***Ear*witness Memory**

The overview that follows is offered to help the trier of fact understand earwitness memory — the identification of a person by voice — and what the research means when a witness claims to recognize a speaker heard during a crime. Voice identification arises in kidnappings, ransom and threatening calls, robberies committed in the dark or by a masked offender, and any case in which a witness heard but could not see the perpetrator. Its purpose is to define the terms a court is likely to encounter, to summarize what the research shows about the reliability of voice identification and the conditions that affect it, and to do so in plain language and without favoring any party.

Recognizing a Person by Voice

An earwitness is a witness who identifies, or attempts to identify, a person by the sound of the voice. As with visual identification, two different tasks are involved. Describing a voice — its pitch, accent, or rate of speech — is a recall task that depends on a vocabulary most people lack; voices are notoriously hard to put into words, and descriptions tend to be brief and vague. Recognizing a voice in a lineup is a recognition task that does not require words. The two abilities are loosely related at best, so a thin description neither establishes nor refutes the witness's ability to recognize the voice. The central scientific point is that memory for an unfamiliar voice is, in general, considerably weaker and more error-prone than memory for an unfamiliar face.

How Accurate Is Voice Identification?

The reliability of earwitness identification depends heavily on whether the voice was already familiar. For an unfamiliar speaker heard only once and briefly, accuracy is poor: across studies, correct identification of such speakers is often below 50%, and misidentification of strangers — including confidently identifying the wrong person, or judging a stranger to be familiar — is common (Yarmey, 1995). The earliest research on the question, prompted by a famous case discussed below, already found voice recognition to be unreliable (McGehee, 1937), and decades of work since have confirmed that listeners are frequently mistaken about voices they heard only in passing (Clifford, 1980). Familiar voices are recognized far more accurately, and the recognition of a well-known voice is comparatively robust; but even highly familiar voices are sometimes misidentified, and familiarity is a matter of degree rather than a guarantee.

A cautionary illustration comes from the 1932 Lindbergh kidnapping. Charles Lindbergh heard a few shouted words — "Hey, Doctor!" — from roughly a football field's distance, in the dark, on a single occasion. Nearly two and a half years later, in a highly suggestive setting in which the man had already been identified to him as the accused, Lindbergh identified Bruno Hauptmann's voice as that speaker. The identification combined every condition the research associates with unreliability — a brief exposure to an unfamiliar voice, a long delay, and a suggestive procedure — and it helped launch the scientific study of earwitness memory.

What Affects Earwitness Accuracy

Several conditions shape how reliable a voice identification will be, and they parallel the factors known to affect visual identification. Familiarity is the most important: the recognition of familiar and unfamiliar voices appears to rely on partly different processes, with familiar-voice recognition relatively durable and unfamiliar-voice discrimination weak (Stevenage, 2018). Exposure matters as well — longer speech samples and more of them generally improve accuracy, while a few words afford little basis for recognition. Memory for an unfamiliar voice also fades quickly with delay, more steeply than memory for faces (Saslove & Yarmey, 1980).

Other factors are specific to the auditory domain. Disguise is especially damaging: whispering, faking an accent, or otherwise altering the voice sharply reduces identification accuracy, and a perpetrator who spoke in an unusual register during the crime presents a moving target. A change in emotional tone between the crime and the lineup has a similar effect, and listeners' memories for properties such as pitch and speaking rate are themselves distorted over time. Language familiarity is a further, often overlooked factor: listeners identify speakers far better in their own language than in an unfamiliar one, so a witness attempting to identify a speaker of a language the witness does not understand is at a marked disadvantage (Goggin, Thompson, Strube, & Simental, 1991; Philippon, Cherryman, Bull, & Vrij, 2007). Finally, voices first heard over the telephone or other transmission lose information that aids recognition.

As with eyewitness identification, an earwitness's confidence is a weak guide to accuracy. Studies comparing the two find that confidence is, if anything, even less diagnostic for voices than for faces, and that earwitnesses tend to be overconfident — sure of identifications that are wrong (Olsson, Juslin, & Winman, 1998). A witness's certainty that "I would know that voice anywhere" therefore cannot substitute for the objective conditions under which the voice was heard and tested.

