2015

Earwitness Evidence: The Reliability of Voice Identifications

Christopher Sherrin

Follow this and additional works at: http://digitalcommons.osgoode.yorku.ca/olsrps

Part of the Courts Commons

Earwitness Evidence: The Reliability of Voice Identifications

Christopher Sherrin

Abstract:
This article discusses the reliability of non-expert voice identification evidence. While much attention has been paid to the frailties of eyewitness evidence, little attention has been given to the frailties of 'earwitness' evidence, even though it has been tendered in several wrongful conviction cases. The author reviews the results of the empirical literature that has examined the reliability of earwitness evidence. The author also analyzes the principal factors used by Canadian criminal courts to assess earwitness reliability in light of the empirical study of those factors. The general conclusions are that earwitness evidence can often be quite unreliable and that the courts have not always properly assessed its reliability.

Keywords:
Earwitness, voice identification, wrongful conviction

Author(s):
Christopher Sherrin
University of Western Ontario
E: csherrin@uwo.ca

Earwitness Evidence: The Reliability of Voice Identifications

Christopher Sherrin·

This article discusses the reliability of non-expert voice identification evidence. While much attention has been paid to the frailties of eyewitness evidence, little attention has been given to the frailties of 'earwitness' evidence, even though it has been tendered in several wrongful conviction cases. The author reviews the results of the empirical literature that has examined the reliability of earwitness evidence. The author also analyzes the principal factors used by Canadian criminal courts to assess earwitness reliability in light of the empirical study of those factors. The general conclusions are that earwitness evidence can often be quite unreliable and that the courts have not always properly assessed its reliability.

On November 3, 2006, two armed men, one white, one black, invaded the southwestern Ontario home of Duane and Deirdre Hicks.[1] They tied up the occupants, demanded money, and ransacked the house for electronics. The black intruder even took Mr. Hicks to the bank to withdraw money. The police quickly charged Joseph Webber with the robbery, alleging that he was the white intruder. That intruder wore a balaclava during the robbery but the area around his eyes remained exposed. At Webber's trial, Mr. Hicks testified that he was able to recognize Webber by his eyes, the way he walked, and his voice. Hicks had the opportunity during the robbery to see and hear the white intruder when he was only a few feet away. Hicks had known Webber for many years, had occasionally socialized with him, and two years earlier had hired him to do work on the Hicks' house. Hicks also testified that the black intruder called his partner Joe and that the white intruder mentioned he had a drug debt to pay, something that was of interest given that two weeks before the robbery Webber had asked Mr. Hicks for a loan to repay a drug debt. Based on this evidence, the trial judge convicted Webber and sentenced him to seven-and-a-half years' imprisonment. The only problem was that Webber was innocent. The

[1] Most of the information in this paragraph was obtained from the facta filed in the Court of Appeal for Ontario by the Crown and defence in *R v Webber*, Court File No.C48841.

police themselves discovered this by happenstance when another man named Justin Parry confessed to the crime in June 2008. The police corroborated the confession, located the accomplice (who also identified Parry as the white intruder), and found no connection between Webber and Parry. The Crown eventually asked the Court of Appeal for Ontario to acquit Webber. The Court did so on January 7, 2010, commenting that "the Crown acknowledges that the fresh evidence overwhelmingly shows the appellant did not commit these crimes."[2]

Joseph Webber's case exemplifies the familiar frailties of eyewitness identification but it also highlights another, much less discussed phenomenon: the frailties of voice identification, sometimes called 'earwitness' evidence. Earwitness evidence is less common than eyewitness evidence, and it is often heard in combination with eyewitness evidence, but it can be important and even critical in criminal prosecutions. It is routinely needed when the Crown relies on evidence of intercepted communications, as in many conspiracy and narcotics trials. It can also be crucial in cases where the perpetrator wore a disguise, when the victim was blindfolded or otherwise unable to make reliable observations, or where the offence was committed over the telephone. The evidence is admissible in Canadian courts under the general rules of admissibility. When tendered through a lay witness,[3] it is not considered opinion evidence and there are no special preconditions to admissibility.[4]

The contribution of mistaken eyewitnesses to wrongful convictions has been well documented. The contribution of mistaken earwitnesses has not. But Joseph Webber is not the

---

[2] 2010 ONCA 4 at para 1.
[3] Expert evidence of voice identification is sometimes tendered but it is beyond the scope of this paper. I will be dealing solely with non-expert identification evidence.
[4] *R v Williams* (1995), 98 CCC (3d) 160 (Ont CA); *R v Braumberger* (1967), 62 WWR 285 (BCCA). The only precondition, if one can call it that, is that the Crown must lay a foundation for the witness' ability to identify the voice (by adducing evidence that the witness had more than fleeting exposure to the voice outside of the criminal encounter in circumstances where the witness could connect the voice to a particular person): *R v Portillo* (2003), 176 CCC (3d) 467 at paras 40-42 (Ont CA).

only victim of an incorrect voice identification. At Guy Paul Morin's second trial, the victim's mother identified Morin's voice as the one that uttered "Help me, help me, Oh God, help me" on the night of the victim's funeral − testimony that the prosecution used as evidence of consciousness of guilt. The Commissioner examining Morin's wrongful conviction concluded that the voice identification evidence was "patently unreliable."[5] At least 17 individuals have been wrongly convicted in the United States based in part on faulty voice identification evidence.[6] In at least five cases, the voice identification evidence was critical to the prosecution's case. One of England's oldest documented wrongful convictions was the product of mistaken voice identification. William Hulet was convicted of high treason based substantially on the identification of his voice as that of the masked executioner who hanged King Charles I. Another individual (the regular hangman) subsequently confessed and Hulet was exonerated.[7]

As all of this illustrates, voice identification is not infallible. In fact, those who have studied its reliability have variously characterized the identification task, at least as it relates to unfamiliar voices, as "remarkably difficult"[8] and "error prone"[9] where performance is "very

---

[5] *The Commission on Proceedings Involving Guy Paul Morin: Report*, Hon Fred Kaufman, Commissioner (Toronto: Ontario Ministry of the Attorney General, 1998) at 978. Ivan Henry may be another Canadian example. The Crown's case against him certainly included voice identification evidence: *R v Henry*, 2010 BCCA 462 at paras 114-132. The case was also very weak, an alternative perpetrator has been identified, and Henry is sometimes included in the list of Canada's wrongly convicted: e.g. Joan McEwen, *Innocence on Trial: The Framing of Ivan Henry* (Canada: Heritage House Publishing Company, 2014). But it is arguable that he has not been proven factually innocent.

[6] See the names and sources listed in Appendix A. As is so often true, there may be some debate about the factual innocence of some of these individuals but I limited my list to cases where there was fairly strong, even if not always indisputable, evidence of innocence.

[7] Lawrence Solan and Peter Tiersma, "Hearing Voices: Speaker Identification in Court" (2003) 54 Hastings LJ 373 at 393-394, citing *Trial of William Hulet*, 5 Howell's State Trials 1185 (1660).

[8] Sarah Stevenage, Amy Howland and Anna Tipplet, "Interference in Eyewitness and Earwitness Recognition" (2011) 25 Applied Cognitive Psychology 112 at 112 [hereinafter Stevenage, "Interference"].

[9] A. Daniel Yarmey, "The Psychology of Speaker Identification and Earwitness Memory" in *The Handbook of Eyewitness Psychology, Volume II: Memory for People*, R.C.L. Lindsay et al, eds (Mahwah, New Jersey: Lawrence Erlbaum, 2007) 101 at 102 [hereinafter Yarmey, "Psychology of Speaker Identification"].

poor,"[10] "extremely inaccurate"[11] and "likely to produce high false identifications."[12] This may seem surprising, given the common experience we all have of easily recognizing voices. But care must be taken not to generalize from experience with familiar voices in everyday situations to assumptions about identification accuracy with less familiar voices in other contexts.[13] Indeed, one should not even assume that identification of familiar voices is inevitably accurate.[14]

In recent years, Canadian criminal courts have demonstrated an appreciation of both the fallibility of earwitness testimony and some of the reasons for it. However, their analysis has been informed by experience, intuition, and assumption rather than empirical study. That is not to say that the courts have always got it wrong; on the contrary, they have often got it right. But their assessment of earwitness testimony can only be improved by an understanding of the results of the scientific study of voice identification. The point of this paper is to offer some assistance in that regard. I discuss the results of the empirical studies. I compare the ways in which the approach of the courts has coincided with and deviated from those results. I offer some thoughts on how courts could alter and ultimately improve their assessment of earwitness testimony.

1. The (In)Accuracy of Voice Identification

As stated above, many researchers studying voice identification have expressed concerns over its reliability. This stems in large part from the fact that studies have so often produced very low accuracy rates. One study, for example, obtained a "hit rate" of 9%, meaning that only 9% of

---

[10] Lisa Öhman, Anders Eriksson and Pär Anders Granhag, "Angry Voices from the Past and Present: Effects on Adults' and Children's Earwitness Testimony" (2013) 10 Journal of Investigative Psychology and Offender Profiling 57 at 59.

[11] Howard Saslove and A. Daniel Yarmey, "Long-Term Auditory Memory: Speaker Identification" (1980) 65 Journal of Applied Psychology 111 at 114.

[12] A. Daniel Yarmey, A. Linda Yarmey and Meagan Yarmey, "Face and Voice Identifications in Showups and Lineups" (1994) 8 Applied Cognitive Psychology 453 at 463 [hereinafter Yarmey, "Face and Voice Identifications"].

[13] Brian Clifford, "Voice Identification by Human Listeners" (1980) 4 Law and Human Behavior 373 at 374.

[14] See, *infra*, section 3(a).

the test subjects were able to correctly recognize a voice they had previously heard (the target voice) from amongst a group of other voices in a voice lineup.[15] Some other studies have obtained hit rates, collapsed across experimental conditions, of 12%,[16] 13%,[17] 15-20%,[18] 19-20%,[19] and 24%.[20] It is not uncommon for hit rates to be lower than the percentage of test subjects who incorrectly indicate that the target voice is not present in the lineup[21] and/or who choose the voice of a lineup foil.[22]

Of even greater concern is the "false alarm rate" obtained in numerous studies: the percentage of test subjects who, when presented with a set of voices, incorrectly identified someone other than the speaker they had previously heard. An inaccurate positive identification, after all, can lead to the prosecution of the wrong person. Test-subjects routinely select foils in a target-present lineup (one that includes the previously-heard voice).[23] More importantly, they very often believe they recognize a voice in a target-absent lineup, i.e. a lineup that does *not* contain the previously-heard voice. In such lineups, studies have obtained false alarm rates,

---

[15] Yarmey, "Face and Voice Identifications," *supra* note XXX at 458. The hit rate was 28% (lower than chance) in the voice showup condition where test subjects only heard a single voice and were asked if they recognized it as one they had heard five minutes before.

[16] Öhman, "Angry Voices," *supra* note XXX at 63. This was chance level performance.

[17] Lisa Öhman, Anders Eriksson and Pär Anders Granhag, "Mobile Phone Quality vs. Direct Quality: How the Presentation Format Affects Earwitness Identification Accuracy" (2010) 2 The European Journal of Psychology Applied to Legal Context 161 at 172 [hereinafter Öhman, "Mobile Phone Quality"]. This was chance level performance.

[18] Nils Olsson, Peter Juslin and Anders Winman, "Realism of Confidence in Earwitness versus Eyewitness Identification" (1998) 4 Journal of Experimental Psychology: Applied 101 at 107.

[19] Lisa Öhman, Anders Eriksson and Pär Anders Granhag, "Enhancing Adults' and Children's Earwitness Memory: Examining Three Types of Interviews" (2013) 20 Psychiatry, Psychology and Law 216 at 223.

[20] José Kerstholt et al, "Earwitnesses: Effects of Accent, Retention and Telephone" (2006) 20 Applied Cognitive Psychology 187 at 192 [hereinafter Kerstholt, "Earwitnesses"]. This was the hit rate when test subjects were allowed to indicate they did not know if the target was present amongst the voices (which 24% of them indicated). When subjects were forced to make a selection, the hit rate increased to 34%.

[21] See, e.g., Öhman "Enhancing," *supra* note XXX at 224; Olsson, *supra* note XXX at 106-107; Öhman, "Angry Voices," *supra* note XXX at 63; Lisa Öhman, Anders Eriksson and Pär Anders Granhag, "Overhearing the Planning of a Crime: Do Adults Outperform Children as Earwitnesses?" (2011) 26 Journal of Police and Criminal Psychology 118 at 123.

[22] See, e.g., Öhman, "Enhancing," *supra* note XXX at 224; Olsson, *supra* note XXX at 106-107; Yarmey, "Face and Voice Identifications," *supra* note XXX at 458; Öhman, "Angry Voices," *supra* note XXX at 63; Öhman, "Overhearing the Planning," *ibid* at 123.

[23] See all the sources cited *ibid*.

across experimental conditions, of 99%,[24] 98%,[25] 85%,[26] 43-93%,[27] 70%,[28] 67%,[29] 56%,[30] 53%,[31] and around 50%.[32] Those figures do not translate to the likelihood that in a real-life situation an innocent suspect will be identified, since many of the false alarms would be foil identifications ignored by the police, but they do illustrate the propensity of earwitnesses to mistakenly identify voices. Furthermore, studies that pre-selected a particular voice to serve as the innocent suspect in a voice lineup have still found substantial "misidentification rates," such as 13-14%,[33] 19%[34] and 22%.[35] Those rates were *over and above* all the foil identifications.

