comparison across languages either by familiars or by non-familiars (for example, jurors and judges) is not the same as the already highly variable process of familiar voice identification or unfamiliar voice comparison within languages. Familiar voice identification relies on both idiosyncrasies of expression as well as memory for the acoustical properties of the voice. Both of these types of cues to identity are under threat in the case of cross-language identification, particularly when the second language is not familiar to, or spoken by, the listener making the comparison (as is true for the jurors in *Korgbara*<sup>94</sup> and *Nguyen*<sup>95</sup>). Ultimately, cross-lingual identifications (ie from incriminating Igbo or Vietnamese recordings to non-incriminating English speech) are only possible if there are sufficient language-independent cues which remain constant across

---

<sup>88</sup> Clifford above n 81.

<sup>89</sup> Daniel Read and Fergus I M Craik, 'Earwitness Identification: Some Influences on Voice Recognition' (1995) 1 *Journal of Experimental Psychology: Applied* 6.

<sup>90</sup> Ibid; Saslove and Yarmey, above n 83.

<sup>91</sup> (2002) 131 A Crim R 341.

<sup>92</sup> (2007) 71 NSWLR 187.

<sup>93</sup> (2002) 131 A Crim R 341.

<sup>94</sup> (2007) 71 NSWLR 187.

<sup>95</sup> (2002) 131 A Crim R 341.

both languages. Potential cues of this type include sex, age and physiology of the vocal and nasal structures.<sup>96</sup> Even so, performance on cross-lingual comparisons can be highly variable with identification accuracy ('hit') rates peaking at 70 per cent and dropping to 12 per cent depending on the conditions. More significantly, some studies report a false alarm rate (that is, where an 'innocent' foil is misidentified) of above 67 per cent.<sup>97</sup>

Further, in both cases the voice comparisons were based on different types of phone calls and speech. Experimental evidence suggests that the content and tone of expression across speaker samples can be important for the accurate identification of voices. When a voice is incidentally heard and the memory test is unexpected, recognition accuracy is significantly impaired both by changes in tone and in content.<sup>98</sup> This general pattern of declining performance is also evident in the identification of whispered voices,<sup>99</sup> or where tone is changed from angry or hostile to neutral across exposures, as might be expected when comparing the tone used to arrange some kind of drug deal with, say, a conversation with one's wife, the NRMA, a real estate agent, or indeed when speaking in court.<sup>100</sup> Tonal variation and the subsequent impairment of identification accuracy across languages is not inevitable, although it is highly likely in cases where comparisons are being made between voices speaking tonal languages (eg Vietnamese and Igbo) and non-tonal languages (eg English), where the variable use of tone across languages has the potential to undermine the diagnosticity of what is otherwise a fairly stable and therefore useful cue to identity.<sup>101</sup> Voice identification tends to be much more accurate in a language spoken by the listener.<sup>102</sup> Moreover, identification of familiar voices is often also facilitated by the idiosyncratic use of language, such as unique greetings, turns of phrase or nicknames. The extent to which these cues can be relied upon across languages depends both on the language and the specific nature of the idiosyncrasies themselves.

In this context, there is a real risk that alleged similarities and absence of differences will be equated, or closely associated with, identity. In *Korgbara*, the jury was told that it would be sufficient,

---

<sup>96</sup> The problem is that the person of interest tends to be unknown.

<sup>97</sup> Axelle C Philippon et al, 'Earwitness Identification Performance: The Effect of Language, Target, Deliberate Strategies and Indirect Measures' (2007) 21 *Applied Cognitive Psychology* 539; Judith P Goggin et al, 'The role of language familiarity in voice identification' (1991) 19 *Memory & Cognition* 448.

<sup>98</sup> Read and Craik, above n 89.

<sup>99</sup> Orchard and Yarmey above n 83.

<sup>100</sup> Saslove and Yarmey above n 83.

