2003

Hearing Voices: Speaker Identification in Court

Lawrence Solan

Peter Tiersma

Hearing Voices:
Speaker Identification in Court

by
LAWRENCE M. SOLAN* AND PETER M. TIERSMA**†

Introduction

On April 3, 1936, Bruno Richard Hauptmann was executed in Trenton, New Jersey, for kidnapping and murdering the Lindbergh baby. The most dramatic moment in the trial came when Charles Lindbergh identified Hauptmann's voice as that of his baby's kidnapper. Three years earlier, still hoping to get his son back alive, Lindbergh had accompanied Dr. John Condon to St. Raymond's cemetery in the Bronx. Condon had gone there to deliver ransom money to the kidnapper, while Lindbergh waited some seventy to one hundred yards away in the car.[1] Out of the darkness came the words, "Hey, doktor! Over here, over here," pronounced with a foreign accent.[2] Twenty-nine months after the encounter in the cemetery, in September 1934, Lindbergh told a Bronx grand jury that "[i]t would be very difficult to sit here and say that I could pick a man by that voice."[3] Nonetheless, the district attorney asked Lindbergh later that day: "Would you like to see the man who kidnapped your son?"[4] The next morning, while Lindbergh sat in the back of the D.A.'s office among a group of detectives, Hauptmann was brought in and

---

   1. LUDOVIC KENNEDY, THE AIRMAN AND THE CARPENTER: THE LINBERG KIDNAPPING AND THE FRAMING OF RICHARD HAUPTMANN 266-67 (1985).
   2. JOHN F. CONDON, JAFSIE TELLS ALL! REVEALING THE INSIDE STORY OF THE LINDBERGH-HAUPTMANN CASE 149 (1936).
   3. JIM FISHER, THE LINDBERGH CASE 248 (1987).
   4. *Id.*

asked to repeat the words, "Hey, doctor. Here, doctor, over here."[5] Lindbergh told the prosecutor that he recognized Hauptmann's voice as that of the kidnapper.[6]

At trial, Lindbergh recounted the events at the cemetery.[7] He then testified:

> Q. Whose voice was it, Colonel, that you heard in the vicinity of St. Raymond's Cemetery that night, saying, "Hey, Doctor?"
>
> A. That was Hauptmann's voice.
>
> Q. You heard it again the second time where?
>
> A. At District Attorney Foley's office in New York, in the Bronx.[8]

Lindbergh's lawyer commented: "The minute Lindbergh 'pointed his finger' at Hauptmann, the trial was over. 'Jesus Christ' himself said he was *convinced* this was the man who killed his son. Who was anybody to doubt him or deny him justice?"[9]

Sixty-five years later, Hauptmann's conviction and execution remain controversial. Was he truly guilty of the kidnapping, or was he wrongly convicted, perhaps in part because of the anti-German sentiment then prevalent in the United States? We will not attempt to answer the ultimate question of Hauptmann's guilt or innocence. But we will address others that the Lindbergh case raised: Are people really able to remember a voice that they have only heard once? Are three syllables enough of a sample? Isn't twenty-nine months a long time? Does the stress of the situation make memory better or worse? What effect does a foreign accent have on our ability to identify voices?

Hauptmann's attorney was in no position to present expert evidence on any of these questions. There was no relevant expertise at that time. But the case did trigger experimentation by psychologists into the question of how good we are at identifying voices.[10] Since then we have learned a great deal more. Nonetheless, the law has virtually stood still since Hauptmann's execution. Generally, people who believe they recognize a voice are simply

---

5. *Id.* at 249.

6. *Id.* at 249–50.

7. In his testimony, Lindbergh left out the words, "over here." Transcript at 109, State v. Hauptmann, 180 A. 809 (N.J. 1935) (No. 99). In fact, there is dispute over what the kidnapper actually said. Lindbergh had apparently told the grand jury that the words were, "Hey, doc," KENNEDY, *supra* note 1, at 209. Others quote the kidnapper as having said, "Hey, Doctor! Hey, Doctor, over here." GEORGE WALLER, KIDNAP: THE STORY OF THE LINDBERGH CASE 75 (1961). We thank Ronelle Delmont for providing us with a compact disk containing the transcript.