---

[24] Lori van Wallendael et al, "'Earwitness' Voice Recognition: Factors Affecting Accuracy and Impact on Jurors" (1994) 8 Applied Cognitive Psychology 661 at 666. Some of the test subjects had heard the target voice only a short time before the lineup was administered. All test subjects were cautioned (albeit subtly) that the target might not be present in the lineup.

[25] Yarmey, "Face and Voice Identifications," *supra* note XXX at 458. Test subjects were exposed to the voice during an innocuous encounter that lasted about 15 seconds. Testing occurred about five minutes later. Subjects were warned that the target might not be present in the lineup.

[26] Harriet Smith and Thom Baguley, "Unfamiliar voice identification: Effect of post-event information on accuracy and voice ratings" (2014) 5 Journal of European Psychology Students 59 at 62 and 64. Some of the test subjects were given misleading information about the pitch of the target voice prior to selecting a voice from the lineup, but the misinformation had no effect on overall accuracy.

[27] Sarah Stevenage, Gabriella Clarke and Allan McNeill, "The 'other-accent' effect in voice recognition" (2012) 24 Journal of Cognitive Psychology 647 at 650 [hereinafter Stevenage, "The 'other-accent' effect"].

[28] A. Daniel Yarmey, "Earwitness descriptions and speaker identification" (2001) 8 Forensic Linguistics 113 at 118 [hereinafter Yarmey, "Earwitness descriptions"].

[29] Axelle Philippon et al, "Earwitness Identification Performance: The Effect of Language, Target, Deliberate Strategies and Indirect Measures" (2007) 21 Applied Cognitive Psychology 539 at 545. That is the false alarm rate for voices speaking a familiar language. The rate for voices speaking an unfamiliar language was even higher at 93%. All test subjects had heard the target voice about 30 minutes earlier, were instructed to pay attention to the voice, and were told that the target voice may not be present in the lineup.

[30] Charles Thompson, "A Language Effect in Voice Identification" (1987) 1 Applied Cognitive Psychology 121 at 126.

[31] Öhman, "Overhearing the Planning," *supra* note XXX at 123.

[32] José Kerstholt et al, "Earwitnesses: Effects of Speech Duration, Retention Interval and Acoustic Environment" (2004) 18 Applied Cognitive Psychology 327 at 331 (51%) [hereinafter Kerstholt, "Effects of Speech Duration"]; Kerstholt, "Earwitnesses," *supra* note XXX at 192 (50%). In both studies, the test subjects knew the experiment concerned voice identification and were warned that the target voice might not be in the lineup.

[33] A. Daniel Yarmey, "Earwitness identification over the telephone and in field settings" (2003) 10 Forensic Linguistics 62 at 69 [hereinafter Yarmey, "Earwitness identification"].

[34] Yarmey, "Face and Voice Identifications," *supra* note XXX at 458. Subjects were tested within minutes of hearing the target voice and were told that the voice might not be in the lineup.

[35] Tara 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 at 254. This is the false identification rate collapsed across all experimental conditions but only for non-distinctive voices. The equivalent rate for distinctive voices was 6%. Test subjects were told to listen carefully to the target voice for purposes of later identification. They were also told that the target voice might not be present in the lineup.

Not all studies have obtained poor accuracy rates. Indeed, some have obtained fairly good rates. For specific lineups, one occasionally sees hit rates of or in excess of 90% and false alarm rates of less than 10%.[36] More commonly, one can find hit rates in the 35-65% range and false alarm rates in the 20-40% range, at least in certain experimental conditions (albeit sometimes with poorer rates in other conditions).[37] Some researchers in the 1980s posited an overall accuracy rate of 60-70%,[38] although more recent estimates are much less optimistic.[39] In the end, the variety of results and methodologies in the different earwitness studies probably makes it impossible to come up with an overall, generally applicable, accuracy rate. Accuracy is very much a product of specific factors and specific circumstances. Earwitnesses can be reliable.[40] In a variety of situations, however, they can be quite inaccurate and prone to misidentify. As a result, one must endorse the conclusion of many researchers that courts should exercise great caution and care before placing reliance on earwitness testimony.[41]

2. Voice Identification Evidence in Canadian Law

---

[36] See, e.g., Orchard, *ibid* at 253-254 (perfect scores in one of the 8 minute exposure conditions); van Wallendael, *supra* note XXX at 666 (in two of four target-present lineups, although the authors cautioned that the targets in those lineups had distinctive voices, and the same study also produced extremely high false alarm rates in the target-absent lineups, usually 100%).

[37] E.g., Judith Goggin et al, "The role of language familiarity in voice identification" (1991) 19 Memory & Cognition 448 at 451 (most hit rates between 35% and 65% and false alarm rates lower than 40%); Susan Cook and John Wilding, "Earwitness Testimony: Never Mind the Variety, Hear the Length" (1997) 11 Applied Cognitive Psychology 95 at 101 (overall hit rate of 38% for unfamiliar voices); Yarmey, "Earwitness descriptions," *supra* note XXX at 118 (hit rates of 47% and 55% and false alarm rates of 33% and 38% in target-present lineups); Kersholt, "Effects of Speech Duration," *supra* note XXX at 331 (hit rate of 42% and false alarm rate of 24% in target-present lineup, albeit with a 51% false alarm rate in the target-absent lineup).

[38] Ray Bull and Brian Clifford, "Earwitness voice recognition accuracy" in *Eyewitness Testimony: Psychological Perspectives*, Gary Wells and Elizabeth Loftus, eds. (Cambridge: Cambridge University Press, 1984) at 120. See also Thompson, *supra* note XXX at 125.

[39] See, e.g., Öhman, "Angry Voices," *supra* note XXX at 67: poor performance in the study "concurs with the suggestion that real-life conditions will result in earwitness performance at chance level"; Daniel Read and Fergus Craik, "Earwitness Identification: Some Influences on Voice Recognition" (1995) 1 Journal of Experimental Psychology: Applied 6 at 16: "In more realistic circumstances, … the probability that the correct speaker will be chosen is no better than chance."

[40] Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 102.

[41] See, e.g., Kersholt, "Effects of Speech Duration," *supra* note XXX at 334; Öhman, Angry Voices," *supra* note XXX at 68.

In recent years, Canadian criminal courts have recognized the need to proceed cautiously when it comes to voice identification evidence. In the 2015 case of *R v Dodd*, for example, the Ontario Court of Appeal commented that such evidence "ought to be treated with extreme caution."[42] Courts in other provinces, both trial and appellate, have adopted a similar position.[43]

This speaks well of the Canadian position with respect to earwitness evidence. But four qualifications must be added. First, one can certainly find cases where courts have treated voice identification evidence with something less than "extreme caution." The 2006 case of *R v Campbell* is illustrative.[44] Campbell was charged with robbing a video store. The issue in the case was identity. The store clerk was the only person to give identification evidence. The robber was previously unknown to her and she interacted with him on the date in question for five to ten minutes. A month later she claimed to see him at a local mall. She recognized him by his appearance and his voice. The trial judge cautioned himself regarding the frailties of eyewitness evidence but said nothing about the weaknesses of earwitness evidence. On the contrary, he only used the victim's voice identification to help overcome any weaknesses with her visual identification.[45] On appeal, Campbell claimed that his conviction was unreasonable, in part because the trial judge "gave undue weight to [the victim's] recognition of the appellant's voice as confirming her identification of him."[46] The British Columbia Court of Appeal said nothing about that submission and only used the earwitness testimony to help justify the reasonableness

---

[42] 2015 ONCA 286 at para 79. From the same court, see also *R v Clouthier*, 2012 ONCA 636 at para 19; *R v Quidley*, 2008 ONCA 501 at para 36; *R v Badgerow*, 2008 ONCA 605 at para 9; *R v Brown*, [2003] OJ No 4592 at para 4 (CA).
[43] E.g., *R v Saddleback*, 2013 ABCA 250 at paras 18 and 35; *R v Pabla*, 2013 BCSC 913 at para 23; *R v J.E.*, [2011] NJ No 38 at para 27 (Prov Ct).
[44] 2006 BCCA 109.
[45] *Ibid* at para 35.
[46] *Ibid* at para 68.

of the trial judge's decision.[47] I do not claim that *Campbell* is typical of how courts treat voice identification but it is also not so anomalous as to be extraordinary. A number of other court decisions have demonstrated some insensitivity to the potential weaknesses of earwitness evidence.[48]

The Ontario Court of Appeal has also taken the position that a warning to the jury about the frailties of voice identification evidence is not always mandatory. In *R v Masters*, the Court stated that it was not convinced a caution was mandatory in the case given the "significant body of other evidence at trial capable of implicating the appellant as one of the speakers on the wiretaps."[49] This curious holding suggests that when the trier of law determines that the corroborative evidence of identification is strong the trier of fact need not be warned that the direct evidence of identification may be weak. That not only blurs the roles between the two triers in a case, it also leaves a jury unassisted in assessing the weight to be given to a piece of evidence that the jury may feel is a necessary element in finding proof beyond a reasonable doubt.

Canadian courts also sometimes treat earwitness evidence no differently than eyewitness evidence. One sees this in some of the general cautionary statements. In the 2012 case of *R v Clouthier*, for example, the Ontario Court of Appeal wrote that "voice identification evidence, *like any identification evidence*, ought to be treated with extreme caution."[50] The same court

---

[47] *Ibid* at para 69. Only Finch CJBC, in dissent, addressed this issue but it is safe to assume the majority agreed since they only said they differed from the Chief Justice in their view of how the trial judge used another piece of evidence.

[48] E.g., *R v Mackinaw*, 2010 ABCA 359 (forgiving a trial judge's complete lack of self-caution regarding important earwitness evidence); *R v Wu*, 2010 ABCA 337 at paras 29, 48 and 52 (responding to a concern over the frailties of voice identification evidence by simply noting that the trial judge acknowledged the witness had no training in voice recognition and that the witness was subject to "the usual scrubbing of cross-examination"); *R v Masters*, 2014 ONCA 556 (discussed immediately below).

[49] *Ibid* at para 46.

[50] *Supra* note XXX at para 19 (emphasis added).

more recently commented that "[v]oice identification evidence is even more fraught with dangers than eyewitness identification evidence"[51] but its overall record is mixed. For example, it has not repudiated its pronouncement in 1995 that "[t]here is no pattern of instruction that the trial judge must give to the jury or to himself, if he is the trier of fact, with respect to this type of identification evidence as distinct from the more usual identification evidence given by an eyewitness."[52] One can also see the same approach in the analysis of voice identification evidence in particular cases. *R v Henry*, for example, was a prosecution that depended heavily on voice identification as well as, and possibly even more than, visual identification.[53] In assessing the reasonableness of the verdict, however, the British Columbia Court of Appeal focused solely on the weaknesses of eyewitness evidence, adding nothing specific about earwitness evidence.[54] The Court in *Henry* ultimately declared the verdict unreasonable but, from an analytical standpoint, it was inappropriate to effectively assimilate the two kinds of evidence. Studies have shown that earwitness evidence is even *less* reliable than eyewitness evidence.[55] It must therefore be treated with *greater* caution than eyewitness evidence, not with the same amount.[56]

The last reason why one cannot be overly sanguine about the Canadian position regarding earwitness evidence is that the courts do not always properly assess the variables that affect the reliability of earwitness testimony. General cautions are helpful but what really matters is how courts treat the various factors that relate to the trustworthiness of the specific earwitness evidence in a given case. Courts certainly take into account a number of appropriate factors but

---

[51] Dodd, *supra* note XXX at para 79.
[52] *Williams*, *supra* note XXX at para 22. See, similarly, *R v Rowbotham (No.4)*, (1977) 33 CCC (2d) 411 at para 46 (Ont Co Ct): "the jury must be cautioned with respect to the frailties surrounding such evidence in like manner to the caution normally given in the case of visual identification."
[53] See *supra* note XXX at paras 113-138.
[54] See, especially, *supra* note XXX at paras 77 and 139-141.
[55] E.g., Yarmey, "Face and Voice Identifications," *supra* note XXX; Olsson, *supra* note XXX.
[56] See *R v Pinch*, 2011 ONSC 5484 at para 76.

they do not always do so with the nuance required. They also sometimes rely on unhelpful factors and fail to consider helpful ones. In the next part of this paper, I turn to an analysis of some of the most important factors. I emphasize at the start that, in order to analyze the performance of Canadian courts, I have had to isolate their use of a particular factor from the rest of the evidence considered in a given decision. The reader must keep in mind that courts rarely rely solely on voice identification evidence to convict and that any frailties in the analysis of such evidence may not have had a significant impact on the overall validity of the ultimate verdict.

