The reliability of voice recognition by 'ear witnesses': An overview of research findings

Helen Fraser

University of New England, Australia

10.21747/21833745/lanlaw/6_2a1

Abstract. *An ear witness is an individual lacking training in relevant branches of phonetic science, who hears a voice related to a crime, and claims to be able to identify the speaker. As with eye witness evidence, confident ear witness testimony is known to be highly compelling to triers of fact. But how reliable is it? This paper provides an overview of research findings, focusing mainly but not exclusively on situations where the witness hears an unfamiliar voice, then recognises someone heard at a later date as having been the speaker. The overview starts by outlining research demonstrating the unreliability of eye witness evidence, and the measures now commonly used in trials to counter its acknowledged weaknesses. It then reviews evidence from long-standing research across several disciplines indicating that ear witness evidence is considerably less reliable than eye witness evidence, and that its weaknesses are harder to cure.*

Keywords: *Ear witness, eye witness, speaker recognition, forensic speaker identification, forensic voice comparison.*

Resumo. *Uma testemunha auricular é uma pessoa sem formação em fonética que, ouvindo uma voz implicada num crime, se admite capaz de identificar o falante. Tal como com as testemunhas oculares, as testemunhas auriculares seguras são essenciais para os juízes. No entanto, serão fiáveis? Este artigo apresenta uma revisão da literatura centrada sobretudo, mas não exclusivamente, em situações nas quais as testemunhas ouvem uma voz desconhecida e reconhecem alguém que ouvem posteriormente como sendo o mesmo falante. Esta revisão começa por elencar estudos que mostram a falta de fiabilidade da prova testemunhal ocular e as medidas utilizadas habitualmente nos tribunais para combater as fraquezas identificadas. De seguida, aborda resultados de investigação fundamentados obtidos em várias áreas científicas que indicam que a prova testemunhal auricular é significativamente menos fiável do que a prova testemunhal ocular e que muito dificilmente é possível ultrapassar as suas fraquezas.*

Palavras-chave: *Testemunha auricular, testemunha ocular, reconhecimento de falantes, identificação de falantes forense, comparação de voz forense.*

Forensic voice recognition and ear witness identification evidence

Forensic voice recognition (or forensic speaker identification) refers to any evidence that purports to identify the speaker of utterances relevant to a crime by comparing his or her voice (the 'unknown', 'questioned' or 'disputed' voice) with the voice of a known or identifiable speaker (the 'known' voice). This form of evidence can appear in a wide range of scenarios, each with different factors affecting the reliability of the evidence (Watt, 2010; Foulkes and French, 2012).

The present overview focuses on scenarios in which:

- the evidence is provided by a 'lay' witness (i.e. one lacking expertise in relevant branches of forensic phonetics);
- the disputed voice is heard 'live' (as opposed to via a recording), so comparison is possible only via the memory of the hearer;
- the disputed voice is unfamiliar to the hearer at the time of hearing; and
- comparison with the known voice occurs at a later date.

This scenario is one of several in which the hearer can be called an 'ear witness', on analogy with the more common term 'eye witness'. It is worth briefly considering recent developments in the study of eye witness evidence, as a basis for discussion of the reliability of ear witness identification.

Brief background on eye witness research

Eye witnesses are highly credible

In everyday language, an eye witness is someone who has observed an event personally, as opposed to learning about it from a secondary source. Eye witness identification, considered 'direct' (as opposed to 'circumstantial') evidence, has always been given high status in courts of law. An observer who has seen a perpetrator 'with their own eyes', and is able to identify a suspect confidently, is extremely compelling in court (Semmler *et al.*, 2012). Presumably this is because triers of fact imagine themselves in the role of the eye witness and take on their sense of personal confidence (Heller, 2006).