8. Transcript at 113–14, *Hauptmann* (No. 99).

9. A. SCOTT BERG, LINDBERGH 315 (1998).

10. This is literally true. *See* Frances McGehee, *The Reliability of the Identification of the Human Voice*, 17 J. GEN. PSYCHOL. 249 (1937), discussed *infra* note 169.

allowed to take the stand and say so. As we will see, the American
legal system is all too often wrong in its assumptions about people's
ability to recognize voices. Other countries fare no better. In 1995, a
Canadian appellate court exonerated Guy Paul Morin, whose
conviction three years earlier of raping and murdering a young girl
was largely based on an inaccurate identification of his voice.[11] Post-
conviction DNA testing excluded him as the perpetrator.[12]

Many of the problems that people have identifying speakers
from their voices are similar to those that people have as
eyewitnesses.   The amount of exposure, the nature of the
identification process, and the number of exposures all matter in
determining how likely a witness is to be correct.[13] Yet while the
reliability of *eye*witness identification has been a focal point in the
news, the scholarly literature, and the courts,[14] the unreliability of
*ear*witness identification has gone virtually unnoticed in the case law
and legal literature. The reluctance of the legal system to deal with
this problem stems from a confluence of ignorance, rigid adherence to
historical positions that are no longer tenable, and some interesting
judicial missteps concerning the accuracy of "voiceprints" that have
made courts unreceptive to voice identification research.

Part I of this Article examines the foundation required for a
voice identification to be admissible. The Supreme Court has held
that a party offering eyewitness identification made under suggestive
circumstances must demonstrate indicia of reliability to comply with
due process standards.[15] While courts sometimes apply these same
standards to earwitness identification cases, they often fail to give
earwitness testimony even that much scrutiny, basing their decisions
on the rules of evidence governing authentication, which require very

---

11. *See* Regina v. Morin, 37 C.R. (4th) 395 (Ontario Ct. App. 1995). The voice
identification aspect of the case is discussed in A. Daniel Yarmey, A. Linda Yarmey,
Meagan J. Yarmey & Lisa Parliament, *Commonsense Beliefs and the Identification of
Familiar Voices*, 15 APPLIED COGNITIVE PSYCHOL. 283 (2001).

12. *Morin*, 37 C.R. (4th) 395.

13. There is extensive literature on problems with eyewitness testimony. For a
collection of articles that provide a good introduction to some of the research in this area,
see EYEWITNESS TESTIMONY: PSYCHOLOGICAL PERSPECTIVES (Gary L. Wells &
Elizabeth F. Loftus eds., 1984). *See also* BRIAN L. CUTLER & STEVEN D. PENROD,
MISTAKEN IDENTIFICATION: THE EYEWITNESS, PSYCHOLOGY AND THE LAW (1995).
One study has compared people's ability to identify a target whom they had both seen and
heard for fifteen seconds, five minutes before the test. A. Daniel Yarmey, A. Linda
Yarmey & Meagan J. Yarmey, *Face and Voice Identifications in Showups and Lineups*, 8
APPLIED COGNITIVE PSYCHOL. 453 (1994). The result was that people were more
accurate in visual identification than they were identifying the individual by voice. *Id.*

14. For example, New York has recently joined the growing number of states that
permit expert evidence by defendants to point out potentially unreliable aspects of
eyewitness testimony. *See* People v. Lee, 750 N.E.2d 63 (N.Y. 2001).

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

little foundation.[16] We will argue that some of the cases require so little evidence of reliability that they are unconstitutional even under current standards. We will also point out serious risks of misidentification that call for some adjustment in the legal rules.

In Part II we summarize some of the empirical research regarding the reliability of voice identification, research that the legal system has by and large ignored. This omission is significant. For one thing, if reliability is the basis for constitutional analysis of suggestive identifications, knowing what makes an identification more or less reliable is a prerequisite for intelligent inquiry. Moreover, in establishing such a low threshold for admissibility the system leaves judgments of reliability to the jury. But unless information is presented to jurors through other actors in the legal system, whether through cross-examination, the introduction of expert testimony, or informative jury instructions, jurors will have no reasonable basis for making this judgment.