3. Factors Bearing Upon the Reliability of Voice Identifications

a) Familiarity

The factor most commonly take into account by Canadian courts is almost certainly familiarity. Courts routinely note that an earwitness had significant prior exposure to the accused's voice and thus was in a good position to recognize it during the commission of the crime.[57] Some courts have even divided earwitness evidence into recognition vs. identification evidence in an effort to highlight the presumed reliability of the former and distinguish instances where people are asked to identify the voice of a stranger.[58]

Familiarity is an entirely appropriate factor to take into account. Empirical studies have supported the common sense assumption that significant prior exposure to a voice enhances the

---

[57] E.g., *R v Dobbin*, (2013) 334 Nfld & PEIR 284 at para 32 (Prov Ct); *R v Fitur*, [2011] MJ No 94 at para 36 (QB); *R v Shea*, 2010 NSPC 69 at paras 11 and 15. Courts have also taken into account lack of familiarity: e.g., *Clouthier*, *supra* note XXX at para 20; *Dodd*, *supra* note XXX at paras 80-81.
[58] E.g., *R v Herman*, 2010 BCSC 1068 at para 98; *JE*, *supra* note XXX at para 30.

ability to identify it on a given occasion.[59] Indeed, some studies have obtained results of near perfect accuracy.[60]

It is important, however, not to assume that familiarity necessarily imports accuracy. Familiar voices can be misidentified, sometimes at significant rates. One study obtained a 10% misidentification rate from members of a close social network who had known each other for almost two years and who currently or recently lived in the same residence.[61] In other words, the test subjects, hearing the voices of a series of close friends, incorrectly attributed a voice to the wrong person 10% of the time.[62] The hit rate was 68% and the other 22% of the time the test subjects offered no opinion regarding the identity of the speaker. Another study of immediate family members and best friends obtained a 5% misidentification rate.[63] A judge who resolves the issue of voice identification solely on the basis of familiarity, therefore, proceeds too quickly. The judge also needs to consider the other factors that bear upon reliability. Hearing someone speak only a small number of words, for example, has been shown to affect identification accuracy of even very familiar listeners.[64]

There are also different degrees of familiarity. All the studies mentioned in the preceding two paragraphs used test subjects who had amongst the greatest degree of familiarity with the

---

[59] See Diane van Lancker, Jody Kreiman and Karen Emmorey, "Familiar voice recognition: patterns and parameters" (1983) 13 Journal of Phonetics 19 at 20-21.

[60] E.g., Harry Hollien, Wojciech Majewski and E. Thomas Dogerty, "Perceptual identification of voices under normal, stress and disguise speaking conditions" (1982) 10 Journal of Phonetics 139.

[61] Paul Foulkes and Anthony Barron, "Telephone speaker recognition amongst members of a close social network" (2000) 7 Forensic Linguistics 180.

[62] This was based on hearing about 9 seconds of speech over the telephone. Test subjects were simply asked to listen to a voice and, if they recognized it, to write down the name of the speaker.

[63] A. Daniel Yarmey et al, "Commonsense Beliefs and the Identification of Familiar Voices" (2001) 15 Applied Cognitive Psychology 283 [hereinafter Yarmey, "Commonsense Beliefs"]. As in the Foulkes and Barron study, test subjects were simply asked to listen to a voice and, if they recognized it, to state the name of the speaker. Identifications were made as soon as the listener thought s/he recognized the voice – after hearing anywhere from 1 word to two minutes of speech. The hit rate was 85%.

[64] This issue is discussed, *infra*, in section 3(b). See *R v Meier*, 2012 SKPC 41 at paras 23-26, 49-54, and 93, for an example of a case where earwitness testimony was accepted largely on the basis of familiarity despite the fact that during the crime the witness, who was highly stressed, only heard the perpetrator utter about 15 words.

speakers' voices. The accuracy of test subjects who do not have quite the same degree of familiarity is generally lower. The aforementioned study of immediate family members and best friends, who the study categorized as "high-familiar" subjects, compared them to "moderate familiar" (co-workers, teammates, club-mates and general friends) and "low-familiar" subjects (casual acquaintances who only spoke with the target for a few minutes on occasion in any week over the preceding year).[65] The respective rates were 85%, 79% and 49% for hits and 5%, 13% and 23% for misidentifications.[66] Another study showed that people who considered themselves to be extremely familiar with a target's voice, having heard it for more than one hour, identified it only 31% of the time.[67] It is not clear that Canadian judges have been adequately alive to the significance of different gradations of familiarity. They have sometimes found and relied on familiarity despite fairly vague evidence of the degree of prior exposure.[68] They may also have occasionally slotted a witness into a higher category of familiarity than appears to have been warranted.[69] The exact magnitude of the problem is difficult to discern[70] but hopefully greater familiarity with the relevant empirical literature will encourage proper sensitivity in the future.

---

[65] The study actually referred to the groups as high-, moderate- and low-familiar "speakers" but it is clear that it was the familiarity of the listeners with the speakers' voices that was being measured.

[66] Yarmey, "Commonsense Beliefs," *supra* note XXX at 287-288.

[67] Read and Craik, *supra* note XXX at 11-13. Thirty-one percent is the hit rate averaged across the three experimental conditions.

[68] See, e.g., *Mackinaw*, *supra* note XXX, in which the trial judge found that the victim recognized the accused's voice from an uncertain number of past encounters (possibly only one) at inexact dates in the past; *R v Sanghera*, 2012 BCSC 733 at para 145 (accepting a police officer's identification evidence even though his personal dealings with the accused "had been relatively limited, including, he estimated, two or three face to face meetings"); *R v Lamarche*, 2010 YKTC 28 ("With respect to Mr. Brown's ability to identify Mr. Lamarche's voice over the phone, even though Mr. Brown had difficulty giving specific details about the timing and frequency of his dealings with Mr. Lamarche, I am satisfied, based on his evidence, that he was sufficiently familiar with Mr. Lamarche to be able to recognize his voice"). I remind the reader that comments like these in this paper are only directed at the courts' use of a specific factor (in this case, familiarity) and not at the overall legitimacy of the judgments. It so happens, for example, that the judge in *Sanghera* ultimately acquitted and the judge in *Lamarche* relied on significant corroborating evidence.

[69] See, e.g., *Dobbin*, *supra* note XXX, where the court found that an identifying witness was "very familiar" with the accused's voice after having heard it as little as 10 times over the preceding two or more years during the course of a professional social worker-client relationship. In fairness, it is unclear that the court in *Dobbin*, or courts in any other

Finally, there is the issue of expectation. An "elementary principle of psychology is that context and expectations influence an individual's perceptions and interpretations of what he observes."[71] When circumstances lead people to expect some fact, they tend to perceive that fact in the face of ambiguous information. This can lead to error biased in the direction of the expectation.[72] This phenomenon of expectancy bias happens naturally and quite unconsciously but it can be powerful.[73]

Researchers have suggested that expectancy bias can find application in the context of voice identification. For example, Prof. Yarmey has written:

> Misidentification of a familiar speaker … can occur through witness expectations. Recognition of familiar persons often depends on the closed set of people likely to be encountered in particular settings. People tend to hear who they expect to hear. Thus, if observers (police) expect to hear a particular person answer a telephone, misidentification of a familiar speaker may occur if someone else actually answers the call.[74]

To my knowledge, only one experiment has specifically sought to determine if expectation had an effect on voice identification (finding that it did),[75] but the phenomenon is so well-established[76] that it would be surprising if it did not have an effect. Indeed, some courts have recognized that it is a factor that should be taken into account in assessing earwitness

---

case, used the term 'very familiar' to refer specifically to the highest category of familiarity studied in the psychological literature.

[70] Courts do sometimes explicitly recognize that there are different degrees of familiarity. See, e.g., *J.E.*, *supra* note XXX at para 30 ("in recognition cases the degree and circumstances of past associations between the witness and the accused play a pivotal role in the weight the trier of fact must give to the evidence").

[71] D. Michael Risenger et al, "The *Daubert/Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion" (2002) 90 Cal L Rev 1 at 12.

[72] Keith A. Findley and Michael S. Scott, "The Multiple Dimensions of Tunnel Vision in Criminal Cases" [2006] Wis L Rev 291 at 308.

[73] See Saul Kassin, Itiel Dror and Jeff Kukucka, "The forensic confirmation bias: Problems, perspectives, and proposed solutions" (2013) 2 Journal of Applied Research in Memory and Cognition 42.

[74] Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 116.

[75] Peter Ladefoged, "Expectation Affects Identification by Listening" (1978) 21 Language and Speech 373.

[76] See Kassin, *supra* note XXX at 44.

testimony.[77] The problem is that it is rare to find a case where consideration of expectation actually had an impact, even though the prosecution frequently tenders identification evidence that could have been tainted by expectation. Witnesses, usually (but not always) police officers involved in the investigation of the alleged offences, often testify to voice identifications first made *after* they know the accused has been arrested or singled out as a suspect – such as when they listen to a post-arrest interview and compare it (and only it) to a voice heard on intercepted communications.[78] In such situations, witnesses could be affected by the expectation generated by the arrest or investigation that the voice they will hear is the voice they heard previously.[79] Yet judges have commonly assessed and even accepted such identification evidence without any clear consideration of the potential biasing effect of the context.[80] Even when judges rely on additional independent evidence of identification (as they often do), the weight they give to the potentially tainted evidence should factor in the impact of expectation. Indeed, when an officer familiar with a voice on intercepted communications is asked to listen to only the accused speak to see if it is the same voice, the identification process is tantamount to a showup. Showups in the voice identification context have been shown to be especially prone to error.[81] Less weight

---

[77] E.g., *Saddleback*, *supra* note XXX at para 25; *R v Willis*, 2002 MBCA 138 at para 23, per Twaddle JA, dissenting; *R v Chan*, 2001 BCSC 1180 at para 31.

[78] E.g., *Wu*, *supra* note XXX at para 29 ("Known samples of the appellants' voices were acquired by [Sgt.] Wieschorster in a post-arrest interview"). Many cases have reviewed the constitutionality of the police using the post-arrest procedures to provide an opportunity for voice identification: see *R v Oliynyk*, 2008 BCCA 132 at paras 19-36, and cases cited therein. See *R v Badgerow*, 2010 ONSC 937 at paras 13-15, for an example of a case where lay witnesses, after learning that the accused had been arrested, were asked if they recognized his voice on a tape-recorded phone call.

[79] They could also be affected by the related psychological phenomenon of motivational bias, where a person's perceptions and interpretations are tainted by preference for a particular conclusion. See Kassin, *supra* note XXX at 45.

[80] See, e.g., *R v Alcantra*, 2009 ABQB 524 paras 86-89; *R v Nestorovski*, 2013 ONSC 2523 at paras 32-41.

[81] Two studies have shown that false alarms are significantly higher in showups than in lineups: Yarmey, "Face and Voice Identifications," *supra* note XXX at 459; Yarmey, "Earwitness identification," *supra* note XXX at 70-71. In the first study, the hit rate was lower than that expected by chance.

would normally be given to *eye*witness identification evidence collected in a showup.[82] It is difficult to understand why earwitness evidence should be treated differently.[83]

b) Exposure Length

In assessing the reliability of earwitness testimony, courts have frequently considered the length of time that the witness was exposed to the perpetrator's voice on the date(s) the crime was committed. The British Columbia Supreme Court in *R v Savoy*, for example, held that voice identification had not been proven partly because the witness only heard a "few sentences uttered by the robber in the course of a robbery which he, [the witness], estimates to have had a duration of about a half-minute."[84]

Empirical studies have generally affirmed that exposure length is related to accuracy. The research results have not been entirely consistent but on the whole they show that "[t]he longer the opportunity to listen to a speaker, the greater the accuracy of identification."[85] Unfortunately, there is no magical tipping point, before which voice identifications are not reliable and after which they are. It depends in part on familiarity. If someone has had significant prior exposure to a voice (along with the identity of the person associated with it) there is a good chance, in ideal conditions, that s/he will be able to recognize it after hearing only a sentence or two.[86] If the

---

[82] *R v Miaponoose* (1996), 30 OR (3d) 419 (CA).

[83] In the context of a civilian earwitness, the Ontario Court of Appeal has commented that a showup procedure was "seriously flawed": *Clouthier, supra* note XXX at para 22.

[84] 2000 BCSC 296 at para 35. For two other examples, see *Dodd, supra* note XXX at para 80; *R v Whalen*, 2007 NLTD 79 at para 65.

[85] Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 120. Kenneth Deffenbacher et al have suggested that the results of studies not showing an effect of exposure length may be partly attributable to the fact that most of the studies tested immediate rather than (the more forensically relevant) delayed recall: "Relevance of Voice Identification Research to Criteria for Evaluating Reliability of an Identification" (1989) 123 Journal of Psychology 109 at 111.

[86] I. Pollack, J. Pickett and W. Sumby, "On the Identification of Speakers by Voice" (1954) 26 Journal of the Acoustical Society of America 403; Peter Bricker and Sandra Pruzansky, "Effects of Stimulus Content and Duration on Talker Identification" (1966) 40 Journal of the Acoustical Society of America 1441; Phil Rose and Sally Duncan, "Naïve auditory identification and discrimination of similar voices by familiar listeners" (1995) 2 Forensic

speaker is a stranger to the listener, one does not usually see hit rates in post-event lineups greater than 50% until more (sometimes much more) than a minute of exposure during the event.[87] False alarm rates for unfamiliar voices in target-absent lineups do not seem to change much as exposure increases from around 20 seconds to eight minutes, often hovering around 50%.[88]

Some early research suggested that variability matters more than length, such that people may be able to accurately identify voices based on relatively short speech samples as long as the samples contain sufficient phonemic variety.[89] To some extent this is a distinction without a difference: increased variety will often accompany increased duration. The research findings are also not entirely consistent[90] and the early research may only apply to familiar listeners,[91] so it would be unsafe for courts to ignore exposure length. Certainly, as a practical matter, length is easier for a court to gauge than phonemic variety. The research findings regarding variability are more useful in providing reason to be extra cautious when considering short speech samples of

Linguistics 1. Some early research suggested that familiar listeners can accurately identify voices after hearing less than a single word: e.g. Bricker and Pruzansky, *ibid*. More recent (and methodologically sound) research shows that suggestion to be overly optimistic: see, e.g., Yarmey, "Commonsense Beliefs," *supra* note XXX at 295 (finding only a 30% hit rate and a 30% false alarm rate by familiar listeners after hearing one word). In *R v Herman*, 2010 BCSC 1068, the trial judge accepted that a familiar witness was able to identify a voice based on twice hearing the words "100 times". That may have been overly optimistic.