Eye witnesses are often unreliable

Over recent decades, however, it has come to be widely accepted that there are substantial problems with eye witness identification. Contrary to popular belief, observing an event personally does not automatically confer ability to recount it accurately. A great number of experiments, perhaps most famously by Elizabeth Loftus (e.g. Loftus, 1979), have sought to measure the reliability of eye witnesses objectively. This is typically done by setting up scenarios, similar to real crimes except that the 'ground truth' about what happened, and who did it, is definitively known, and then questioning witnesses about what they observed (cf. Münsterberg, 1908).

A long chain of research (e.g. National Research Council, 2015), increasingly recognised within the law (e.g. Supreme Judicial Court Study Group on Eyewitness Evidence, 2013), has shown that eye witnesses are very often inaccurate in their description and identification of individuals. This inaccuracy can be explained by several factors. One major factor is that eye witnesses typically observe less than they think they do at the time of an event. Much of the information they feel derives from direct observation is really reconstructed later. Another major factor is that this reconstruction can be unwittingly 'contaminated' by misleading information obtained from other sources, most

notably via discussion with other witnesses, or by contextual information obtained (possibly unconsciously) after the events.

Interestingly, eye witnesses themselves are typically unable to distinguish which aspects of their recount come from their own direct observation, and which from information (misleading or otherwise) originating from a separate source: the data all merge together into one 'memory'. This means eye witnesses generally have poor ability to evaluate the accuracy of their own recount of events they have witnessed.

Eye witness confidence is a poor guide to accuracy

The above research can be summarised as showing that eye witnesses can be highly confident while also being completely wrong; or, conversely, unconfident but right. In other words, eye witnesses' confidence correlates poorly with their accuracy.

This creates a major problem in court, where, of course, ground truth is not objectively known. It is up to the jury to determine the facts of the matter, as best they can, by evaluating evidence from various sources, of which eye witness testimony is one. The fact that juries are liable to give credence to confident but possibly wrong eye witnesses can be expected to lead them to erroneous verdicts. Indeed, while researchers have long urged the courts to be more cautious regarding admission of eye witness testimony (Wells *et al.*, 2006), recent statistical studies have demonstrated conclusively that eye witness misidentification is a major cause of injustice, responsible for around three quarters of acknowledged wrongful convictions in the US (Gould *et al.*, 2012). Perhaps more important than the demonstration itself is the publicity it has gained, notably via the Innocence Project (http://www.innocenceproject.org).

Measures have been instituted to limit the effects of unreliable eye witnesses

Courts around the world have responded with several kinds of requirement (Vidmar and Schuller, 1989; Wells and Quinlivan, 2009) intended to minimize the risk of juries accepting confident but inaccurate eye witness identification. For example:

- eye witnesses may be expected to provide some kind of demonstration of the reliability of their recognition, typically by being able to pick the individual they claim to recognize from a lineup or photo gallery;
- judges may be expected to direct juries that eye witness evidence can be unreliable and should be treated with caution;
- expert evidence about the general unreliability of eye witnesses may be called to assist a jury in assessing testimony in the particular case being tried.

These measures are known not to be fully effective

While these measures represent an improvement on earlier practice, considerable independent research has shown them to be insufficient to counter the problems described above. For example:

- lineups and other tests of reliability are not always effective indicators of accuracy, as it is hard to control conditions that are known to influence the outcome of the test (Semmler *et al.*, 2004; Charman and Wells, 2008);
- judicial instructions can be surprisingly ineffective in reducing acceptance of confident but inaccurate eye witness testimony (Wykes, 2014; Berman, 2015);

- expert witnesses giving general background about eye witness unreliability may have more impact than judicial instructions, but cannot be guaranteed to be fully effective in ensuring appropriate caution in juries (Martire and Kemp, 2009, 2011);
- factors that statistically correlate with eye witness accuracy may be less important than individual differences in how observant the witness is (Balsdon *et al.*, 2018).

Indeed it has been argued convincingly that, rather than trying to cure the effects of poor quality eye witness evidence during a trial, a better approach is to ensure that initial collection of eye witness evidence is done in ways that maximise good quality observation and minimise risks of contamination (Paterson, 2018).