We discuss a number of factors that have an impact on reliability. Some people are good at identifying voices—others are terrible at it. Memory for voices stays with us for a few weeks, but then degrades if not reinforced. Longer exposures produce better recollection—up to a point. Repeated exposure helps a great deal. People have trouble recognizing voices that they earlier heard with a different tone of voice. People between the ages of sixteen and forty are best at recognizing voices. People are better at remembering voices that they heard live rather than on tape or over the telephone. People are generally bad at recognizing the voices of those speaking foreign languages and are not very good at recognizing the voices of those speaking their own language with a foreign accent.

Part III considers the possible role of experts in speaker identification cases. The legal system's excessive confidence in voice identification technology in the 1960s and 1970s has made the system especially suspicious of experts even several decades later. Courts had placed increasing hope on the ability of machines, called sound spectrographs, to create "voiceprints" unique to each speaker. Earlier, voiceprint analysis had been rejected by courts as unreliable. With the adoption of the Federal Rules of Evidence in 1975 came a softening in judicial resistance. At about the same time, certain prominent phoneticians who had opposed the use of sound spectrography in courts began to support it as the technology improved. Some, but not all courts, began to permit voice comparisons by spectrographic experts. Then, in 1979, the National Academy of Science issued an influential report, arguing that there is insufficient evidence that spectrographic analysis is reliable enough

---

16. FED. R. EVID. 901(b)(5).

for forensic application.[17]   The result was abandonment of spectrographic analysis in court proceedings in most jurisdictions, even those whose courts had approved it just a few years earlier. Nonetheless, a few courts, relying on a superficial understanding of this history, continue to allow it. The Supreme Court of Alaska has only recently permitted its use for the first time.[18]

Unfortunately, this focus on spectrographic analysis has deflected serious inquiry into how good ordinary people are at identifying speakers from their voices and whether experts are any better than untrained people. A great deal has been written on these issues, especially in the past ten years, but almost none of it has made its way into American legal discourse, which associates questions of speaker identification with the voiceprint issue. To be fair, much of the basic research has been conducted outside the United States and is not readily accessible to those who do not know where to look. But it is now time for the system to consider this information and to put it to appropriate use.

As for what experts can do, there is still debate. But experts are at least somewhat better than lay people at comparing voice samples aurally, which may be especially significant in cases requiring comparison of a voice on a tape recording to that of a witness or the defendant. Although courts have been divided on the matter since the Supreme Court's 1993 *Daubert*[19] decision, we do not recommend that the system accept the testimony of those who hold themselves out as voiceprint experts. However, when used in limited circumstances by experts in phonetics, spectrographic information can sometimes augment aural voice analysis in ways that can be helpful to the legal system. Moreover, experts can bring the weaknesses of lay identification to the attention of judges and juries. Courts have long recognized that identification of a person by voice alone presents "grave dangers of prejudice to the suspect."[20]

Finally, in Part IV, the Article concludes with a number of suggestions for improving the legal system's handling of voice identification. Among our recommendations are the adoption of non-suggestive identification procedures, the use of reliability criteria in tune with the scientific research, the admission of expert testimony to bring to the jury's attention aspects of an identification that reduce its reliability, and informative jury instructions.

---

17. COMMITTEE ON EVALUATION OF SOUND SPECTROGRAMS, ASSEMBLY OF BEHAVIORAL AND SOC. SCIENCES, NAT'L RESEARCH COUNCIL, ON THE THEORY AND PRACTICE OF VOICE IDENTIFICATION (1979).

18. State v. Coon, 974 P.2d 386, 400–02 (Alaska 1999).

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

20. Palmer v. Peyton, 359 F.2d 199, 201 (4th Cir. 1966).

I. Legal Standards for Identifying Speakers

Lindbergh's identification of Hauptmann could not have been more suggestive. Having just sworn to a grand jury that he would not likely be able to identify the kidnapper's voice, in a one-on-one encounter Lindbergh was presented with an individual whom the D.A. said was the person who killed Lindbergh's son. That person was Hauptmann. Lindbergh indeed identified Hauptmann's voice and testified at trial that he had done so. Since the Hauptmann trial, the basic evidentiary principle has remained the same: Someone who heard a speaker's voice can take the witness stand and identify that voice, subject to the opposing party's right to cross-examine.