[87] A. Daniel Yarmey and Eva Matthys, "Voice Identification of an Abductor" (1992) 6 Applied Cognitive Psychology 367 at 370-371 (hit rates close to 30% after 18 and 36 seconds of exposure, close to 50% after two and six minutes); Orchard, *supra* note XXX at 253 (hit rates mostly below 50% after 30 seconds, mostly above 50% after 8 minutes); A. Daniel Yarmey, "Earwitness Identification Over the Telephone" (1991) Journal of Applied Social Psychology 1868 at 1872 (hit rate of 48% after 7.8 minutes but only 24% and 30% after 3.2 and 4.3 minutes) [hereinafter Yarmey, "Over the Telephone"]; Kersholt, "Effects of Speech Duration," *supra* note XXX at 333 (hit rates all above 50% after 70 seconds, half below 50% after 30 seconds). Clifford, *supra* note XXX at 378-379, obtained much better hit rates.

[88] Yarmey and Matthys, *supra* note XXX at 373 (false alarm rates around 50% after 18 seconds, 36 seconds, two minutes and six minutes); Yarmey, *ibid* at 1873-1874 (false alarm rates of 48%, 51% and 44% at 3.2, 4.3 and 7.8 minutes, respectively); Kersholt, "Effects of Speech Duration," *supra* note XXX at 333-334 (overall false alarm rate of 51% across 30 and 70 seconds). Curiously, in Orchard, *supra* note XXX at 254, false alarm rates were sometimes worse after 8 minutes than after 30 seconds.

[89] See, e.g., Pollack, *supra* note XXX at 406; Bricker and Pruzansky, *supra* note XXX at 1444.

[90] See, especially, Susan Cook and John Wilding, "Earwitness testimony: Effects of exposure and attention on the Face Overshadowing Effect" (2001) British Journal of Psychology 617 at 622-623 [hereinafter Cook and Wilding, "Earwitness testimony"].

[91] Cook and Wilding, "Earwitness testimony," *supra* note XXX at 618.

limited variability. In *R v Aulakh*, for example, the evidence disclosed that the perpetrator of a sexual assault repeatedly uttered a single three word phrase during the attack, which the complainant claimed to recognize from encounters with the accused somewhat earlier in the evening. [92] While the corroborating evidence of identification in *Aulakh* was extremely powerful,[93] voice identification evidence of that sort should probably be given no weight.[94]

Another important finding in the research is that people are generally quite poor at estimating the length of time they heard someone speak on a particular occasion, often overestimating it by a significant margin.[95] In many cases, the only evidence of exposure length will come from the earwitness. Courts should generally treat such evidence with scepticism.

c) Retention Interval

A factor listed by some courts as relevant to the assessment of earwitness testimony is retention interval: the amount of time between the criminal event and the witness' closest other exposure(s) to the accused's voice.[96] The supposition is that, all things being equal, the reliability of the testimony decreases as the retention interval increases. In truth, consideration of this factor is not terribly common. While some courts may list it as a relevant factor, it is not often given much emphasis in the analysis.[97] Some decisions address evidence relevant to retention interval

---

[92] 2012 BCCA 340 at para 19.

[93] It suggested that the accused had exclusive opportunity to commit the offence.

[94] The trial judge in *Aulakh* gave the identification evidence some weight, albeit not much. The Court of Appeal held that the evidence of exclusive opportunity was determinative. See *supra*, note XXX at paras 38 and 88.

[95] This may be particularly true for short speech samples. Orchard, *supra* note XXX at 257 (mean estimates for 30 second and 480 second speech samples were 102 seconds and 517 seconds, respectively; 93% of test subjects overestimated the 30 second sample); A. Daniel Yarmey and Eva Matthys, "Retrospective duration estimates of an abductor's speech" (1990) 29 Bulletin of the Psychonomic Society 231 (98% of test subjects overestimated a 72 second speech sample, by an average of 4:1; 51% of test subjects underestimated a five minute speech sample).

[96] *Chan*, *supra* note XXX at para 31; *Pabla*, *supra* note XXX at para 23; *Pinch*, *supra* note XXX at para 75. The other exposure(s) will occur before the criminal event in the case of a familiar voice and after the event in the case of an unfamiliar voice, for example when the witness hears a voice lineup or showup.

[97] It occasionally is. See, e.g., *Saddleback*, *supra* note XXX at paras 14, 26, 29, and 34; *Dodd*, *supra* note XXX at paras 80-82.

mostly or entirely in the context of discussing other issues.[98] Some decisions discount concerns about retention interval.[99]

It is difficult to assess the mixed performance of the courts. The results of empirical studies, which have focused almost entirely on unfamiliar voices, have been somewhat mixed as to the impact of retention interval. Most show no real impact of intervals up to two weeks.[100] The results of the few studies that have investigated intervals longer than that have been inconsistent, with one study finding a significant impact after three weeks[101] and two others finding no or limited impact after three and even eight weeks.[102] To further complicate matters, in the real world intervals longer than two or eight weeks will occur.[103] This is an area where the empirical studies do not offer maximal assistance to the courts. The general assumption is that memory for at least an unfamiliar voice, just like for most other things, tends to decline over time[104] but more research is needed into the effect of intervals of several weeks and months, for both familiar and unfamiliar voices. For now, courts would probably be wise to assume that after a few weeks

---

[98] See, e.g., *Mackinaw*, *supra* note XXX at 4, 10-13, where evidence relevant to retention interval was mentioned during the analysis of whether, rather than when, the earwitness had heard the accused's voice before. Evidence of retention interval is often mentioned during the analysis of familiarity. See, e.g., *Dobbin*, *supra* note XXX at paras 2 and 32. Given the paucity of empirical studies of retention interval involving familiar voices (as discussed below), it is hard to say whether the two issues are substantially distinct.

[99] See, e.g., *Lamarche*, *supra* note XXX at paras 24-26, in which the trial judge found a voice identification to be accurate even though the witness "had difficulty giving specific details about the timing and frequency of his dealings" with the accused. There was significant corroborating evidence but the finding was not dependent on that evidence.

[100] E.g., Öhman, "Angry Voices," *supra* note XXX at 63-64 (no significant effect of a two week interval on adults, although hits by children decreased); van Wallendael, *supra* note XXX at 666 (no significant differences across 0, 7 and 14 day retention periods); Read and Craik, *supra* note XXX at 12.

[101] Clifford 1980, *supra* note XXX at 384-385 (accuracy not significantly different after two weeks vs one week but significantly different – reduced to chance level – after three weeks).

[102] Kerstholt, "Earwitnesses," *supra* note XXX at 193-194 (finding a marginally significant decline in accuracy in a target-present lineup over one, three and eight week intervals but an improved performance in a target-absent lineup after three and eight week intervals versus a one week interval); A. Broeders and A. Rietveld, "Speaker Identification by Earwitnesses" in *Studies in Forensic Phonetics*, Angelika Braun and Jens-Peter Koster, eds (Trier: Wissenschaftlicher Verlag, 1995) 24 at 34 (little difference in performance after three weeks vs. after one week).

[103] See, e.g., *Mackinaw*, *supra* note XXX at para 4, where there was (or at least may have been) an interval of "a few months".

[104] E.g., Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 121.

retention interval has an increasingly detrimental, but not necessarily profound, effect on identification accuracy[105] – albeit one that may be less severe in the case of familiar voices.

d) Confidence

Another factor that has been of some interest to the courts is how confident the earwitness is about his or her identification. Some judgments have been ambivalent about this factor, acknowledging that confidence is not determinative of reliability[106] and that a confident witness can be mistaken.[107] But the courts have not dismissed confidence as irrelevant and it does seem to have factored into some decisions regarding reliability.[108]

This approach to confidence is similar to the approach taken in the eyewitness identification context, where the courts have repeatedly cautioned against placing much weight on confidence.[109] Curiously, the approach of the courts may even be stricter in the eyewitness context, where the Supreme Court of Canada has taken the position that there is a "very weak link between the confidence level of a witness and the accuracy of that witness."[110] This is curious because in the eyewitness identification context empirical studies have shown some (qualified) correlation between confidence and accuracy,[111] whereas the same cannot be said in the earwitness context. Numerous studies have investigated the relationship between confidence

---

[105] See Kersholt, "Earwitnesses," *supra* note XXX at 195.

[106] *Masters*, *supra* note XXX at para 47; *R v Danaii-Asil*, 2011 ONSC 2230 at para 20.

[107] *Saddleback*, *supra* note XXX at para 18. Arguably, in *Pinch*, *supra* note XXX at para 75, the Ontario Superior Court went further in saying that "confidence does not substantially correlate to reliability." The Ontario Court of Appeal in *Masters*, however, seemed to interpret *Pinch* as saying that "witness confidence in voice identification does not make the identification reliable": *supra* note XXX at para 47.

[108] E.g., *R v Sharifi*, [2011] OJ No 3985 at para 13 (SCJ); *Sanghera*, *supra* note XXX at paras 145 & 176. In neither case was confidence considered to be the sole determinant of reliability but it was one factor taken into account.

[109] See, e.g., *R. v. Jack*, 2013 ONCA 80 at para 30 (discussing the importance of "a caution about the diminished correlation between a witness' confidence level and his or her accuracy").

[110] *R v Hibbert*, 2002 SCC 39 at para 52. Arguably, the courts have sometimes even gone further and recommended that no weight should be given to witness confidence. See, e.g., *MacDonald* 2014 ONCA 610 at para 15; *R v Knox*, [2006] OJ No 1976 at paras 54-55 (CA).

[111] See, e.g., Neil Brewer and Gary Wells, "Eyewitness Identification" (2011) 20 Current Directions in Psychological Science 24 at 24-25.

and voice identification accuracy and the general conclusion is that there is very little if any

correlation between the two.[112] The findings have not been entirely consistent but many studies

have found no significant correlation[113] and many others have found only a modest correlation in

only some experimental conditions.[114] In a few instances, studies have even found a negative

correlation, with confident witnesses being *less* likely to be correct than unconfident

witnesses.[115] This is yet another illustration of why earwitness testimony must be treated with

greater caution than eyewitness testimony, not with the same or less caution.[116] In general, the

safest and wisest position for the courts to adopt is to completely disregard confidence as any

indicator of the accuracy of voice identifications.[117]

There may be one qualification to add to the above. There is some evidence that

confidence may have a useful correlation with accuracy when it comes to quite familiar voices.

Two studies found a significant correlation between confidence and accuracy for very familiar

---

[112] Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 110: "Confidence-accuracy correlations in most studies on voice identification for unfamiliar speakers are nonsignificant or are relatively low."

[113] E.g., Öhman, "Overhearing the Planning," *supra* note XXX at 124; Yarmey, "Earwitness descriptions," *supra* note XXX at 119; Goggin, *supra* note XXX at 457; Kerstholt, "Effects of Speech Duration," *supra* note XXX at 335; Kerstholt, "Earwitnesses," *supra* note XXX at 194-195.

[114] E.g., Yarmey, "Earwitness identification," *supra* note XXX at 70 (small but significant correlation in only one of four experimental conditions, with no correlation in the other three); Yarmey, "Over the Telephone," *supra* note XXX at 1871-1872 (modest correlation found only for target-absent lineup, not for target-present lineup); Thompson, *supra* note XXX at 128 (correlation only approached or reached significance when outliers were included); Philippon, *supra* note XXX at 547 (significant correlation found for target-absent lineups but not for target-present lineups or overall); Smith and Baguley, *supra* note XXX at 63 (significant correlation only found for male-voice, target-present lineup; no significant correlation for female-voice or for target-absent lineups); Timothy Perfect, Laura Hunt and Christopher Harris, "Verbal Overshadowing in Voice Recognition" (2002) 19 Applied Cognitive Psychology 973 at 977.

[115] Yarmey and Matthys, *supra* note XXX at 373 (significant negative correlations found in some experimental conditions relating to voice samples of 18 and 36 seconds, although modest positive correlations found in some experimental conditions relating to longer voice samples); Orchard, *supra* note XXX at 256-257 (negative correlations found in experiments involving distinctive voices, although small positive correlations found when results were collapsed across all experimental conditions).

[116] See Olsson, *supra* note XXX at 116: "In comparable information-processing circumstances, the forensic system should place more trust in eyewitness than in earwitness confidence." See also Stevenage, "Interference," *supra* note XXX at 115-117.