Calls for further restraint in admission of eye witness evidence are increasing

Despite the measures discussed above, overconfidence, both of eye witnesses in their own recognition, and in eye witness evidence by the courts, remains a substantial problem. Internationally, calls for restraint in admission of eye witness testimony are increasing, not just from researchers but also from within the law (e.g. Coyle *et al.*, 2008).

Ear witnesses in comparison to eye witnesses

Introduction

Ear witness evidence, though less common than eyewitness evidence, has been treated by the courts in a similar manner. Based on the issues raised above, that analogy in itself should give cause for concern over the reliability of ear witness evidence. However, research findings going back nearly a century (see Kreiman and Sidtis, 2011, Chapter 7, for a valuable summary) demonstrate that ear witnesses are even less reliable than eye witnesses. Unfortunately, findings like these remain insufficiently known outside academia, particularly within the law. The following sections outline some relevant considerations.

Listeners are confident of their ability to recognise voices

Everyday life affords listeners many experiences of instantly, confidently and usually accurately recognising the voices of people they interact with. Common knowledge offers a ready two-step explanation for this experience: (a) individual voices are assumed to have uniquely identifying features, sometimes said to be similar to a fingerprint; (b) listeners are assumed to be able to pick up on these features to recognise the speaker. However, well-established findings from several branches of phonetic science indicate that both steps of this explanation are false beliefs (Yarmey, 1995, 2004).

Objective tests confirm that listeners are actually poor at recognising voices

Both visual and auditory perception are crucial to many aspects of human life. However, when it comes to explicit memory for specific objects and people, humans are generally far more oriented to the visual domain than to the auditory domain (Cohen *et al.*, 2009). In particular, measured as a general capability, human ability to recognise individuals from their voices is far inferior to ability to recognise individuals from their faces (Barsics, 2014).

Findings like these give reason to expect that ear witnesses will be less reliable than eye witnesses. This expectation is borne out by long-standing experimental research demonstrating that when listeners are forced to rely only on the voice, ability to recognise even familiar speakers is surprisingly poor (McGehee, 1944; Ladefoged and Ladefoged, 1980).

Voices rarely have uniquely identifying features

These findings are familiar to phonetic scientists, who know that, contrary to popular belief, few voices are unique. This is the reason phoneticians universally deprecate the term 'voiceprint' (technically, 'spectrogram') for the false analogy it suggests between voices and fingerprints (Foulkes and French, 2012; Edmond *et al.*, 2011). In fact, to this date, despite decades of well-funded research, no features have been discovered that reliably allow identification of voices in anything remotely like the manner of fingerprints, which indeed are themselves less reliable than commonly assumed (Walvisch, 2017).

These observations are often found surprising or hard to believe, but informal appreciation of their validity can be gained by considering the enormous financial gains that would accrue to developers of a system that allowed users to access smart phones, ATM machines and so on by speaking their personal identification number (PIN) rather than by entering it secretly via a keyboard. The fact we still do not have widespread voice identification facilities in these highly lucrative contexts confirms that reliable identification of a voice from an open population is not (yet) possible. (It is true that speaker verification systems are becoming increasingly familiar, but these typically require the speaker to use particular words, to make a clear recording, and to provide additional information, such as a tax file number or date of birth, which reduces the population among which the system must discriminate. Conditions like these are clearly unrealistic in a forensic speaker identification context.)

Given the difficulties of identifying voices, many experts in forensic phonetics agree that the responsible evidence is not speaker identification but speaker comparison, in which recorded samples of known and disputed voices are analysed, and conclusions regarding similarities and differences are given appropriate weight via appropriate use of statistics (Watt, 2010; Foulkes and French, 2012). Of course, providing speaker comparison evidence of this kind requires availability of recordings, not reliance on memory. More importantly, it requires substantial expertise, not just for the technical analysis, but to express conclusions with appropriate statistical caution (Rose, 2005). Highly confident voice identification is generally viewed with suspicion by genuine experts.