<sup>101</sup> John W Mullennix et al, 'Earwitness Memory: Distortions for Voice Pitch and Speaking Rate', (2010) 24 *Applied Cognitive Psychology* 513

<sup>102</sup> Philippon et al, above n 97.

taking into account the other evidence, that the voice on the tape, speaking Igbo, ‘sounds like’ the accused speaking English.<sup>103</sup> In *Nguyen*, Malcolm CJ was emboldened by the perceived failure of the defence, explaining:

It was not suggested that the jury did not have ample opportunity to listen to the recordings of the relevant calls. Counsel for each side had the opportunity to point out similarities or dissimilarities to the jury.<sup>104</sup>

In both cases, the court seems to imply that similarities (or lack of differences) suggest proof of identity, just as differences might tend to suggest another speaker. The problem is, and this problem is accentuated where languages are unfamiliar (and we are ignorant of differences, however subtle or manifest), that apparent similarities might not be particularly discriminating. In the absence of information about voice properties and features of a language, it might be that apparent similarities in one voice (or differences in another voice), and even popular expressions or tendencies, do not assist with discrimination or identification. Similarities do not simply equate with identity, and impressions of similarities and dissimilarities are likely to be mediated by other evidence and opinions, though not necessarily in a simple or conscious manner.

***B Contextual Cues (for  
Investigators and  
Compounded for Juries)  
and Synergistic Effects***

None of the identifications in the cases discussed were undertaken in neutral conditions. CN was part of an investigation and knew about the main suspects. The juries in both cases undertook their analysis in possession of a range of incriminating evidence. Just as idiosyncratic information helps to facilitate the accurate ‘recognition’ and identification of previously heard voices, such contextual information can also facilitate the perception that a correct ‘match’ has been made between two voice samples that did not originate from the same source. Context effects are particularly relevant when interpreting ambiguous information (such as deciding whether two voice samples in different languages originate from the same speaker) and may make the listener more likely to perceive a match, even if the

---

<sup>103</sup> *Korgbara* (2007) 71 NSWLR 187, 193 [20] (emphasis altered).

<sup>104</sup> *Nguyen* (2002) 131 A Crim R 341, 360 [78].

contextual information is, in fact, inaccurate. The term ‘confirmation bias’ describes the situation where people are inclined to interpret evidence in a manner consistent with their expectations, rather than at face value. These expectations may arise as a consequence of being presented with information about a case which supports a particular interpretation of the evidence, but which is not actually necessary for that interpretation. In *Nguyen*,<sup>105</sup> a large amount of this type of peripheral information was available to the interpreter who was asked to perform a comparison between the suspect and the target voices (much of which was subsequently available to the jury when asked to undertake its own comparison). Specifically, CN was aware of extraneous similarities between the suspect and the perpetrator, of the existence of a warrant for surveillance (also applicable in *Korgbara*<sup>106</sup>), and probably other incriminating information through his interactions — including informal discussions — with investigating police. In the context of any comparison, this information, which implies that the source and target voices share a common origin, makes it much more likely that the person comparing the voices, guided by these cues, will confirm the clearly held suspicion — that two voices share a common origin — rather than contradict that suspicion.<sup>107</sup> This response is likely even if the suspicion is mistaken — that the voices are from different sources. Extraneous cues make it more likely that the interpreter or the jury will *perceive* a match irrespective of whether the recorded voices are from the same person.<sup>108</sup>

In addition to the contextual information, in *Nguyen*<sup>109</sup> the opinions of interpreters and jury comparisons presumably interacted (with each other and other evidence). That is, the jury was asked to make a comparison after being told the ‘answer’. While the jury is to weigh all of the evidence, if some evidence dramatically skews the other evidence in a way that might not be entirely attributable to its probative value then asking juries to make comparisons in such circumstances is undesirable.

---

<sup>105</sup> (2002) 131 A Crim R 341.

<sup>106</sup> (2007) 71 NSWLR 187.

<sup>107</sup> See *R v Harris (No 3)* [1990] VR 310.