However, during the past three decades, the Supreme Court has held that due process considerations require inquiry into the reliability of an identification when it is found to be suggestive. In Section A, we look at the application of these requirements in earwitness identification cases. In Section B, we examine a set of cases in which the voice being identified is on tape. In those cases, courts apply rules governing authentication and require only minimal familiarity with a voice for an identification to stand. As we will see, courts sometimes pay too little attention to suggestiveness and reliability issues in tape cases and occasionally even apply the minimal standards of tape cases to live identifications.

A. Due Process Requirements

(1) The Biggers Criteria

The seminal case in both eyewitness and earwitness identification is *Neil v. Biggers*, decided by the Supreme Court in 1972.[21] That case involved a crime victim's identification of the defendant at a "showup"—a procedure in which the police march the suspect in front of the victim and ask for identification, without the safeguard of requiring the victim to choose the defendant from among a group of people in a lineup. Approximately seven months after the crime occurred, officers showed the defendant to a rape victim at a police station, where she had an opportunity to look him over and hear him utter the words "shut up or I'll kill you."[22] Based on his appearance and voice, she testified at trial that she had "no doubt" that the defendant was her assailant.[23]

The Court concentrated on the eyewitness aspect of the identification, and established a framework for evaluating claims that

---

21. 409 U.S. 188.
22. *Id.* at 195.
23. *Id.*

a defendant's right to due process was violated by a suggestive identification procedure.[24] The focus should be on the risk of a false identification:

It is the likelihood of misidentification which violates a defendant's right to due process. . . . Suggestive confrontations are disapproved because they increase the likelihood of misidentification, and unnecessarily suggestive ones are condemned for the further reason that the increased chance of misidentification is gratuitous. But as *Stovall* makes clear, the admission of evidence of a showup without more does not violate due process.[25]

The Court then articulated criteria for evaluating the likelihood of misidentification:

We turn, then, to the central question, whether under the "totality of the circumstances" the identification was reliable even though the confrontation procedure was suggestive. As indicated by our cases, the factors to be considered in evaluating the likelihood of misidentification include the opportunity of the witness to view the criminal at the time of the crime, the witness' degree of attention, the accuracy of the witness' prior description of the criminal, the level of certainty demonstrated by the witness at the confrontation, and the length of time between the crime and the confrontation.[26]

Using these criteria, the Court held that the rape victim's identification of the defendant as her assailant was good enough to pass muster.[27] It noted that the defendant's appearance was consistent with a description she had given police shortly after the crime occurred.[28] Moreover, she had previously been shown several other suspects and had failed to single out any one of them as the rapist.[29] The Court was also impressed with her level of confidence in the identification.[30]

In a subsequent case, *Manson v. Brathwaite*, the Court elaborated on the decision in *Biggers*.[31] It held that "the corrupting effect" of suggestive procedures must be balanced against indicia of reliability, which is "the linchpin in determining the admissibility of identification testimony."[32] If the identification is reliable, then it

---

24. *Id.* at 198.

25. *Id.*

26. *Id.* at 199–200. How much influence the *Biggers* criteria have on jurors is not entirely clear. For a study showing that jurors appear to care about the interaction between attention and certainty in evaluating eyewitness testimony, see Amy J. Bradfield & Gary L. Wells, *The Perceived Validity of Eyewitness Identification Testimony: A Test of the Five* Biggers *Criteria*, 24 LAW & HUM. BEHAV. 581 (2000).

27. *Biggers*, 409 U.S. at 200–01.

28. *Id.* at 200.

29. *Id.* at 201.

30. *Id.* at 200–01.

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

32. *Id.* at 114.

should be allowed notwithstanding improperly suggestive procedures.[33] This creates a two-step analysis in cases of this sort. First, a court should ask whether the identification was suggestive, and second, if it was suggestive, whether it was nonetheless reliable under the criteria set forth in *Biggers*.

Although *Biggers* and *Manson* both concentrated on eyewitness identification, it is worth bearing in mind that the victim in *Biggers* was exposed not only to the defendant's appearance, but also to his voice. There is no reason for courts to limit the holdings of these cases to eyewitness identification, and they do not. As we will see below, some courts have applied *Biggers* and *Manson* to voice identification, and no court has said that the two-step analysis should not apply when auditory identification is in issue. Thus, *Biggers* and *Manson* set the constitutional standard for admitting voice identification evidence in earwitness cases.