[117] See, e.g., Kerstholt, "Earwitnesses," *supra* note XXX at 196.

listeners: family members and close friends.[118] Another study found a correlation when using the voices of generally well known public figures.[119] It may be premature to make too much of these findings, though. Methodological differences make an overall assessment of the studies somewhat difficult.[120] One of the aforementioned studies also obtained inconsistent results across two experiments.[121] And an entirely different study found no significant correlation at all, although that result may be the result of not using any test subjects who were as familiar as family members and close friends.[122] If the courts are to consider confidence when assessing identifications made by familiar earwitnesses, they should only do so in relation to the most familiar category of witnesses.

An interesting question not really addressed in the empirical literature is whether *lack* of confidence should be taken into consideration.[123] Courts certainly consider lack of confidence to be a relevant and sometimes important factor undermining the reliability of identifications,[124] and the position may be defensible simply as a function of the burden of proof.[125] But as matters currently stand we do not truly know whether lack of confidence is indicative of lack of accuracy. It may be that confidence of any sort is entirely unrelated to accuracy. Further assistance from earwitness researchers would be welcome.

---

[118] Rose and Duncan, *supra* note XXX at 14; Yarmey, "Commonsense Beliefs," *supra* note XXX at 289 (only when speakers spoke in a normal tone of voice).

[119] van Lancker, *supra* note XXX at 28-29.

[120] For example, unlike other studies, Rose and Duncan, *supra* note XXX, included several familiar voices (rather than just one) in the lineups, and included more familiar voices than foils.

[121] In the second of two experiments, Yarmey and his colleagues did not find a significant confidence-accuracy correlation for very familiar speakers (at least using one method of analysis). Indeed, the results of the second experiment were, in one respect, the converse of the results of the first, a result the researchers found "difficult to explain": "Commonsense Beliefs," *supra* note XXX at 297.

[122] Read and Craik, *supra* note XXX at 16.

[123] Studies occasionally provide some relevant information but no sustained research into the issue has been undertaken.

[124] See, e.g., *Pinch*, *supra* note XXX at para 75; *J.E.*, *supra* note XXX at para 44,

[125] Although one might ask whether the burden of proof compels a particular response to facts the *value*, rather than the existence, of which is in question.

e) Attention

A factor to which the courts make surprisingly little reference is attention: the degree to which the earwitness made a conscious effort during the crime to pay attention to the characteristics of the perpetrator's voice. The case law lists it as a relevant factor[126] and courts sometimes explicitly consider it.[127] But, given that it could be relevant to essentially every single voice identification case, it is surprising that it is not more often given more explicit consideration.[128]

Judges would be wise to pay more attention to earwitness attention. Several studies have determined that subjects who are told in advance of an experiment that they will be tested on their ability to identify a voice perform better than subjects who are not told.[129] The positive effect of attention has not always been found[130] and it is certainly not the case that a witness who specifically attended to the perpetrator's voice will always be correct (or, conversely, that a witness who did not will always be wrong). But, all things being equal, paying attention to the characteristics of a voice improves the chances of later being able to identify it, and vice-versa.

It may be tempting to apply this conclusion to cases where the perpetrator's voice is captured on tape and someone – be it a police officer, judge or jury member – later compares it to the voice of the accused, on the basis that in those cases attention is being closely paid to the

---

[126] E.g., *Chan*, *supra* note XXX at para 31.

[127] E.g., *Bubar*, *supra* note XXX at paras 94 and 101.

[128] In fairness, courts may sometimes be implicitly considering it, perhaps especially in wiretap cases where officers listen to a sample of the accused's voice for the specific purpose of comparing it to a voice recorded on tape.

[129] Saslove, *supra* note XXX at 113; Yarmey, "Earwitness identification," *supra* note XXX at 69; Heather Armstrong and Stuart McKelvie, "Effect of Face Context on Recognition Memory for Voices" (1996) 123 Journal of General Psychology 259 at 265. Perfect and his colleagues found an effect in the same direction but it did not quite reach statistical significance: *supra* note XXX at 976. See also Clifford 1980, *supra* note XXX at 382-383.

[130] Cook and Wilding, "Earwitness testimony," *supra* note XXX at 621.

characteristics of the voice. That would be unwise. As explained above,[131] 'earwitnesses' in those cases can be affected by expectancy bias. No study on earwitness attention employed a scenario where test subjects were told in advance whose voice it might be (never mind, that it is alleged to be). It is simply not known whether the positive effect of attention would be attenuated or even completely undone by the effect of expectation. Until we know more, the findings relating to the effect of attention should only be applied to cases where earwitnesses attend to the characteristics of a voice *without* previously being told who the speaker could be – or, perhaps more frequently, to cases where earwitnesses do *not* specifically attend to the characteristics of a voice.

f) Abnormal Speaking

Someone perpetrating a crime may not use his or her normal voice during the event. S/he may deliberately attempt to disguise it so as not to be identified. S/he may also speak in a way that is unusual for situational reasons: because of the demands of the event (which require yelling or the use of an angry tone), because of the stress and emotionality of the situation,[132] or because of intoxication.[133] Canadian courts have included disguise as a factor that diminishes the accuracy of voice identification.[134] They have also tried to take into account situation-driven changes in voice.[135] In both respects, they have been correct to do so. However, it may be that courts should be even more cautious.

---

[131] See section 3(a).
[132] Bull and Clifford, *supra* note XXX at 111.
[133] Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 118.
[134] E.g., *Saddleback*, *supra* note XXX at para 25; *Bubar*, *supra* note XXX at para 86; *Whalen*, *supra* note XXX at para 65; *Pinch*, *supra* note XXX at para 75.
[135] E.g., *Saddleback*, *supra* note XXX at para 29; *Bubar*, *supra* note XXX at paras 86 and 93; *Herman*, *supra* note XXX at para 106; *Pinch*, *supra* note XXX at para 75.

Four different studies have investigated the effect of having speakers intentionally disguise their voices in almost any way they wanted. The hit rate for extremely familiar listeners was 79% (leaving a 21% error rate).[136] The hit rate for unfamiliar listeners in one study was 61%[137] but in three others was quite low: subjects comparing a disguised voice with a speaker's normal voice made correct identifications at most one-quarter of the time.[138] Regretfully, methodological issues make it a little difficult to interpret those results. The authors of the study with the highest hit rate for unfamiliar listeners acknowledged that the rate was likely inflated by the fact that the study simply asked listeners to decide whether individual pairs of speech samples came from the same speaker.[139] Conversely, it is possible that the methodology of one of the other studies depressed the hit rates.[140]

That disguise can have a substantial impact on identification accuracy, however, is supported by the results of two other studies that examined the use of one particular form of disguise: whispering. Whispering can act as a disguise because it conceals vocal characteristics such as pitch, inflection, and intonation.[141] When test subjects heard whispered unfamiliar voices

---

[136] Hollien, *supra* note XXX at 142. Test subjects were classified as "extremely familiar" with the target voices when they not only knew the speakers but also correctly identified all ten of the speakers, without any errors, in a pre-test assessment.

[137] Alan Reich and James Duke, "Effects of selected vocal disguises upon speaker identification by listening" (1979) Journal of the Acoustical Society of America 1023 at 1025. This was the hit rate when speakers were permitted to disguise their voices in a manner they thought would best conceal their identity. The hit rates were slightly higher when speakers were directed to use specific forms of disguise, reaching as high as 70% when speakers spoke at an extremely slow rate.

[138] Clifford, *supra* note XXX at 382 (26% hit rate); Hollien, *supra* note XXX at 142 (18%, 17%, 25% and 21%, across four different groups); Read and Craik, *supra* note XXX at 13-14 (22% and 26%).

[139] Reich and Duke, *supra* note XXX at 1027.

[140] Hollien, *supra* note XXX at 141. Listeners were not asked to select a previously heard voice from a voice lineup. Instead, listeners were given pre-test training on the names associated with the ten different unfamiliar speakers and then, in the experiment, later asked to choose the correct name from a supplied list when hearing the voices played on tape. Test subjects may have found it difficult to digest the pre-test information and keep track of the various voice-name associations. That said, subjects were given a preliminary identification test using only undisguised voices and even those who score quite well performed poorly in the actual experiment.

[141] Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 118.

they again generally achieved hit rates less than or equal to 25%.[142] False alarm rates were also high when subjects heard a voice lineup[143] and the misidentification rate was over 50% when subjects were simply exposed to an unfamiliar voice and asked if they could identify who was speaking (the presumption being that they could not).[144] These findings held true even when test subjects were specifically directed to attend to the voice for purposes of later identification,[145] when distinctive voices were used,[146] and/or when exposure length was substantial.[147] Familiar voices did not always substantially change the results. In one study, test subjects hearing very familiar voices achieved a 77% hit rate (and a 15% misidentification rate) but achieved a 35% hit rate and a 39% misidentification rate when hearing moderately familiar voices.[148]

Similar or even worse results have been obtained in experiments where speakers employ tones of voice (anger or emotionality) that may often be used in the commission of crimes. Across six different experiments, the highest hit rate obtained was 33%; most rates were at the level of chance.[149] This was true even when test subjects were informed in advance that it was a voice identification test,[150] were tested immediately after exposure to the voice,[151] and/or were

---

[142] Orchard, *supra* note XXX at 253 (across four experimental conditions, hit rates were 25%, 25%, 17% and 33%, with the highest rate achieved only after subjects heard the whispered voice for eight minutes); Yarmey, "Commonsense Beliefs," *supra* note XXX at 292 (20% hit rate after two minutes of speech).

[143] Orchard, *supra* note XXX at 254 and 256 (53% to 83% across four experimental conditions).

[144] Yarmey, "Commonsense Beliefs," *supra* note XXX at 292 (58% misidentification rate after hearing the whispered voice for two minutes). In Orchard's study using target-absent voice lineups, lineup members specifically chosen to serve as the innocent suspects were selected 12% of the time across the four lineups.

[145] Orchard, *supra* note XXX at 252; Yarmey, "Commonsense Beliefs," *supra* note XXX at 291.

[146] Orchard, *supra* note XXX at 253-254.

[147] Orchard, *supra* note XXX at 253-254 (eight minutes); Yarmey, "Commonsense Beliefs," *supra* note XXX at 287 (two minutes).

[148] Yarmey, "Commonsense Beliefs," *supra* note XXX at 292. A moderately familiar speaker was a co-worker, teammate, club-mate or general friend of the listener. Pollack, *supra* note XXX at 405, found that for familiar voices equivalent hit rates were not obtained until subjects heard speech at least three times as long as a non-whispered sample.

[149] Saslove, *supra* note XXX at 113 and 115 (13%-33% hit rate, rates that were sometimes lower than chance); Read and Craik, *supra* note XXX at 10, 12 and 14-15 (19%, 20% and 22% across three experiments – all not significantly different from chance); Bull and Clifford, *supra* note XXX at 114 (33%); Öhman, "Angry Voices," *supra* note XXX at 64 and 66 (chance level 5%-14%).

[150] Saslove, *supra* note XXX at 113.

exposed to as much as 40 seconds of speech.[152] The one study to record false alarms obtained rates between 43-53%.[153] Even using listeners who were mildly familiar with the speakers did not improve the results.[154]

All the results reported above reflect the accuracy rates achieved when test subjects were asked to compare the voice of an individual using an abnormal voice with the voice of an individual using his or her normal speaking voice. This is important. The available evidence does not indicate that disguised, angry or emotional voices are inherently more difficult to identify than unemotional voices. The generally poor results were more likely a product of comparing normal to abnormal voices. Hit rates in studies where test subjects hear an abnormal voice both initially and at test, for example, were generally (albeit not always) significantly higher.[155] One experiment even specifically tested to see whether voices were differently memorable depending on whether they were emotional, finding no evidence that they were.[156]

This suggests that greater reliance can be placed on identifications by earwitnesses who are exposed to the same kind or tone of voice both during the crime and at another time; the similarity of exposure can assist them in making an accurate comparison and identification.[157] Even then, however, the identification can be suspect. The one study that specifically sought to explore the benefits of similar exposures found that hit rates for both unfamiliar and mildly

---

[152] Saslove, *supra* note XXX at 113; Öhman, "Angry Voices," *supra* note XXX at 64.
[152] Öhman, "Angry Voices," *supra* note XXX at 61.
[153] Öhman, "Angry Voices," *supra* note XXX at 64.
[154] Read and Craik, *supra* note XXX at 11-13. Mildly familiar listeners had previously heard the speaker for more than one hour. The study excluded anyone who was familiar enough with a speaker's voice to identify it on initial exposure.
[155] Read and Craik, *supra* note XXX at 10; Saslove, *supra* note XXX at 113; Orchard, *supra* note XXX at 254 (one experimental condition only).
[156] Read and Craik, *supra* note XXX at 14-15.
[157] Some courts seem to have assumed as much. See, e.g., *J.E.*, *supra* note XXX at paras 13 and 38; *Saddleback*, *supra* note XXX at para 29.

familiar earwitnesses, while improved, still did not usually rise above 50%;[158] the rates were even lower when the two exposures were not perfectly identical.[159] Regretfully, the initial exposure in that study was quite brief (only four seconds long) so the generalizability of its conclusions is uncertain. But even if longer exposures would produce better results (as seems likely), it may not be common in the real world for the various exposures to be similar enough to generate high levels of accuracy. The authorities cannot force a suspect to speak at all,[160] never mind in a particular way, and it may be difficult for them to obtain, through observation of the suspect, a voice sample similar in kind or tone to that used during the commission of the offence, especially if it was a deliberately disguised voice.

g) Telephone Transmission

Courts have been concerned about the possibility that hearing a voice over the telephone can interfere with the ability to accurately identify it. In *R v Garofalo*, for example, it was stated that "[v]oice identification, by itself, is fraught with problems. It is even more uncertain when a voice is identified over the telephone."[161] Commentators have likewise been concerned, hypothesizing that recognition may be impaired by the limited range of sound frequencies transmitted over the phone.[162]

---

[158] Read and Craik, *supra* note XXX at 10 and 13-14. The most familiar listeners had previously heard the speaker's voice for more than one hour. All the other listeners had previously heard the voice, if at all, for at most an hour.