<sup>108</sup> For a demonstration of this phenomena in the forensic sciences see Itiel E Dror et al, ‘When Emotions Get the Better of Us: The Effect of Contextual Top-Down Processing on Matching Fingerprints’ (2005) 19 *Applied Cognitive Psychology* 799; Itiel E Dror, David Charlton and Ailsa E Péron, ‘Contextual Information Renders Expert Vulnerable to Making Erroneous Identifications’ (2006) 156 *Forensic Science International* 74; D Michael Risinger et al, ‘The *Daubert/Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion’ (2002) 90 *California Law Review* 1. In addition, the simple fact that the accused speaks the same language as the source may be misread by the jury as having some evidentiary force, rather than, as it is, a precondition for his (or her) inclusion within the class of suspects.

<sup>109</sup> (2002) 131 A Crim R 341.

In situations where jurors are being informed by the testimony of an expert there will undoubtedly be some form of synergistic effect whereby their perceptions of the facts in evidence will be, to some extent, changed. It does not *necessarily* follow that fact finders, having heard the expert's view, will abdicate their responsibility to critically evaluate all evidence. Research on juror perceptions of experts and their testimony demonstrates that jurors will thoughtfully consider many dimensions of the message they are receiving, including: 'expert qualifications, use of good reasoning, familiarity with the facts of the case, and appearance of impartiality, as well as the party who retained the expert'.<sup>110</sup> Empirical research indicates that decision makers do have regard to the complexity of the evidence presented, the internal consistency of that evidence and the completeness of the experts' account.<sup>111</sup> But despite these positive signs there is also empirical evidence that the critical evaluation of expert opinion evidence, and its persuasiveness, is moderated by several factors which have no necessary bearing on the value of the evidence itself. One example is assessment of expert credibility, a complex construct incorporating a number of factors including believability, credentials, likeability and confidence.<sup>112</sup> Moreover, where expert testimony has *already* been presented describing evidence as a match, this, in combination with these other factors, will effectively undermine the likelihood that juror evaluations of that same evidence will be independent of, or unfettered by, the earlier opinion or judgement. These concerns about expert influence are accentuated in circumstances where the task confronting the decision maker is difficult.<sup>113</sup> Further, given that jury verdicts are the outcome of group deliberation, it is entirely possible that the endorsement by some jurors of a persuasive, though factually inaccurate expert (and/or the interpretation of the voices), will

---

<sup>110</sup> Daniel W Shuman and Bruce D Sales, 'The Admissibility of Expert Testimony Based Upon Clinical Judgment and Scientific Research' (1998) 4 *Psychology, Public Policy, and Law* 1226, 1243.

<sup>111</sup> Sanja Kutnjak Ivković and Valerie P Hans, 'Jurors' Evaluation of Expert Testimony: Judging the Messenger and the Message', (2003) 28 *Law & Social Inquiry* 441.

<sup>112</sup> Robert J Cramer, Stanley L Brodsky and Jamie DeCoster, 'Expert Witness Confidence and Juror Personality: Their Impact on Credibility and Persuasion in the Courtroom' (2009) 37 *Journal of the American Academy of Psychiatry and the Law* 63. While there is some indication that confident witnesses are perceived as more credible, the relationship is not linear, suggesting a complex process of interplay between confidence and credibility in juror decision making.

<sup>113</sup> R Kemp, S Heidecker and N Johnston, 'Identification of Suspects from Video: Facial Mapping Experts and the Impact of their Evidence' (Paper presented at the 18<sup>th</sup> Conference of the European Association of Psychology and Law, Maastricht University, 4-5 July 2008). Participants deferred to the interpretation of the image comparison *expert* even when that assessment was *factually wrong*, hinting at the potential influence of prosecution experts.

influence and indeed alter the opinions held by jurors who were originally unconvinced by the testimony.<sup>114</sup>

## V Some Legal Issues and Implications

These issues, informed by published scientific research raise serious problems for legal practice and conventional ways of legal knowing. In this section we consider traditional trial practice and safeguards while drawing upon the insights from the scientific research described in the previous section before concluding with an analysis of the efficacy of directions and warnings for the jury.