(2) Due Process Analysis in Voice Identification Cases

The initial question in a *Biggers/Manson* analysis is whether voice identification procedures were overly suggestive. If the procedure is held not to have been suggestive, courts generally do not apply the due process analysis and must decide whether to admit the identification under ordinary rules of evidence. The threshold for admission of a voice identification under these rules is very low.[34]

The surest way to guarantee that voice identifications are not excessively suggestive is to use an appropriately constituted voice lineup. Consider the description by the Supreme Judicial Court of Massachusetts of a permissible procedure using a five-voice lineup:

> After consulting with the office of the district attorney, the police used a voice identification procedure that adequately protected the defendant's rights. There was no one-on-one confrontation between the victim and the defendant. The victim could not see the participants during the procedure, nor could they see her. The defendant selected the order in which he would read. The participants read the same innocuous passage from a fifth-grade reader. Defense counsel attended the procedure and, although consulted, never objected to it. In addition, we have viewed a videotape of the voice identification procedure, and conclude that the procedure was not impermissibly suggestive. The defendant's voice did not stand out because of his age, nor did any other aspect of the procedure direct undue attention to the defendant's voice. Hence, we conclude that the judge properly denied the defendant's motion to suppress the voice identification.[35]

---

33. *See id.*
34. *See infra* note 90 (discussing FED. R. EVID. 901(b)(5)).
35. Commonwealth v. Miles, 648 N.E.2d 719, 728–29 (Mass. 1995).

We do not endorse the Massachusetts procedure as flawless. For example, there probably should be more than five voices in a lineup.[36] But the court was clearly making a reasonable effort to ensure that fair procedures would be used.

In contrast, one-on-one voice identifications are almost inherently suggestive. For example, in *Yeatman v. Inland Property Management, Inc.*, a federal district court rejected an identification when "[o]nly one tape containing only one female voice was played." [37] Moreover, the witness "already knew the critical need to give an affirmative answer to the question that she was being asked. And no opportunity was given to [the opposing parties] to participate in or to monitor the procedure."[38] The court likened the identification process to a "card trick" where "a magician forces on the person chosen from the audience the card that the magician intends the person to select, and then the magician purports to 'divine' the card that the person has chosen."[39] Similarly, *State v. Johnson*, a New Jersey case, involved a woman who was raped by a man whose face she could not see, but whose voice she heard.[40] She was asked to come to the police station, and through an open door she heard the voice of a suspect talking to a detective.[41] After some initial hesitation, she identified the suspect as her assailant.[42] On appeal, the court noted that the constitutional safeguards that apply to visual identifications "are equally applicable to identification of a voice through auditory senses," particularly because the risks of misidentification "are even more apparent where the identification is by voice alone."[43] It concluded that the identification procedure was sufficiently suggestive to require a *Biggers* analysis of reliability.[44]

In a Florida case, a man forced a woman off the road with his van.[45] She never got a good look at his face but heard him say, "lady, I'm going to rape you and kill you."[46] Thirty-two days later she was presented with a short tape recording of a detective interviewing a suspect, who mainly said nothing beyond invoking his right to silence.[47] She identified him as her assailant.[48] Later she attended a

---

36. See *infra* notes 229–31 for literature concerning the appropriate criteria for assembling voice lineups.
37. 845 F. Supp. 625, 628 (N.D. Ill. 1994).
38. *Id.*
39. *Id.*
40. 351 A.2d 787 (N.J. Super. Ct. App. Div. 1976).
41. *Id.* at 789.
42. *Id.*
43. *Id.* at 788.
44. *Id.* at 790–91.
45. Macias v. State, 673 So. 2d 176 (Fla. Dist. Ct. App. 1996).
46. *Id.* at 178.
47. *Id.* at 179.

court session in which the same suspect was arguing that his bond should be reduced, and she again identified his voice as that of her assailant.[49] Because the victim was presented with only one possible voice in each situation, the appellate court held the procedure overly suggestive.[50]