Speaker recognition depends heavily on context and content

These findings raise the question of how it can be that everyday speaker recognition is so often effortlessly successful. The answer requires appreciation of the fact that, though listeners do recognise people's voices, they do not recognise people by their voices.

Everyday speaker identification relies not on the voice alone, but on a great deal of additional information, including prior expectations deriving from the context in which the voice is heard, and the content of what the speaker is saying. Listeners are typically unaware of these influences, but they certainly must, and do, play a substantial role in voice recognition (Ladefoged, 1978).

Ear witnesses are easily misled by unreliable contextual information

In the vast majority of everyday situations, information derived from the context is in alignment with information derived from the voice. In the rare cases that a mismatch causes erroneous recognition, this is easily and quickly discovered when the true identity is revealed.

However, in the forensic context, where the whole point is that the speakers' true identity is not definitively known, the unnoticed role of context in speaker recognition

makes it easy for listeners to be unwittingly misled by inaccurate contextual assumptions. This can happen both prospectively, at the time of perception, and retrospectively, via information provided later (Smith and Baguley, 2014). The effects are similar to those discussed above in relation to eye witnesses.

However, the situation with ear witnesses is substantially more problematic than that of eye witnesses. One reason is that, in addition to background, or external, context, speech also has internal context, derived from its content, or linguistic meaning. As with external context, internal context provides a powerful source of information that guides, or primes, listeners' recognition of speakers. The problem is that, while this kind of priming is generally helpful in everyday situations where contextual expectations are in alignment with ground truth, it is potentially highly misleading in forensic situations (Philippon *et al.*, 2007).

Earwitnesses have very limited metalinguistic awareness

The whole point about speech, as a symbolic system, is that it directs users' attention away from its form, and towards its content, or linguistic meaning. This and other factors mean that most people have extremely limited metalinguistic awareness (ability to describe the form of speech). It takes many years of training for linguists to be able to describe the nature of speech in a clear and consistent manner.

This means that when ear witnesses go beyond a mere statement of recognition and attempt to describe the voice, their description is typically even less reliable than eye witnesses' description of persons, even when they are noting apparently obvious features such as pitch and regional accent (Tomkinson and Watt, 2018).

Confident ear witnesses have a compelling effect on juries

Taken together, the facts just discussed create a mismatch between confidence and accuracy in ear witnesses that is even greater than that of eye witnesses (Olsson *et al.*, 1998).

Nevertheless, juries are known to trust confident ear witness testimony at least as much as they trust confident eye witness testimony (Laub *et al.*, 2016; Van Wallendael *et al.*, 1994). This creates a clear potential for unfairness or injustice, and indeed cases of wrongful conviction due to unreliable ear witness testimony are known (e.g. Sherrin, 2016).

Effects of confident but unreliable ear witnesses are very hard to cure

The considerations outlined so far may suggest a need for curative measures similar to those used for eye witnesses to be used in court. However, there are good reasons to caution that such measures are unlikely to offer even the limited success they enjoy in relation to eye witness evidence. For example:

- Testing ear witnesses by their ability to identify a target from a lineup is even more problematic than similar tests of eye witnesses. While lineups can produce reliable results under very particular circumstances (Nolan, 2007), they can also elicit highly unreliable responses (Sarwar *et al.*, 2014). Attempts at guidelines (Hollien, 2012; McGorrery and McMahon, 2016) have not so far delivered methods that meet with general scientific approval.
- Judicial directions, or even expert evidence, are less likely with ear witnesses than with eyewitnesses to cure juries' over-confidence (Laub, 2010; Laub *et al.*, 2013).