### *A Lack of Engagement with Relevant Scientific Literatures*

Perhaps the most conspicuous feature of the appellate decisions discussed above, and characteristic more generally of appellate decisions in this area, is the almost complete absence of relevant scientific research and specialist literatures.<sup>115</sup> Cases are prosecuted and defended at trial, and convictions reviewed on appeal, in courts apparently deprived of relevant scientific knowledge. There appear to be a range of professional/ideological, institutional and cognitive limitations on legal practice that prevent access to relevant exogenous knowledge and expertise. Such engagement is exceedingly rare and not facilitated within existing cultures, traditions and procedures.<sup>116</sup> Consequently, decisions about prosecution, defence and guilt were all

---

<sup>114</sup> See the overview in Solomon E Asch 'Opinions and Social Pressure' in Elliot Aronson (ed) *Readings About the Social Animal* 13 (W H Freeman, 7<sup>th</sup> ed, 1995) 13.

<sup>115</sup> Most of our discussion of scientific literature has been based on experimental work in psychology. There are, however, other groups involved in studying language and sound. The following is taken from the of the International Association for Forensic Phonetics and Acoustics *Code of Practice* (2004) [6]:

(a) Members should exercise particular caution if carrying out forensic analysis of any kind on recordings containing speech in languages of which they are not native speakers.

(b) In carrying out forensic speaker identification/elimination work, Members should exercise particular caution if the samples for comparison are in different languages.

Without wanting to promote qualitative voice comparison (by linguists and phoneticians), where professional bodies caution their members courts should not encourage such individuals (investigators) or juries to fill evidentiary gaps.

<sup>116</sup> See Gary Edmond and David Hamer, 'Evidence Law' in Peter Cane and Herbert M Kritzer (eds), *The Oxford Handbook of Empirical Legal Research* (Oxford University Press, 2010) 652.

undertaken on the basis of prevalent, though sometimes mistaken, views about the ordinariness and accuracy of voice comparison and identification. There are few attempts to distinguish, and no apparent recognition of, the especial difficulties of cross-lingual comparison as compared to ‘ordinary’ (and error-prone) listening to unfamiliar speakers for the purpose of identification. There are no serious attempts to identify or engage with relevant scientific literatures of individuals with genuine expertise in voice comparison, as opposed to the impressions of investigators. All ‘knowledge’ about the various opinions and comparison exercises seems to have been derived from earlier legal decisions and/or trial and appellate judges’ own experience.<sup>117</sup> In supporting the jury’s cross-lingual voice comparison the various judges draw on earlier decisions involving identification evidence more generally, voice comparisons in English, and even the jury’s access to trial exhibits.<sup>118</sup> There is no evidence that scientific literature or knowledge directly or overtly influenced legal assumptions, practice or decision-making.

Though institutionally orthodox, this focus on legal authority diverts attention from more informative exogenous knowledge that could assist with the development and application of rules and practices in ways that are consistent with system values and aspirations — particularly factual rectitude.

### ***B Admissibility of Opinions of Interpreters, Translators and Investigators***

Cross-lingual jury comparisons may, as in *Nguyen*,<sup>119</sup> be supplemented by the opinions of translators, interpreters, investigators and even linguists treated as experts in voice comparison. The admissibility of CN’s incriminating opinion was not addressed at trial or substantially on appeal. Nevertheless, both Malcolm CJ and Anderson J were confident about CN’s ability to express that opinion on the basis of qualifications and experience as an interpreter in conjunction with the large number of intercepted calls he had been exposed to during a two-month investigation.

There was no doubt about CN’s ability or expertise as an interpreter. However, whether this extended to voice comparison and identification (as opposed to translation), whether there is a ‘field’ of

---

<sup>117</sup> See eg *Nguyen* (2002) 131 A Crim R 341, 370–1 [134]; 374 [145].

<sup>118</sup> *Ibid* [138]–[139] citing *Bulejcik* (1996) 185 CLR 375, *R v Smith* (1990) 50 A Crim R 434 and *Barker v The Queen* [2002] WASCA 127.