[159] Read and Craik, *supra* note XXX at 13-14. In one variation of the experiments, test subjects were exposed to emotional statements both initially and at exposure but the test statement was a re-recording, by the same speaker, of the original statement rather than a replaying of the original recorded statement. The hit rate never exceeded 38%.

[160] *R v Turcotte*, 2005 SCC 50 at para 52. The suspect also cannot be compelled to participate in a lineup: *Henry*, *supra* note XXX at paras 40-58.

[161] 2012 ONSC 6351 at para 111. See, similarly, *Pinch*, *supra* note XXX at para 75 (referring to "difficulties which the telephone imposes on voice identification because of the loss of acoustic information"); *Whalen*, *supra* note XXX at para 61.

[162] See, e.g., Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 122 ("Speaker identifications involving the use of the telephone may have particular problems to overcome ... Degradations of the speech signal ... are common because telephone lines typically transmit a band of frequencies between 300 and 3,400 Hertz whereas human voices may contain components up to about 12,000 Hertz").

In addition, concerns have also been expressed about distortions introduced when earwitnesses attempt to compare speech heard live with speech heard over the telephone. These concerns are less commonly found in the case law[163] (indeed, they frequently appear to be entirely absent)[164] but one does occasionally come across indications that courts are at least comforted by evidence that an earwitness purporting to identify a voice heard over the phone had previously heard the accused speak over the same medium.[165]

Surprisingly, neither of these concerns has been validated by empirical research. Only limited research has been undertaken but the majority of the existing studies have failed to find a significant difference in identification accuracy between voices heard over the telephone and voices heard live.[166] This is true using both target present and target absent lineups.[167] Furthermore, no research currently supports the notion that accuracy is impaired when an earwitness compares a voice heard over the telephone to a voice heard live, or vice-versa.[168]

This does not mean that judges should not concern themselves with possible impairments to identification accuracy when voices are heard over the telephone. The quality of a particular

---

[163] They are found in the literature. See, e.g., H. Rathborn, R. Bull and B. Clifford, "Voice Recognition over the Telephone" (1981) 9 Journal of Police Science and Administration 280 at 282-284.

[164] Identifications of voices heard on telephone wiretaps are commonly made based on comparisons with live speech without courts considering the potential effect of witnessing across different mediums. See, e.g., *Wu*, *supra* note XXX at paras 29-30 and 47-53.

[165] For example, in a case where the accused was found to have uttered a threat to kill over the telephone, the trial judge mentioned that the recipient of the threat had previously spoken with the accused, "on the phone and in person, ten to fifteen times": *Dobbin*, *supra* note XXX at para 2.

[166] Yarmey, "Earwitness identification," *supra* note XXX at 69 and 71; Perfect, *supra* note XXX at 976; Öhman, "Mobile Phone Quality," *supra* note XXX at 172-174. Rathborn, Bull and Clifford, however, found that witnesses who heard telephone voices either at exposure or in the lineup, or both, were significantly less accurate than witnesses who were only exposed to non-telephone voices: *supra* note XXX at 283.

[167] Yarmey, "Earwitness identification," *supra* note XXX at 69.

[168] Rathborn, *supra* note XXX at 283-284 (no significant difference in accuracy between telephone–telephone condition and telephone–non telephone conditions); Kerstholt, "Earwitnesses," *supra* note XXX at 194 (no significant differences between telephone–non-telephone condition and either telephone–telephone or non-telephone–non-telephone condition); Öhman, "Mobile Phone Quality," *supra* note XXX at 175 ("our results … imply … that using a mobile phone recorded voice line-up when the voice is originally heard over a mobile phone is not likely to improve identification accuracy").

transmission may be poor, there may be lots of ambient noise (for example, when someone uses a mobile or public telephone),[169] and some individuals may even speak differently over the phone.[170] What the empirical studies do suggest, however, is that the mere fact that the communication is over a telephone is not necessarily cause for concern. They also indicate that police forces need not worry too much about specifically using telephone voices when constructing voice lineups[171] and courts need not worry when the police do not use them. All that said, more research is needed before any definitive conclusions can be drawn.[172] For instance, it may be important that no study has compared exposure to truly live telephone voices with exposure to truly live voices both initially and at test; subjects are generally exposed only to tape-recorded voices.[173] There are very understandable practical reasons for this[174] but, given that the studies are trying to ascertain the possible impact of a medium of transmission (the telephone), it is a bit worrying that the studies have used in the experiments a second medium of transmission (the tape recorder) which may have its own impact.

h) Unfamiliar Accents and Languages

Witnesses are sometimes asked to identify the voice of a criminal perpetrator who spoke with an accent or even in a foreign language. A few judges have been concerned about the opportunity for error in these circumstances. Justice Twaddle of the Manitoba Court of Appeal,

---

[169] For a related example, see *Dodd*, *supra* note XXX, where voice was heard over a baby monitor through static and the moaning of another man.

[170] See Foulkes and Barron, *supra* note XXX at 182: "It is ... well known that some speakers adopt a 'telephone voice', modifying their rate of speech, segmental pronunciations, and/or voice quality."

[171] The police sometimes do this. See, e.g., *R v Nichols*, [2004] OJ No 6186 at paras 18-26 (CJ).

[172] Kersholt and his colleagues, for example, were unable to offer an explanation for why the results differed between the Rathborn, Bull and Clifford study and the remaining empirical studies: *supra* note XXX at 196. I am in the same position.

[173] Yarmey used live voices during initial exposure but tape-recorded voices for the lineups: "Earwitness identification," *supra* note XXX at 64-67.

[174] It could be very expensive, time-consuming and organizationally challenging to use truly live voices. It could also introduce methodological problems in maintaining consistency across tests.

for example, once wrote of the "danger, where the accused has an accent, that the witness is identifying the accent rather than the particular voice of the accused."[175] That is, in essence, a concern that people may not be able to distinguish readily between voices speaking in a manner that is unfamiliar to the witness. It is a concern that has been borne out by the empirical literature.

A number of studies have tested the accuracy of individuals trying to identify the voice of someone who speaks the same language as the listener but with a different accent. With English speech, the listeners tested have come from the United States, England, Scotland and Australia, and the accents studied have included Spanish, British, Scottish and Taiwanese accents. Other experiments have used non-English speech, listeners from other countries, and other accents. In most cases the bottom line result has been the same: people are worse at identifying accented voices than non-accented voices.[176] The difference has not always reached statistical significance (although it often has) but the overall trend is consistent. In many cases, the difference in accuracy has been in the range of 20%. Interestingly, the problem does not just arise when people are hearing accents that are very foreign or different than their own. In fact, the study that

---

[175] *Willis*, *supra* note XXX at para 24, per Twaddle J, dissenting.

[176] Stevenage, "The 'other-accent' effect," *supra* note XXX at 650-651 (listeners from Glasgow, Scotland and Southampton, England significantly worse at identifying speakers from the foreign vs. home location in target-present lineups; listeners from Southampton, but not Glasgow, made significantly more errors with Glaswegian vs. Southampton speakers in target-absent lineups); Thea Vanags, Marie Carroll and Timothy Perfect, "Verbal Overshadowing: A Sound Theory in Voice Recognition?" (2005) 19 Applied Cognitive Psychology 1127 at 1133-1134 (Australian listeners significantly better at identifying speakers from Australia vs. England); Kersholt, "Earwitnesses," *supra* note XXX at 193 (Dutch listeners better at identifying "ordinary standard-accented" Dutch voice than "regionally and socially strongly marked (The Hague) accent," although the difference was only marginally significant); Goggin, *supra* note XXX at 454 (monolingual English listeners from Texas significantly better at identifying mid-western American speakers using their normal voice than when they used a heavy Spanish accent, although the same was not true for English-Spanish bilingual listeners); Thompson, *supra* note XXX at 124 and 126 (mid-western American listeners better at identifying mid-western American English speakers vs. the same speakers using a heavy Spanish accent, although the difference did not reach statistical significance and was only found in a target-present lineup); Alvin Goldstein et al, "Recognition memory for accented and unaccented voices" (1981) 17 Bulletin of the Psychonomic Society 217 at 218-219 (identification by mid-western American listeners of Taiwanese-accented vs. mid-western American voices significantly worse when initial exposure was a single word, although not when it was a single sentence).

obtained the greatest difference in accuracy was one that asked listeners in southern England to identify voices of people from Glasgow.[177]

Studies testing the accuracy of individuals trying to identify voices of people speaking a foreign language that the listener does not understand have obtained similar results. English speaking listeners have been tested when hearing people speaking Spanish, German and French. Polish, German, Chinese and Spanish speaking listeners have been exposed to people speaking German or English. In most cases listeners were worse at identifying voices speaking an unfamiliar language than a familiar language.[178] A few studies have tried to investigate whether accuracy is improved by partial, albeit incomplete, familiarity with a language. It does seem to help.[179] Unfortunately, it does not always eliminate the difference in accuracy[180] and current research does not indicate how much familiarity is necessary to have a substantial impact.[181]

---

[177] Stevenage, "The 'other-accent' effect," *supra* note XXX at 650 (50% difference in a target-absent lineup for both correct rejections and false alarms).

[178] Olaf Köster and Niels Schiller, "Different influences of the native language of a listener on speaker recognition" (1997) 4 Forensic Linguistics 18 at 23-25 (Spanish and Chinese speaking listeners significantly worse than German speaking listeners at identifying a voice speaking German); Goggin, *supra* note XXX at 451-454 (English and German speaking listeners both significantly less accurate when hearing the unfamiliar vs. the familiar language; English speaking listeners also significantly less accurate when identifying Spanish vs. English speakers); Neils Schiller and Olaf Köster, "Evaluation of a foreign speaker in forensic phonetics: a report" (1996) 3 Forensic Linguistics 176 at 179 (listeners who only spoke English were significantly worse than native German speakers at identifying German speaking voices); Philippon, *supra* note XXX at 545 (English speaking listeners from the United Kingdom had a substantially higher false alarm rate when hearing a French vs. English speaker in both target-present and target-absent lineups, although the hit rate was identical); Thompson, *supra* note XXX at 124, 126, and 128 (in two of three experiments, English speaking Americans were significantly or marginally more accurate in identifying English speakers than Spanish speakers; no difference was found in the third experiment that only used a target-absent lineup but the overall false alarm rate in that lineup was 56% and the overall correct rejection rate was only 13%); Hollien, *supra* note XXX at 142-144 (Polish-only speakers less accurate than English speakers in identifying a voice speaking English).

[179] Schiller and Köster, *ibid* at 178-179 (English speaking listeners with some knowledge of German better than English-only listeners at identifying a voice speaking German); Köster and Schiller, *supra* note XXX at 24 ("In all cases (English, Spanish, Chinese), subjects with a knowledge of the target language (German) were able to identify a German speaker better than subjects without any knowledge of the target language"); Kirk Sullivan and Frank Schlichting, "Speaker discrimination in a foreign language: first language environment, second language learners" (2000) 7 Forensic Linguistics 95 at 98-99 and 105-107 (some knowledge of Swedish improved the accuracy of native English speakers when identifying a voice speaking Swedish, although the improvement did not increase as knowledge of Swedish increased).

[180] Köster and Schiller, *supra* note XXX at 23 (Chinese and Spanish listeners with some knowledge of German both significantly less accurate than native German listeners when identifying a voice speaking German). But see Schiller

Decreased accuracy in identifying accented and foreign language-speaking voices is probably attributable to an expertise effect.[182] By repeated exposure to a language or accent, listeners learn to perceive, understand and attend to subtle but important differences between speakers in syntax, pronunciation, emphasis, and so forth. Listeners unfamiliar with the meaning or manner of speech miss those clues and thus make more errors in identification.[183] As stated above, judges have sometimes implicitly recognized this problem. Some judges, however, have inappropriately labelled this a problem of "cross-racial" voice identification.[184] They are presumably borrowing from the eyewitness identification context in which "cross-racial" identification has been classified as a matter of particular concern[185] but it would be better to refer to, and focus on, *unfamiliar* accents and languages. This was perhaps best illustrated by a 1998 study that found that British and American listeners, Caucasian and non-Caucasian alike, were all significantly better at identifying voices of various races from their home country than from a foreign country.[186] Lack of familiarity can coincide with dissimilarity of race but it is really a function of differences in geography, culture, education, etc.[187]

i) Distinctiveness

---

and Köster, *supra* note XXX at 180 (English speaking listeners with some knowledge of German as accurate as native German speaking listeners at identifying a voice speaking German). Regretfully, Sullivan and Schlichting, *ibid*, did not compare their results with native Swedish speaking listeners.