Conclusion

The history of ear witness research is characterised by researchers across several disciplines (notably psychology and phonetics, but also law) consistently urging the courts to curtail use of ear witness identification due to its strong tendency to be unreliable and misleading, and its consequent potential to contribute to unfairness or miscarriage of justice. Thorough reviews have been provided for the UK by Clifford (1980), for the US by Solan and Tiersma (2003), for Canada by Sherrin (2016), and for Australia by McGorrery and McMahon (2016). Despite this, ear witness evidence continues to be used, with effective restraints rarely implemented in court even to the limited extent they are for eye witnesses.

References

Balsdon, T., Summersby, S., Kemp, R. I. and White, D. (2018). Improving face identification with specialist teams. *Cognitive Research Principles and Implications*, 3(25), 1–13.

Barsics, C. G. (2014). Person recognition is easier from faces than from voices. *Psychologica Belgica*, 54(3), 244–254.

Berman, M. K. (2015). *Eyewitness identification jury instructions: Do they enhance evidence evaluation?* , CUNY.

Charman, S. and Wells, G. (2008). Can eyewitnesses correct for external influences on their lineup identifications? The actual/counterfactual assessment paradigm. *Journal of Experimental Psychology: Applied*, 14(1), 5–20.

Clifford, B. R. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior*, 4(4), 373.

Cohen, M. A., Horowitz, T. S. and Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Science*, 106(14), 6008–6010.

Coyle, I., Field, D. and Miller, G. (2008). The blindness of the eye-witness. *Australian Law Journal*, 82, 471–498.

Edmond, G., Martire, K. A. and San Roque, M. (2011). Unsound Law: Issues with "expert" voice comparison evidence. *Melbourne University Law Review*, 35, 52–112.

Foulkes, P. and French, P. (2012). Forensic speaker comparison: a linguistic-acoustic perspective. In L. M. Solan, Ed., *The Oxford handbook of language and law*. Oxford: Oxford University Press, 557–572.

Gould, J. B., Carrano, J., Leo, R. and Young, J. (2012). *Predicting erroneous convictions: A social science approach to miscarriages of justice*. Washington DC: National Institute of Justice.

Heller, K. J. (2006). The cognitive psychology of circumstantial evidence. *Michigan Law Review*, 105(2), 241–305.

Hollien, H. (2012). On earwitness lineups. *Investigative Sciences Journal*, 4(1).

Kreiman, J. and Sidtis, D. (2011). Identifying unfamiliar voices in forensic contexts. In *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Oxford: Wiley Blackwell, chapter 7, 237–259.

Ladefoged, P. (1978). Expectation affects identification by listening. *Language and Speech*, 21(4), 373–374.

Ladefoged, P. and Ladefoged, J. (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics*, 49, 43–51.

Laub, C. E. (2010). *Can earwitness limitations be overcome by the court system?* , University of Nebraska, Lincoln.

Laub, C. E., Kimbrough, C. D. and Bornstein, B. H. (2016). Mock juror perceptions of eye witnesses vs ear witnesses: Do safeguards help? *American Journal of Forensic Psychology*, 34(2).

Laub, C. E., Wylie, L. E. and Bornstein, B. H. (2013). Can the courts tell an ear from an eye? Legal approaches to voice identification evidence. *Law and Psychology Review*, 37, 119–158.

Loftus, E. F. (1979). *Eyewitness Testimony*. Cambridge, MA: Harvard University Press.

Martire, K. A. and Kemp, R. I. (2009). The impact of eyewitness expert evidence and judicial instruction on juror ability to evaluate eyewitness testimony. *Law and Human Behavior*, 33(3), 225–236.

Martire, K. A. and Kemp, R. I. (2011). Can experts help jurors to evaluate eyewitness evidence? A review of eyewitness expert effects. *Legal and Criminological Psychology*, 16(1), 24–36.

McGehee, F. (1944). An experimental study of voice recognition. *The Journal of General Psychology*, 31(1), 53–65.