<sup>119</sup> (2002) 131 A Crim R 341.



voice comparison, and whether CN was an expert in that ‘field’ are completely different issues. There is substantial slippage between expertise in interpretation and translation and expertise in voice comparison (and discrimination). Rather than focus on relevant qualifications, experience or the existence of a field (or reliability), the Court in *Nguyen* in effect qualified CN as an *ad hoc expert*, placing emphasis on the amount of material that CN had listened to and the fact that there was no delay involved in the comparison.<sup>120</sup> However, absent a demonstrably reliable method or capacity to make meaningful comparisons, there is no place for non-experts to be afforded the privileges of expertise for reasons of expediency or convenience, especially where the accused bears the non-trivial risk of error.<sup>121</sup>

There is, in addition to the *expertise* of any interpreter (or police officer or linguist), an issue flowing from the combination of the admission of *expert* opinion evidence and allowing *the jury* to make its own comparison.<sup>122</sup> If juries are capable of reliably making voice comparisons — because jurors form part of the ‘commonplace of human experience’ or because they are ‘as capable of forming a correct view as anyone else’ — then why is opinion evidence (from an interpreter) required or tolerated?<sup>123</sup> While it might be argued that CN had much greater exposure than the officers in *Smith* and also spoke Vietnamese, this does not answer the question of why the jury were allowed to undertake the task as well — especially given the dangers.<sup>124</sup>

### *C The Practical Weakness of Trial Safeguards (in the Absence of Scientific Knowledge)*

Trial safeguards are far less likely to be effective, particularly in the quotidian trial and appeal, than most lawyers, judges and

---

<sup>120</sup> Ibid. See Edmond and San Roque above n 32.

<sup>121</sup> See Edmond et al above n 4.

<sup>122</sup> Jurors are not permitted, in contrast, to make their own comparisons based on fingerprints or electropherograms.

<sup>123</sup> In some ways this resembles the issue in *Smith v The Queen* (2001) 206 CLR 650, where a majority of the High Court characterised the opinions of police officers, with some limited exposure to Smith, to be irrelevant. See also *Dasreef Pty Ltd v Hawchar* (2011) 277 ALR 611.

<sup>124</sup> However, we do not mean necessarily to endorse the majority position in *Smith v The Queen* (2001) 206 CLR 650 nor suggest that we should be allowing investigators to make cross-lingual voice comparisons. See below n 136.

commentators tend to appreciate (or concede).<sup>125</sup> Trial safeguards tend to be weak whether individually or in combination. For a variety of reasons, they are unlikely to expose or convey problems with incriminating opinion evidence and problems are unlikely to be appreciated by relevant audiences (ie juries and appellate courts).

In the absence of actual knowledge about voice comparison evidence and the problems created by cross-language comparisons, the ability to cross-examine witnesses, such as CN, robustly, is substantially compromised.<sup>126</sup> Consider CN's response to questions about his voice evidence.

Counsel: Why do you say that — let's just take the example of the person called Huy — the person speaking is always Huy? What particular characteristics identify him as Huy?

CN: Well, sir, if you listen to the same voice every day for two or three months I think you have some guts feeling. I don't have to be an expert, sir, but I know that's him.<sup>127</sup>

In response to *ipse dixit* all the examiner can do is try to challenge the credibility of the *expert* witness — allowed by the court to testify — or adduce evidence, for example rebuttal experts, that might throw doubt on the reliability of their interpretive practices. Once an incriminating opinion is admitted, or a voice comparison exercise facilitated, the accused becomes responsible for identifying and exposing limitations, during the course of the accusatorial proceeding where any move they make, and any witness or evidence they call, tends to have a motivated appearance.

There are, in addition, no substantial references to the mandatory and exclusionary discretions in *Nguyen*<sup>128</sup> or *Korgbara*.<sup>129</sup> This is revealing because in these cases *expert* comparison evidence and cross-lingual voice comparisons were expected to be undertaken by lay jurors where the probative value of the evidence was unknown, the risk of error high, and the dangers associated with a range of biases and exaggerated confidence, significant and unlikely to be conveyed to the jury. Yet exclusionary powers designed to protect the accused from the danger of unfair prejudice lay idle.<sup>130</sup>

<sup>125</sup> More generally, see Gary Edmond and Andrew Roberts, 'Procedural Fairness, the Criminal Trial and Forensic Science and Medicine' (2011) 33 *Sydney Law Review* 359.