Likewise, a Massachusetts court held that requiring a suspect to utter the words of the perpetrator in open court, followed by the victim's identifying the voice as that of the defendant, was impermissibly suggestive.[51] In addition, a Pennsylvania court found it excessively suggestive when a rape victim identified her assailant based on hearing his taped telephone confession.[52] Courts have also found it improper to have a rape victim overhear a suspect being interviewed at a police station and then ask her if the man was her assailant.[53] Also improper was allowing a witness to see the defendant during the voice identification process, casting doubt as to whether it was really the defendant's voice that the witness was identifying.[54]

Other courts have accepted some questionable identification procedures. One Connecticut case held that a lineup consisting of just two voices was not overly suggestive.[55] The same result was reached in a Louisiana case when a defendant's voice was one of three in a voice lineup.[56]

The second part of the two-step *Biggers/Manson* process is to apply the *Biggers* criteria to determine whether a suggestive voice identification was nonetheless sufficiently reliable.

For example, in *Commonwealth v. Marini*, the Massachusetts court, having found identification procedures unduly suggestive, went on to find the identification unreliable when nine months had elapsed between the crime and the identification.[57]

---

48. *Id.*

49. *Id.*

50. *Id.* at 180.

51. 378 N.E.2d 51 (Mass. 1978).

52. Commonwealth v. Vanderlin, 580 A.2d 820 (Pa. Super. Ct. 1990). The court seems to have largely assumed suggestiveness, at least for the sake of argument, but held that the identification was sufficiently reliable nonetheless. *See also* Jackson v. State, 594 So. 2d 1289, 1295 (Ala. Crim. App. 1991) (detective played only a single tape of a statement by a "suspect" to the victim, who identified his voice; held impermissibly suggestive).

53. State v. Pendergrass, 586 P.2d 691, 696 (Mont. 1978); *see also* State v. Johnson, 674 P.2d 1077 (Mont. 1983) (victim allowed to listen to man being interviewed at police station and identified him as her rapist; procedure held suggestive but sufficiently reliable under *Biggers*).

54. State v. Atkins, Nos. 03C01-9302-CR-00058, 03C01-9302-CR00059, 1994 WL 81524 (Tenn. Crim. App. Mar. 3, 1994).

55. State v. Blevins, 536 A.2d 1002 (Conn. App. Ct. 1988).

56. State v. Pickney, 714 So. 2d 854 (La. Ct. App. 1998).

57. 378 N.E.2d 51, 59 (Mass. 1978).

Most of the time, however, courts find adequate reliability. *United States v. Duran*[58] is typical. There, the Ninth Circuit affirmed a conviction for bank robbery.[59] Key evidence consisted of the tellers' identification of the defendant's voice at trial.[60] In response to the defendant's argument that the identification was excessively suggestive, the Court applied the *Biggers* factors, which the Ninth Circuit had adopted for in-court eyewitness identifications.[61] The Court concluded:

> Again, both tellers had ample opportunity to listen to Duran's voice during the robbery. Duran ordered the tellers to raise their hands and demanded money. He ordered [a] teller to escort him into the vault and to open it up. Inside the vault, Duran continued to holler at [the teller], demanding the keys to the vault, telling her to hurry, and asking where all the money was. He ordered [the teller] back to the teller area and demanded the keys to the remaining cash drawers. As Duran left, he threatened everyone in the bank: "don't move or we'll kill you." Both tellers were likely very attentive during the robbery given Duran's weapon and threats, as evidenced by their accurate descriptions of Duran and his distinctive voice and the fact that neither teller equivocated in her identification of Duran's voice. Moreover, the in-court identifications occurred just three months after the bank robbery.[62]

This analysis contains some questionable assumptions. For one, three months may be a long time to remember a voice. For another, the court's conclusion about the degree of attention is only that the tellers were "likely attentive."[63] The opinion demonstrates a judicial recognition that reliability is an issue but does not provide much analysis of what makes an identification reliable or unreliable.[64]

A more convincing case for reliability was made in *United States v. Degaglia.*[65] Although the Seventh Circuit did not explicitly apply the *Biggers* factors, it took seriously the fact that the agent identifying the defendant's voice testified that he had heard it on several occasions for periods of up to one and one-half hours and that the defendant had a very distinctive voice, which the agent described as "high pitched, raspy, and nasal."[66] Here, our intuitions are that the identification is more likely to be reliable. Courts, in fact, frequently

---

58. 4 F.3d 800 (9th Cir. 1993).