[181] The studies were generally imprecise as to the amount of knowledge the test subjects had of the foreign language.

[182] See, e.g., Stevenage, "The 'other-accent' effect," *supra* note XXX.

[183] Identification accuracy is affected not just by repeatedly hearing the sounds of a foreign accent or language but also by understanding the content of the words spoken. See, e.g., Köster and Schiller, *supra* note XXX at 25.

[184] See, e.g., *Pinch, supra* note XXX at para 75 (mentioning "the dangers and potential prejudice of cross-racial voice identification evidence"); *Masters, supra* note XXX at para 47.

[185] E.g., *R v Richards* (2004), 186 CCC (3d) 333 at para 32 (Ont CA).

[186] Nathan Doty, "The influence of nationality on the accuracy of face and voice recognition" (1998) 111 American Journal of Psychology 191 at 196 and 202-204. Voices from the United States, England, Belize and France were used. They were predominantly Caucasian voices but a substantial minority was non-Caucasian. Listeners were Caucasian, African-American, African-English, Hispanic and Middle Eastern.

[187] See, e.g., Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 114: "accents differ as a function of factors other than simply race/ethnicity. Circumstances such as socioeconomic factors, education, historical and political groupings over time, and geographical regions of various sizes may play a role in affecting accents."

Courts quite frequently refer to the distinctiveness of the perpetrator's voice (or lack thereof) as a basis for assessing the reliability of a voice identification. This is not a reference to the perpetrator's manner of speaking, but rather to some inherent characteristic of the voice, such as pitch, nasality, or hoarseness.[188] Several courts, for example, have asked whether there is "some peculiarity or distinctiveness to the [accused's] voice that would make it more readily identifiable."[189] A finding that there is enhances reliability.[190] A finding that there is not sometimes diminishes reliability.[191]

Two studies have specifically tried to examine voice distinctiveness. The first found that distinctive voices were more accurately identified than non-distinctive voices in some experimental conditions but *less* accurately identified in other conditions, with a substantial number of errors being made.[192] The second study found a more consistent (positive) impact of distinctiveness[193] but only in experimental conditions where the target was absent from the lineup.[194]

These two studies leave one in a state of uncertainty regarding the impact of voice distinctiveness. Perhaps more importantly, they really tell us nothing about what specific

---

[188] Courts do sometimes refer to the distinctiveness of the perpetrator's manner of speaking (for example, repeated use of a particular phrase) but I do not examine that factor here due to the paucity of available research on the topic.

[189] E.g., *Saddleback*, *supra* note XXX at para 25; *J.E.*, *supra* note XXX at para 29; *Whalen*, *supra* note XXX at para 47. All these cases are citing language mentioned in *Williams*, *supra* note XXX at para 13.

[190] See, e.g., *Pinch*, *supra* note XXX at para 87.

[191] See, e.g., *Clouthier*, *supra* note XXX at para 20, where the Court discounted evidence by an earwitness that the robber had an Ottawa Valley accent on the basis that "this robbery occurred in a place where such an accent is commonplace."

[192] Orchard, *supra* note XXX at 256 (distinctive voices properly rejected more commonly than non-distinctive voices in a target-absent lineup when speech was whispered both at exposure and at test; distinctive voices less commonly identified than non-distinctive voices in a target-present lineup both when speech was whispered at exposure and normal at test and when speech was normal at exposure and at test).

[193] The study technically examined the effects of typicality rather than distinctiveness *per se*, but I am including the study here on the assumption that an atypical voice is a distinctive voice.

[194] J. Mullenix et al, "Typicality Effects on Memory for Voice: Implications for Earwitness Testimony" (2011) 25 Applied Cognitive Psychology 29 at 31-32 (test subjects more likely to mistake one typical voice for another typical voice than to mistake an atypical voice for another atypical voice, although they were about equally accurate in correctly recognizing both atypical and typical voices).

(allegedly distinctive) voice characteristics facilitate identification. The voices labelled distinctive in the studies were judged to be so by the experimenters and people assisting them but no indication was given as to the characteristic(s) of the voices that made them distinctive, or even whether they were all distinctive in the same way.[195] In the first study, in fact, a distinctive voice was simply defined as "highly striking and not likely to be confused with other voices."[196] The two studies, therefore, really only show that voices exist which are distinctive enough to allow, to some extent, for easier identification, and that a group of people is, to some extent, capable of spotting such voices. That is of little assistance to a court charged with assessing the weight to give to an identification made by a single witness based in whole or in part on that witness' personal belief that the voice is distinctive in some particular or undefined way.

It seems obviously true that some voices will be easier to identify than others but there are real dangers in relying on distinctiveness as a basis for assessing the reliability of an identification. If the particular distinctive characteristics of a voice are not specified by the earwitness, the evidence that the witness *perceives* it to be distinctive is of no assistance on the issue of reliability. Indeed, it raises the danger that a court will determine the reliability of a voice identification by determining credibility – i.e., by reference to the fact that the earwitness *sincerely* asserts that the voice was distinctive and thus identifiable. One cannot forget that a voice might *seem* distinctive to a witness only because it is unfamiliar to the witness, even though there are in fact many similar voices. Someone hearing a voice with a particular accent, for example, may not realize that the accent is common to a whole group of people. Indeed, in some cases a witness may perceive a voice to be distinctive because the witness is not aware of the fact that s/he can*not* distinguish it from other voices (including the accused's voice).

---

[195] Orchard and Yarmey, *supra* note XXX at 251-252; Mullenix, *ibid* at 30.
[196] Orchard and Yarmey, *supra* note XXX at 252.

It is only when a court is told what it is that allegedly distinguishes a voice that the court has any chance of assessing whether the voice truly is more readily identified.[197] Unfortunately, even then a court is not in a great position. Earwitnesses are unlikely to identify many vocal characteristics, even when asked.[198] I know of no research that has sought to quantify the incidence within a population group of any particular vocal characteristic. And very little research has been undertaken to determine what specific voice characteristics actually facilitate (or impair) identification. Some research suggests that high- and low-pitched voices may be more easily identifiable[199] but the research is neither clear nor consistent.[200] There is also the occasional hint in the empirical research that nasality and creakiness do not facilitate identification but the evidence is very far from conclusive.[201] In the end, a judge can usually do little more than guess whether a particular voice characteristic is either intrinsically helpful or sufficiently unusual as to facilitate identification. Unfortunately, experience shows that this can lead judges into error. Cases exist where a judge has deemed a voice to be distinctive at least in

---

[197] See *Whalen*, *supra* note XXX at para 18.

[198] See Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 111-112: test subjects "describe only four or five voice characteristics of the target in spite of being repeatedly prompted to remember additional characteristics … Most witnesses tend to describe specific characteristics, that is, pitch, enunciation, and tone of voice."

[199] Some courts have accepted testimony that a voice was distinctive at least in part because it was high-pitched. See, e.g., *Saddleback*, *supra* note XXX at para 33; *Sanghera*, *supra* note XXX at para 145.

[200] Foulkes and Barron, *supra* note XXX at 194 (amongst a small number of very familiar listeners, the voices with the highest differences from the group average pitch were on the whole more successfully identified than the voices with pitch closer to the group average); John Mullenix et al, "Earwitness Memory: Distortions for Voice Pitch and Speaking Rate" (2010) 24 Applied Cognitive Psychology 513 at 517 and 521 (results showing, on average, fewer errors for low-pitched voices, although no statistical analysis done to see whether the difference was significant; no apparent difference for high-pitched voices); Yarmey, "Earwitness identification," *supra* note XXX at 69 (test subjects significantly more accurate in a showup the higher they rated the pitch of the target's voice); Steven Stern et al, "Distortions in the Memory of the Pitch of Speech"" (2007) 54 Experimental Psychology 148 at 151, 153 and 156 (in two of three experiments, error rates lower for low- and high- vs. moderate-pitched voices, but no statistical analysis done to see whether any of the differences were significant and in the third experiment the error rate was highest for high-pitched voices).

[201] Yarmey, "Earwitness identification," *supra* note XXX at 69, found no correlation between accuracy and the degree to which a test subject rated a voice as nasal; Foulkes and Barron, *supra* note XXX at 190, observed that, relative to a small number of other voices, two voices that "displayed a markedly creaky phonation quality" were not as well identified by close friends.

part because it was accented,[202] even though studies show that accent generally impairs accurate identification.[203]

Tempting as it is to use voice distinctiveness as a basis for assessing the reliability of an identification, judges should be wary of doing so. Simply labelling a voice as distinctive is of no help and we do not know enough about what specific characteristics actually make a voice more readily identifiable. Judges would be wiser to focus on other factors.

4. Conclusion

More frequently than one might imagine, criminal courts in Canada are faced with the task of assessing the reliability of a voice identification made by an earwitness. Empirical research shows that voice identifications can sometimes be accurate but they can also be highly unreliable, even more so (on average) then eyewitness testimony. Indeed, the research results might even lead to the conclusion that identifications of unfamiliar voices should only rarely be given much weight. At the very least, a very cautious approach should be adopted by the courts. Even identifications of familiar voices can frequently be mistaken.

Canadian criminal courts have sometimes displayed impressive sensitivity to the frailties of earwitness testimony but, not surprisingly, they have not always got it quite right. Empirical research into earwitness evidence offers the most reliable source of information regarding the factors that affect accuracy yet it is extremely rare to find any reference to the research in the case law.[204] That is not a criticism of the courts. It is always difficult for scientific information regarding human behaviour and acuity to reach the courts and, to date, there has been little in the

---

[202] *Pabla*, *supra* note XXX at para 26; *R v Meyer*, 2012 ONCJ 791 at paras 32-33.
[203] See, *supra*, section 3(h).
[204] For an exception, see *Bubar*, *supra* note XXX at para 86.

legal literature to bridge that divide. Hopefully, this paper will offer some assistance. A high

degree of familiarity with the accused's voice, extended exposure to the perpetrator's voice at the

time of the crime, and deliberate attentiveness to its characteristics, are all factors that will

generally enhance the likelihood that a witness will make a reliable identification. The witness'

confidence in his or her accuracy, the suggestion that the voice was distinctive, and the fact that

it was heard over the telephone are probably of no significance. If the perpetrator spoke in a

foreign language or with an accent, or if the perpetrator disguised his or her voice or spoke with

anger or emotionality, it is less likely that an earwitness will make an accurate identification.

There are no definitive clues, however. All the factors and all the details must be considered.

Going forward, a more cautious approach and a more nuanced understanding of the

factors impacting reliability should lessen the dangers associated with earwitness evidence. But

the potential weakness of such evidence is so real that we would probably be unwise to leave

matters there.[205] Further reforms are likely necessary in order to safeguard the reliability of the

criminal justice process. I offer some preliminary thoughts to conclude.

Some have suggested that expert testimony could both inform courts about the factors

affecting the reliability of voice identifications and correct any mistaken beliefs that judges or

juries might hold.[206] There is some empirical support for the suggestion that expert testimony

---

[205] One mock jury experiment found that jury instructions on some of the factors affecting earwitness reliability had no effect on the percentage of guilty verdicts across good and poor listening conditions: Cindy Laub, *Can Earwitness Limitations be Overcome by the Court System? Strategies to Help Mock Jurors Appreciate the Limitations of Earwitness Testimony*, unpublished Ph.D. dissertation, University of Nebraska, 2010 at 41. Gary Edmond, Kristy Martire and Mehera San Rogue argue that, at least in cross-lingual voice identification cases, jury instructions and warnings are likely to be "vague imprecations" the primary benefit of which is "to make a trial formally fair": "'Mere guesswork': Cross-Lingual Voice Comparisons and the Jury" (2011) 33 Sydney L Rev 395 at 421-423. I am a little more optimistic, especially for judge-alone trials.

[206] Cindy Laub, Lindsey Wylie and Brian Bornstein, "Can the Courts Tell an Ear from an Eye? Legal Approaches to Voice Identification Evidence" (2013) 37 Law & Psychology Review 119 at 152-154; Eeva Komulainen, "Subjective Voice Identification: The Literal Meaning of 'Talking Yourself Behind Bars'" (1988) 26 Alta L Rev 521 at 544.

would be of assistance[207] but I wonder whether it would be of much value most of the time. Its admissibility is questionable,[208] it is expensive for both the litigants and the trial process, and the available experts are few and far between. How often will the necessary testimony be available, affordable and admissible in the average criminal case?

Another option would be to exclude from evidence voice identifications of dubious reliability. Counsel have occasionally sought to have courts do so in the past.[209] However, their success rate has been very low[210] and the idea of excluding evidence based on reliability concerns runs counter to the general preference in the law for such matters to go to weight.[211] The courts would probably only end up excluding the most glaringly weak identifications, much as they have done in the eyewitness identification context.[212]

The most fruitful option would be to encourage the police to administer pre-trial voice lineups,[213] at least in cases where earwitness testimony will play any significant role and the witness is not a very familiar listener. The police have occasionally conducted voice lineups in the past,[214] and the courts have very occasionally discounted earwitness evidence that was not

---

[207] Laub, *supra* note XXX at 41.