Lay Earwitnesses and Forensic Voice Comparison

Two distinct kinds of evidence are easily confused. This paper concerns the lay earwitness — an ordinary listener who heard the perpetrator and later attempts a recognition. A separate field, forensic voice comparison, involves trained examiners comparing recordings using auditory and acoustic methods, and sometimes automated systems. The reliability of expert voice comparison is itself the subject of ongoing scientific and legal debate and is governed by the rules on expert evidence. The research summarized here addresses the everyday earwitness, whose recognition is a product of ordinary, and fallible, human memory.

Conducting a Fair Voice Lineup

Because voice identification is error-prone, the manner in which it is obtained matters greatly, and researchers have proposed safeguards that parallel best practice for visual lineups. The voice lineup should be administered by someone who does not know which voice is the suspect's, presented from high-quality recordings rather than live, and preceded by an instruction that the perpetrator may or may not be present and that declining to choose is acceptable. The witness's hearing should be checked. The recordings should be of adequate length — on the order of thirty to sixty seconds — and matched in quality, with the suspect and the fillers recorded the same way.

Fillers should be chosen to fit the witness's own description of the perpetrator's voice and matched to the suspect on general vocal qualities and on background characteristics such as age, sex, and dialect, so that the suspect's voice does not stand out; where the identification crosses language or ethnic lines, a phonetician or a listener from the relevant group should help select fillers. To prevent a guilty suspect from disguising particular words, the lineup should use neutral content rather than the phrases spoken during the crime. The witness's certainty should be recorded in the witness's own words at the time, and the fairness of the lineup can be checked in advance by confirming that listeners who never heard the crime do not pick the suspect more often than chance. Where both visual and voice identifications are sought, the research recommends attempting the visual identification first and constructing each lineup with its own suitable fillers.

Earwitness Memory and the Law

The admissibility threshold for lay voice identification is low. Under Federal Rule of Evidence 901(b)(5), a voice may be identified by the opinion of anyone who has heard it "at any time under circumstances that connect it with the alleged speaker," and familiarity acquired even after the

events in question can suffice. But this rule governs only authentication — whether the evidence may come in at all — and sets a deliberately minimal bar. It does not measure reliability, which goes to the weight the fact-finder gives the identification.

Where the identification is challenged as the product of suggestion, the same constitutional framework that governs visual identification applies. Due process requires a court to screen an identification for reliability only when law enforcement used unnecessarily suggestive procedures; in that situation courts weigh the reliability factors of Neil v. Biggers (1972) and Manson v. Brathwaite (1977), and absent such police-arranged suggestion the reliability of a voice identification is for the jury to assess (Perry v. New Hampshire, 2012). Courts have increasingly acknowledged that confident identifications can be mistaken (State v. Henderson, 2011), and commentators and courts have urged that juries be specifically cautioned that voice identification is more error-prone than visual identification. Whether an expert may explain this research is governed by the rules on scientific evidence — Daubert v. Merrell Dow Pharmaceuticals (1993) and Kumho Tire Co. v. Carmichael (1999) in federal court and the states that follow them, and the general-acceptance test of Frye v. United States (1923) elsewhere. Under any standard, such testimony is offered to inform the trier of fact about the limits of earwitness memory — not to opine on whether a particular identification is correct, a judgment the law reserves to the fact-finder.

Special Issues in Legal Practice

The principles described above apply with equal force to every party in a legal proceeding. Memory operates the same way regardless of who is asking the questions, and the science summarized here favors no side. What differs between the parties is not the science but the vantage point — the particular places where memory issues tend to surface, and the particular errors each party is best positioned to guard against. The two summaries below identify those recurring issues. They are offered for orientation, not as litigation strategy; in every case the controlling question is what the evidence in that case will, and will not, support, and the expert's role is to inform the trier of fact rather than to advocate for either side.