<sup>126</sup> See Anderson J, *Nguyen* 368 [124].

<sup>127</sup> *Nguyen* 373 [144] and discussion in 374[145], 355 [59] and [60] (emphasis added).

<sup>128</sup> (2002) 131 A Crim R 341.

<sup>129</sup> The focus in *Korgbara* (2007) 71 NSWLR 187 is on *Evidence Act 1995* (NSW) s 136 rather than either s 135 or s 137.

<sup>130</sup> In relation to *Korgbara* (2007) 71 NSWLR 187 at least, this is consistent with the regrettably limited utility of these sections in assessing "probative value" following decisions such as *R v Shamouil* (2006) 66 NSWLR 228. See also Smith and Odgers, above n 14.

## VI Directions, Instructions, and Warnings to the Jury

The problems with trial safeguards spill over into what have become the primary mechanisms for protecting against problematic forms of admitted evidence: judicial instructions, directions and warnings to the jury. In *Nguyen*<sup>131</sup> and *Korgbara*,<sup>132</sup> the appellate courts accepted the admissibility of the opinions of interpreters and the jury comparisons, while emphasising the desirability of cautionary directions. But, especially when considered against the empirical work discussed above, both cases provide an indication of just how weak legally orthodox directions and warnings can be.

In *Nguyen*, Anderson J explained that it would have been preferable if the jury had been told to ‘bear in mind’ that they ‘were listening to speakers in a foreign language’ when attempting their comparison and warned against ‘an uncritical acceptance of [CN’s] evidence’.<sup>133</sup> In *Korgbara* the jury was told to exercise ‘special caution’ when undertaking its task.<sup>134</sup> However, in comparison to what the scientific literatures reveal, if these represent the appropriate ‘directions’ and ‘warnings’ then they are effectively meaningless. They neither provide information about important influences like confirmation bias, nor information that might enable notorious dangers to be sensibly incorporated into any assessment of the voice evidence. Moreover, it is impractical to warn those involved in voice comparison exercises of the dangers of cues, some of which may operate unconsciously. Simple knowledge of the dangers does not enable a juror (or expert) to transcend them. Similarly, there is no warning capable of undoing the potentially unconscious effects that other information — such as an *expert’s* opinion, knowledge about ownership of the phone, or an awareness of the existence of a warrant for an intercept — will exert on any voice comparison.

Even purportedly sound directions, such as those offered in *Korgbara*, do not provide any indication of the actual effects of contextual factors; just how corrosive comparisons and recollections can be; how limited exposure radically reduces accuracy; how tone and type of speech and recording type influence accuracy; the very high risk of error; the way witness confidence is often misleading; how witness variability might apply in the specific circumstances; how witness interactions and investigator confirmation may produce

---

<sup>131</sup> (2002) 131 A Crim R 341.

<sup>132</sup> (2007) 71 NSWLR 187.

<sup>133</sup> *Nguyen* (2002) 131 A Crim R 341, 372 [140], 371 [136].

<sup>134</sup> *Korgbara* (2007) 71 NSWLR 187 appears in the discussion on admissibility and warnings Judicial Commission of New South Wales, *Criminal Trial Courts Bench Book* (at 12 August 2011) [3.100]–[3.110].

(mistaken) consensus and inflate levels of confidence; and how even the most subtle cues from honest investigators can contaminate virtually any identification.

As things stand, jurors are somehow expected to ‘take into account’ or ‘consider ... most carefully’ a range of factors without information on how such factors might influence accuracy whether individually or collectively. There is an assumption that mere advertence is enough to discharge the obligation with a type of evidence that is demonstrably prone to error, and far less accurate than most jurors and judges seem to appreciate (even after (giving) conventional warnings). Lay persons and experts tend to dramatically underestimate the ability of the process, suggestion or even prior information, to shape their interpretations (and analyses). This (along with an awareness that the accused speaks the relevant language) is important, particularly for jury comparisons undertaken in conjunction with exposure to other incriminating evidence.