[208] In *Badgerow* (2010), *supra* note XXX at paras 6-9, proposed expert evidence on "the inherent frailties in ear witness identification" and "the methods used by the police … when interviewing potential voice identification witnesses" was deemed unnecessary and inadmissible. Expert testimony on the frailties of eyewitness evidence has frequently been excluded: e.g. *R v Frimpong*, 2013 ONCA 243; *R v Woodard*, 2009 MBCA 42. However, Komulainen, *supra* note XXX at 544, reports that an expert was allowed to testify at the preliminary hearing in the unreported early 1980s Alberta case of *R v Laberge*.

[209] See, e.g., *Badegrow* (2010), *supra* note XXX.

[210] Thus far, courts have only been willing to exclude voice identification evidence on the rare occasion when the witness had no real exposure to the suspect's voice before or after the crime: e.g. *Portillo*, *supra* note XXX at paras 40-42.

[211] *R v Buric*, (1996) 106 CCC (3d) 97 (Ont CA), aff'd [1997] 1 SCR 535.

[212] For the most part, only in-dock identifications have been excluded. See, e.g., *R v Holmes* (2002), 169 CCC (3d) 344 at 358-359 (Ont CA); *R v Tebo* (2003), 175 CCC (3d) 116 at 124-125 (Ont CA).

[213] This could be done through court rulings stressing the desirability of voice lineups or through internal police policies directing officers to perform lineups in earwitness as well as eyewitness cases.

[214] See, e.g., *Nichols*, *supra* note XXX. Komulainen, *supra* note XXX at 532, reported in 1988 that more than 150 voice lineups had been held across Canada. At 532-533, she describes three unreported cases.

subjected to a lineup test.[215] Lineups have long been accepted as a valuable tool in the eyewitness context.

The police could borrow from the extensive study of eyewitness lineups for guidelines on how to construct and administer earwitness lineups.[216] Several recommendations of eyewitness researchers would seem obviously transferable.[217] The lineup should be double-blind, meaning that neither the witness nor the lineup administrator should know which voice sample belongs to the suspect. The entire lineup identification process should be accurately recorded, preferably on videotape. The witness should be warned that the perpetrator might not be in the lineup, that it is equally important to clear the innocent as to identify the guilty, and that one of the available options is to not select anyone. If the witness makes a selection, the witness' level of confidence should be ascertained and noted.[218]

Earwitness researchers have also offered additional recommendations specific to the voice identification context.[219] The lineup voices should be tape-recorded and be a minimum of 60 seconds (or 200 words) in duration in order to offer a reasonable representation of the speaker's phonetic and idiosyncratic speech characteristics. In order to avoid deliberate distortion by a guilty suspect, the lineup should not include words or phrases known to have been spoken

---

[215] *Whalen*, *supra* note XXX at para 16.

[216] For a good overview of this research, see Gary Wells, "Eyewitness Identification: Systemic Reforms" [2006] Wis L Rev 615.

[217] See Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 126.

[218] Much attention has been paid in the eyewitness context to the choice between simultaneous and sequential lineups. In the earwitness context, the choice does not arise; all voice lineups are sequential. It has been recommended, however, that the witness not be told how many voices he or she will hear and that the witness only be allowed to go through the lineup once: A. Broeders and A. Amelsvoort, "A Practical Approach to Forensic Earwitness Identification: Constructing a Voice Lineup" (2001) 47 Problems of Forensic Sciences 237 at 239.

[219] See, e.g., Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 124-128; Harry Hollien, "On Earwitness Lineups" (2012) 4 Investigative Sciences Journal 1; A. Broeders and A. Amelsvoort, "Lineup Construction for Forensic Earwitness Identification: A Practical Approach" (1999) Proceedings of the XIVth International Congress of Phonetic Sciences 1373.

by the perpetrator during the crime.[220] However, ideally the lineup should include content using the same tone of speech as used by the perpetrator during the commission of the crime. The witness should be allowed to listen to the tapes in a quiet environment, preferably with headphones. The witness should be questioned about his or her hearing ability.

A particular challenge for voice lineups may be finding appropriate foils. The foils should match any descriptions of the perpetrator's voice given by witnesses and should broadly match the suspect's voice in terms of accent, pitch and speech rate as well as the suspect in terms of age, gender, ethnicity and, if possible, socioeconomic background. Finding such individuals may not be easy to do and may even require the assistance of linguistics experts.[221] Five or six foils is sufficient. They should not be allowed to hear the suspect's voice prior to tape recording so that they cannot mimic or differentiate theirs from the suspect's voice, conciously or not.

Another challenge for voice lineups may be obtaining an adequate voice sample from the suspect. The police can request a sample for purposes of holding a lineup but a suspect has the right to remain silent and may not speak, or speak naturally, knowing that the recording will be used for an identification procedure. An alternative source might be a portion of a recorded police interview but that comes with potential dangers.[222] The words spoken may signal to the listener that the speaker is under suspicion and foils uttering the same words may find it difficult

---

[220] Yarmey, "Earwitness descriptions," *supra* note XXX at 119-120, found that using phrases spoken during the crime did not improve identification accuracy.

[221] In the United Kingdom, the Home Office has suggested that a "suitable source of such material may be other police recorded interview tapes from unconnected cases": Home Office, "Advice on the Use of Identification Parades" (December 5, 2003), online at <https://www.gov.uk/government/publications/advice-on-the-use-of-voice-identification-parades>.

[222] A guilty suspect, aware that voice identification will be critical, might refuse to speak during a recorded interview, but it is possible that the police could surreptitiously record the interview. The courts are divided over whether a suspect has to be informed that he or she is being taped. See, e.g., *R v Hanuse*, [1996] AJ No 615 (QB) (not informing undermines voluntariness); *R v Stoddart*, [2005] OJ No 2207 at para 45 (SCJ) (suspects must be informed they are being videotaped); *R v Dickens*, [1999] AJ No 332 (QB) (not informing does not undermine voluntariness); *R v M.O.J.P.*, [2001] AJ No 1217 (Prov Ct) (not informing does not undermine voluntariness even for a young person).

to replicate the stress and emotionality felt and conveyed by the suspect; one study, using samples from a real police interview, found that it was easy for random listeners to pick out the suspect in the lineup.[223] Ideally, samples of spontaneous, casual speech would be used.[224] Police who have intercepted the suspect's communications during the investigation of the offence would probably have easiest access to such samples. Police in other cases could try to conduct a relaxed, non-accusatorial interview at the police station, clearly separated from any interrogation. More realistically, they may have to obtain judicial authorization to wiretap the accused post-arrest.[225]

I do not minimize the challenges involved in creating voice lineups, and I do not suggest that change can happen overnight.[226] I also acknowledge that, to date, most courts have been largely indifferent to the lack of a lineup.[227] But the general frailty of voice identification for all but the most familiar listeners is so apparent that we would be foolish not to try to seek out the information that lineups can provide. Until we do, perhaps the best advice for a court assessing proof of identity is to focus as much as possible on evidence in the case other than earwitness testimony.

---

[223] A. Laubstein, "Problems of voice lineups" (1997) 4 Forensic Linguistics 262. Yarmey recommends against using a speech sample from a police interview: "Psychology of Speaker Identification," *supra* note XXX at 127.

[224] Yarmey, "Psychology of Speaker Identification," *supra* note XXX at 124-125. Yarmey and Hollien also recommend that the voice lineup include samples of speech read from a selected text: *ibid* at 125; "On Earwitness Lineups," *supra* note XXX at 8. That would seem impossible unless the suspect cooperates.

[225] Komulainen suggests consent wiretaps were used in the three cases she discusses: *supra* note XXX at 536.

[226] Komulainen, *supra* note XXX at 532 estimated that preparation of a voice lineup takes approximately 40-60 hours. That estimate may be premised on the extensive involvement of a linguistics expert. Hopefully, the need for their participation would be minimized by the development of detailed and scientifically-informed guidelines.

[227] See, e.g., *R v Hammerstrom*, 2014 BCSC 94 at paras 24-25 (dismissing the argument that the lack of a voice lineup must contribute to a reasonable doubt); *Bubar*, *supra* note XXX at para 78 (not commenting on the defence submission that "no audio line-up was done"). Conversely, *Pinch*, *supra* note XXX at para 75, indicates that one factor to consider is whether a "voice identification parade" was conducted.

Appendix A

American Wrongful Conviction Cases Based Partly on Voice Identification Evidence

Name of accused	Source(s) identifying that voice identification evidence was used	Source(s) showing accused was factually innocent
Christopher Coleman	*People v Coleman*, 996 NE 2d 617 (Sup Ct 2013)	*People v Coleman*, 996 NE 2d 617 (Sup Ct, 2013); Radley Balko, "This Week in Innocence", *The Washington Post* (14 Mar 2014)
Sedrick Courtney	Innocence Project, "Know the Cases: Sedrick Courtney," online at <http://www.innocenceproject.org/Content/Sedrick_Courtney.php>	Bill Braun, "Wrongly imprisoned Tulsa man declared innocent, eligible to seek compensation from state", *Tulsa World* (24 Jul 2013)
Lesly Jean	*State v Jean*, 311 SE 2d 266 (Sup Ct 1984)	Estes Thompson, "Easley Pardons Ex-Marine Falsely Accused of Rape", *Charlotte Observer* (10 Feb 2001) 2B
Kerry Kotler	Michael Slackman, "Parting Shots Rape Indictment Dismissed, Kotler Raps DA Delay", *Newsday* (15 Dec 1992) 3	Michael Slackman, "Parting Shots Rape Indictment Dismissed, Kotler Raps DA Delay", *Newsday* (15 Dec 1992) 3
Alprentiss Nash	Maurice Possley, "Alprentiss Nash" *The National Registry of Exonerations*, online at < http://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3979>	Maudlyne Ihejirika, "State's attorney vacates conviction in 1995 murder", *Chicago Sun Times* (31 Aug 2012) 10
Kirk Odom	Muarice Possley, "Kirk Odom" *The National Registry of Exonerations*, online at <http://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3943>	Spencer Hsu, "After DNA retesting, Kirk Odom exonerated", *The Washington Post* (14 Jul 2012) B06
David Sheppard	Innocence Project, "Know the Cases: David Sheppard" online at <http://www.innocenceproject.org/Content/David_Shephard.php>	Innocence Project, "Know the Cases: David Sheppard" online at <http://www.innocenceproject.org/Content/David_Shephard.php>
Steve Snipes	Sarah Ovaska, "After 9 years, a pardon to prize: Falsely accused man wins 1 of only 4 issued by Easley since 2001", *Charlotte Observer* (3 Feb 2007) 1B	Sarah Ovaska, "After 9 years, a pardon to prize: Falsely accused man wins 1 of only 4 issued by Easley since 2001", *Charlotte Observer* (3 Feb 2007) 1B
Keith Turner	Bruce Tomaso, "Wrongfully convicted, he's trying to live right: Cleared by DNA test, Dallas man still weighed down by rape he didn't commit", *The Dallas Morning News* (12 Apr 2007), online at 2007 WLNR 8089917	Bruce Tomaso, "Wrongfully convicted, he's trying to live right: Cleared by DNA test, Dallas man still weighed down by rape he didn't commit", *The Dallas Morning News* (12 Apr 2007), online at 2007 WLNR 8089917
Eduardo Velasquez, aka Angel Hernandez	"Briefing", *Providence Journal Bulletin* (17 Aug 2001) 1	Innocence Project, "Know the Cases: Eduardo Velasquez" online at <http://www.innocenceproject.org/Content/Eduardo_Velasquez.php>
James Waller	Jeff Carlton, "Man exonerated of rape in	Jennifer Emily, "Dallas man pardoned in

	Texas", *Alameda Times-Star* (27 Jan 2007)	wrongful rape conviction: DNA exonerated 50-year-old who was jailed for 11 years", *The Dallas Morning News* (10 Mar 2007) 20A
Calvin Washington	Marice Possley, "Calvin E. Washington" *The National Registry of Exonerations*, online at <"http://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3720 >	"Pardon given to inmate cleared by DNA evidence", *Houston Chronicle* (11 Oct 2001) A32
Cathy Watkins	Marice Possley, "Cathy Watkins" *The National Registry of Exonerations*, online at <http://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=4073>	Daniel Beekman, "17-year jail nightmare finally ends", *New York Daily News* (14 Dec 2012) 20
Drew Whitley	Jill King Greenwood, "DNA reversal leaves case at square one", *Pittsburgh Tribune Review* (2 Oct 2006)	Jill King Greenwood, "DNA reversal leaves case at square one", *Pittsburgh Tribune Review* (2 Oct 2006)
Michael Williams	Maurice Possley, "Jailed 24 years, freed by DNA: Innocence Project key to exoneration in Louisiana rape case", *Chicago Tribune* (7 Mar 2005) 1	Maurice Possley, "Jailed 24 years, freed by DNA: Innocence Project key to exoneration in Louisiana rape case", *Chicago Tribune* (7 Mar 2005) 1
Calvin Willis	Innocence Project, "Know the Cases: Calvin Willis" online at <http://www.innocenceproject.org/Content/Calvin_Willis.php>	Henry Weinstein, "DNA Frees Man Jailed for 22 Years", *Los Angeles Times* (20 Sep 2003) 11
Collin Woodley	Robin Topping, "After 9 years in jail on wrongful conviction, Lifer freed" *Newsday* (20 Mar 1998) A05	Robin Topping, "After 9 years in jail on wrongful conviction, Lifer freed" *Newsday* (20 Mar 1998) A05