Special Issues for Prosecutors

For the State, a voice identification is most defensible when the conditions of hearing were favorable and the lineup that produced it was sound. A perpetrator already familiar to the witness, an extended and undisguised exposure, a short interval before testing, and a voice lineup administered blind, with adequate samples, properly matched fillers, neutral content, and a recorded certainty statement, are affirmative indicators of reliability the State is well positioned to establish and document. The errors the prosecution is best positioned to guard against are largely procedural: one-person voice showups, lineups in which the suspect's voice is salient, use of the very words spoken during the crime, suggestive instructions, and delay. An interpretive caution also applies: because describing a voice is difficult, a witness's thin verbal description should not be mistaken for an inability to recognize the voice — but neither does a confident identification, standing alone, establish accuracy, since confidence is a weak guide for voices.

Special Issues for Defense

For the defense, the recurring task is to show how the specific conditions of the case map onto the factors known to make voice identification unreliable — recognizing that an earwitness can be entirely sincere and still be mistaken. Many cases involve precisely the conditions the research flags: an unfamiliar voice, a brief exposure, words heard in the dark or at a distance, disguise such as whispering or a feigned accent, a long delay before testing, transmission over a telephone, or a speaker using a language the witness does not know. The manner of testing is equally open to

scrutiny: suggestive showups, fillers who make the suspect's voice stand out, use of the crime's own phrases, and confirming feedback can all shape an identification after the fact, and the witness's confidence is, for voices, an especially poor indicator of accuracy. The law contemplates that these matters be placed before the trier of fact through cross-examination, through challenge to the manner of the identification, through requests for cautionary instructions, and, where the governing standard is met, through expert testimony (Daubert v. Merrell Dow Pharmaceuticals, 1993; Frye v. United States, 1923). The aim is not to assert that voice identification is never possible — familiar voices in particular can be recognized — but to ensure that the conditions of the particular identification are weighed.

References

Clifford, B. R. (1980). Voice identification by human listeners: On earwitness reliability. Law and Human Behavior, 4(4), 373–394. https://doi.org/10.1007/BF01040628

Goggin, J. P., Thompson, C. P., Strube, G., & Simental, L. R. (1991). The role of language familiarity in voice identification. Memory & Cognition, 19(5), 448–458. https://doi.org/10.3758/BF03199567

McGehee, F. (1937). The reliability of the identification of the human voice. The Journal of General Psychology, 17(2), 249–271. https://doi.org/10.1080/00221309.1937.9917999

Olsson, N., Juslin, P., & Winman, A. (1998). Realism of confidence in earwitness versus eyewitness identification. Journal of Experimental Psychology: Applied, 4(2), 101–118. https://doi.org/10.1037/1076-898X.4.2.101

Philippon, A. C., Cherryman, J., Bull, R., & Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures. Applied Cognitive Psychology, 21(4), 539–550. https://doi.org/10.1002/acp.1296

Saslove, H., & Yarmey, A. D. (1980). Long-term auditory memory: Speaker identification. Journal of Applied Psychology, 65(1), 111–116. https://doi.org/10.1037/0021-9010.65.1.111

Stevenage, S. V. (2018). Drawing a distinction between familiar and unfamiliar voice processing: A review of neuropsychological, clinical and empirical findings. Neuropsychologia, 116(Pt B), 162–178. https://doi.org/10.1016/j.neuropsychologia.2017.07.005

Yarmey, A. D. (1995). Earwitness speaker identification. Psychology, Public Policy, and Law, 1(4), 792–816. https://doi.org/10.1037/1076-8971.1.4.792

Yarmey, A. D. (2007). The psychology of speaker identification and earwitness memory. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), The handbook of eyewitness psychology: Vol. 2. Memory for people (pp. 101–136). Lawrence Erlbaum Associates.

Legal Authorities

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).

Manson v. Brathwaite, 432 U.S. 98 (1977).

Neil v. Biggers, 409 U.S. 188 (1972).

Perry v. New Hampshire, 565 U.S. 228 (2012).

State v. Henderson, 208 N.J. 208, 27 A.3d 872 (2011).

Fed. R. Evid. 901(b)(5).

Related topic papers: Eyewitness Misidentification · Eyewitness Identification Procedures · Lineup Best Practices · Confidence and Accuracy · Cross-Race Effect · Estimator vs. System Variables