In trying to follow conventional directions, how should the jury ‘take into account’ the dangers of cross-lingual or other voice comparison evidence? How should they respond to the need for caution? What is the appropriate response to a cross-language comparison exercise where none of the listeners speaks one (or any) of the languages? How are they to allow for the fact that interpreters, police officers, and most linguists are not voice comparison experts and when they have no information about their abilities or accuracy? If they were told, how would they evaluate or incorporate the insidious effects of suggestion (and other biases)? How are they to exercise caution meaningfully with respect to their own assessment of the voices?

In addition, where witnesses are accepted by the court as experts, whether through formal qualifications or through experience, as ad hoc ‘experts’, the warnings about the risks and errors associated with visual and aural identification more broadly may not be given in relation to the *expert* opinion evidence (as in *Nguyen*<sup>135</sup>) even though they will always be lurking. Rather, juries are likely to be told that there are dangers with experts and that the decision is for them. They are not always told that the individuals expressing opinions do not have validated methods. They may not appreciate the significance of this oversight, nor that lay and ‘expert’ witnesses may not be able to do what they claim, and that some of the witnesses have no relevant expertise.<sup>136</sup>

---

<sup>135</sup> (2002) 131 A Crim R 341.

<sup>136</sup> There is little evidence that police, translators and interpreters and even linguists perform significantly better than average or are particularly accurate at comparisons across the many different conditions confronting earwitnesses and listeners.

There is no expectation that judges will explain every relevant aspect of contested identification evidence in every case. It will ordinarily suffice for a trial judge to canvas the issues broadly, drawing attention to potential limitations, based on legal experience rather than scientific study.<sup>137</sup> There are, revealingly, few judicial references to suggestion and contamination, even though empirical research suggests that these can have incredibly pernicious effects. This means that investigators and witnesses of undoubted integrity can be sincerely mistaken if the evidence is not collected with sensitivity to the risk of contamination.<sup>138</sup> Likewise, witness or juror confidence about similarities and identity may have little or no correspondence with accuracy.<sup>139</sup> Judicial statements rarely warn in these terms and almost never recognise the corrosive potential of such apparently innocuous interactions.<sup>140</sup>

In addition to problems with the content and form of directions and cautionary warnings, the overwhelming majority of available empirical studies suggest that jury directions, instructions and warnings are largely ineffective.<sup>141</sup> Even if judges could provide detailed and scientifically-predicated directions, the empirical research suggests that it would be difficult to understand and apply them to the particular evidence, especially in the overall context of the trial. In the absence of evidence of efficacy and without relevant detail (such as base rates),<sup>142</sup> drawn from relevant and publicly available scientific research, jury instructions and warnings are likely to be vague imprecations. Their primary protection being, perhaps, instrumental. They function to make a trial formally fair, from the perspective of an appellate court, but there is genuine doubt about their ability to assist juries rationally to assess incriminating voice evidence.

---

<sup>137</sup> See *R v Haidley and Alford* [1984] VR 229, 230; *R v Callaghan* (2001) 4 VR 79, 97–8 [35].

<sup>138</sup> Helen M Paterson, Richard I Kemp and Jodie R Ng, ‘Combating Co-Witness Contamination: Attempting to Decrease the Negative Effects of Discussion on Eyewitness Memory’ (2011) 25 *Applied Cognitive Psychology* 43.

<sup>139</sup> Eg discussion in Carolyn Semmler et al, ‘Effects of postidentification feedback on eyewitness identification and nonidentification confidence’ (2004) 89 *Journal of Applied Psychology* 334.

<sup>140</sup> Kemp et al, above n 113.

<sup>141</sup> See New South Wales Law Reform Commission, above n 2; Queensland Law Reform Commission, above n 2; Victorian Law Reform Commission, above n 2.