McGorrery, P. G. and McMahon, M. (2016). A fair "hearing": Earwitness identifications and voice identification parades. *International Journal of Evidence and Proof*, 21(3), 262–286.

Münsterberg, H. (1908). *On the witness stand: Essays on psychology and crime*. New York: MacClure.

National Research Council, (2015). *Identifying the culprit: Assessing eyewitness identification*. Washington DC: National Academies Press.

Nolan, F. (2007). A recent voice parade. *International Journal of Speech Language and the Law*, 10(2), 277–291.

Olsson, N., Juslin, P. and Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. *Journal of Experimental Psychology: Applied*, 4(2), 101–118.

Paterson, H. M. (2018). Why we made iWitnessed, an app to collect evidence. *The Conversation*.

Philippon, A. C., Cherryman, J., Bull, R. and Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures. *Applied Cognitive Psychology*, 21(4), 539–550.

Rose, P. (2005). Forensic speaker recognition at the beginning of the 21st century: An overview and a demonstration. *Australian Journal of Forensic Sciences*, 37, 49–72.

Sarwar, F., Allwood, C. M. and Zetterholm, E. (2014). Earwitnesses: the type of voice lineup affects the proportion of correct identifications and the realism in confidence judgments. *International Journal of Speech Language and the Law*, 21(1), 45–17.

Semmler, C., Brewer, N. and Bradfield Douglas, A. (2012). Jurors believe eyewitnesses. In B. L. Cutler, Ed., *Conviction of the innocent: Lessons from psychological research*. American Psychological Association, 1–39.

Semmler, C., Brewer, N. and Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology*, 89(2), 334–346.

Sherrin, C. (2016). Earwitness evidence: The reliability of voice identifications. *Osgoode Hall Law Journal*, 52, 819–862.

Smith, H. M. J. and Baguley, T. (2014). Unfamiliar voice identification: Effect of post-event information on accuracy and voice ratings. *Journal of European Psychology Students*, 5(1), 59–68.

Solan, L. M. and Tiersma, P. M. (2003). Hearing voices: Speaker identification in court. *Hastings Law Journal*, 54, 373–435.

Supreme Judicial Court Study Group on Eyewitness Evidence, (2013). *Report and recommendations to the Justices*.

Tomkinson, J. and Watt, D. (2018). Assessing the abilities of phonetically untrained listeners to determine pitch and speaker accent in unfamiliar voices. *Language and Law/Linguagem e Direito*, 5(1), 19–37.

Van Wallendael, L. R., Surace, A., Hall Parsons, D. and Brown, M. (1994). Earwitness voice recognition: Factors affecting accuracy and impact on jurors. *Applied Cognitive Psychology*, 8, 661–677.

Vidmar, N. J. and Schuller, R. A. (1989). Juries and expert evidence: Social framework testimony. *Law and Contemporary Problems*, 52(4), 133–176.

Walvisch, J. (2017). Fingerprinting to solve crimes: not as robust as you think. *The Conversation*.

Watt, D. (2010). The identification of the individual through speech. In C. Llamas and D. W. (Eds.), Eds., *Language and Identities*. Edinburgh: Edinburgh University Press, 76–84.

Wells, G. L., Memon, A. and Penrod, S. D. (2006). Eyewitness Evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7(2), 45–75.

Wells, G. L. and Quinlivan, D. S. (2009). Suggestive eyewitness identification procedures and the Supreme Court's reliability test in light of eyewitness science: 30 years later. *Law and Human Behavior*, 33(1), 1–24.

Wykes, T. (2014). *Juror perceptions of eyewitness identification evidence*. , University of Ontario, Canada.

Yarmey, A. D. (1995). Earwitness speaker identification. *Psychology, Public Policy, and Law*, 1(4), 792–816.

Yarmey, A. D. (2004). Common-sense beliefs, recognition and the identification of familiar and unfamiliar speakers from verbal and non-linguistic vocalizations. *International Journal of Speech Language and the Law*, 11(2), 267–277.