<sup>142</sup> Base rates provide a frame of reference which aids in the interpretation of the significance of the presence or absence of particular features or characteristics, such as a feature of speech, that are the subject of an opinion or used in an identification exercise. Examples might include the frequency of a particular feature within a population and the co-occurrence of those features.

## VII Summing up (and Excluding)

The lack of attention paid to the need for substantive controls on the admission of voice identification and comparison evidence is troubling on many levels. It is important to emphasise that where courts are not conversant with limitations and admit incriminating opinion evidence or allow questionable procedures, regardless of the justification, in effect the accused bears the risks. The accused, through counsel, is somehow to overcome — that is, expose and *effectively convey* — limitations with incriminating expert opinion evidence or voice comparisons that are not demonstrably reliable and are known, at least beyond the courts, *to be extremely error prone*. As *Nguyen*<sup>143</sup> and *Korgbara*<sup>144</sup> demonstrate, it is unlikely that lawyers, trial judges and appellate courts will independently appreciate these problems or their significance.<sup>145</sup>

The proliferation of audio technologies and thus the availability of voice recordings to the courts, combined with the need to provide the jury with assistance, should not drive the admissibility of inexperienced opinions of unknown reliability, nor necessitate abandoning such error-prone tasks to lay fact-finders in the context of an accusatorial trial, on the assumption that in both situations safeguards and directions will cure problems. Even in their most detailed and concrete guises, directions, instructions and warnings tend to present highly abstract information to jurors and even technically and epistemologically sound directions tend to be far less efficacious than any safeguard could credibly claim to be. That being the case, it is not sufficient for judges to abdicate their *gatekeeping* function in relation to evidence and procedures of unknown probative value and reliability in the hope that a ‘properly instructed jury’ is best placed to manage the difficulties and frailties of this evidence. In many cases, including the circumstances in *Nguyen*<sup>146</sup> and *Korgbara*,<sup>147</sup> if there is no demonstrably reliable means of analysing the voices then recordings should not be presented to juries for purposes of comparison and identification. If juries are incapable of reliably undertaking cross language voice comparison exercises then voice recordings should not be played at trial. Cross-lingual voice comparison procedures should be abandoned and even voice comparisons in English should be approached cautiously.<sup>148</sup>

---

<sup>143</sup> (2002) 131 A Crim R 341.

<sup>144</sup> (2007) 71 NSWLR 187.

<sup>145</sup> See National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward* (National Academies Press, 2009), which was highly critical of the reliability of most forensic ‘science’ evidence.

<sup>146</sup> (2002) 131 A Crim R 341.

<sup>147</sup> (2007) 71 NSWLR 187.

<sup>148</sup> None of this would prevent the jury considering matters such as ownership of the phone, practices around the use of the phone, names used and discussion of materials that might imply identification. Exclusion of the voice identification evidence would

In concluding, having explained the importance of exclusion, it is appropriate to reiterate that the strength of the case or the existence of other evidence should not ground the admission of expert opinions or voice comparison exercises for the jury. The strength of the case does not inform the reliability (i.e. ability and accuracy) of an *expert's* opinion (or technique) and is likely to contaminate any comparison exercise if known to the listeners. In-trial voice comparisons are actually unfairly prejudicial because they make it *much more likely* that even independent and impartial listeners will mistakenly attribute a voice to the accused. That is, *the listener is likely to associate the incriminating voice with the accused regardless of whether the voice actually belongs to the accused*. High levels of error compounded with the dangers created by highly suggestive procedures and evidence (such as the impressionistic evidence of interpreters and other incriminating evidence) mean jury voice comparisons, particularly cross-lingual comparisons, are risky and prejudicial in a way that is not susceptible to correction and likely to produce irreparable unfair prejudice.

---

not have prevented a prosecution in either *Nguyen* (2002) 131 A Crim R 341 or *Korgbara* (2007) 71 NSWLR 187, although it would have saved court time and prevented substantially unfair procedures. Evidence would still be presumptively admissible even if the translator's beliefs about the names of any speakers should be suppressed. Jurors can be presented with relevant transcripts and real voice comparison experts — to the extent that they exist — might be allowed to testify.