59. *Id.* at 802.

60. *Id.* at 802–03.

61. *Id.* at 803; *see also* United States v. Carbajal, 956 F.2d 924 (9th Cir. 1992).

62. *Duran*, 4 F.3d at 803.

63. *Id.*

64. Occasionally, state courts apply the *Biggers* criteria in voice identification cases, especially when the issue is the distinction between a lineup and a showup. *See* Ohio v. Mallet, No. 76608, 2000 WL 1176880 (Ohio Ct. App. Aug. 17, 2000).

65. 913 F.2d 372 (7th Cir. 1990).

66. *Id.* at 374.

hold identifications to be reliable when the witness testifies to having heard the voice on multiple occasions.[67] We will see that repeated exposures to a voice really does have significant impact on accuracy in experimental studies.[68] The research suggests that the inverse is also true: People do far worse identifying a voice that they have heard only once.[69]

In summary, when an identification has occurred under suggestive circumstances, courts require some indication that the identification was nonetheless reliable as a condition for admissibility. Non-suggestive identifications, in contrast, are not subjected to scrutiny of their reliability. That issue is left to the trier of fact. In Part II, we will explore factors that affect the reliability of voice identification.

B. Voices on Tape

(1) Rule 901 and the Minimalist Approach

When the voice being identified is on tape, courts often do not engage in the *Biggers/Manson* analysis. They do not consider how suggestive the identification was and do not analyze the indicia of reliability even when the identification was suggestive. Rather, applying Rule 901 of the Federal Rules of Evidence or a similar rule, they permit a witness, often a police officer, to identify the voice and leave the question of reliability to the jury.[70]

Rule 901 states:

Requirement of Authentication or Identification

(A) General provision. The requirement of authentication or identification as a condition precedent to admissibility is satisfied by evidence sufficient to support a finding that the matter in question is what its proponent claims.

---

67. *See* Virgin Islands v. Sanes, 57 F.3d 338 (3d Cir. 1995) (finding identification sufficiently reliable when fifteen days elapsed between last attack and positive voice identification and victim making identification testified that she had engaged assailant in conversation during two separate attacks with hope that she would be able to identify his voice); Tate v. Morris, No. 89-3570, 1990 WL 117367 (6th Cir. Aug. 14, 1990) (*Biggers* criteria met when victim heard rapist's voice on many occasions, including phone calls made subsequent to the assault).
68. *See infra* note 158.
69. *See* Part II.A.2, *infra*.
70. The Second Circuit has actually held that the due process considerations relevant to lineups do not apply to a voice identification under Rule 901(b)(5), at least with respect to the facts of the case under consideration. United States v. Albergo, 539 F.2d 860, 864 (2d Cir. 1976).

(B) Illustrations. By way of illustration only, and not by way of limitation, the following are examples of authentication or identification conforming with the requirements of this rule:

. . . .

(5) Voice identification. Identification of a voice, whether heard firsthand or through mechanical or electronic transmission or recording, by opinion based upon hearing the voice at any time under circumstances connecting it with the alleged speaker.[71]

The advisory committee notes accompanying the rule make it clear that experts should generally not be part of the process: "Since aural voice identification is not a subject of expert testimony, the requisite familiarity may be acquired either before or after the particular speaking which is the subject of the identification, in this respect resembling visual identification of a person rather than identification of handwriting."[72]

The rule requires only "evidence sufficient to support a finding" that the tape is what it purports to be, and just about anything is sufficient.[73] The following statement by the United States Court of Appeals for the Ninth Circuit is typical:

Rule 901(b)(5) establishes a low threshold for voice identifications offered to determine the admissibility of recorded conversations. So long as the identifying witness is "minimally familiar" with the voice he identifies, Rule 901(b) is satisfied. The record reveals that Speziale was present in Anchorage at Plunk's initial post-arrest interview. The familiarity that he gained through that exposure was sufficient under Rule 901(b)(5) to support his identification of Plunk's voice on the tape recorded conversations being offered into evidence.[74]

The court took it as a given that the witness had gained sufficient familiarity with the suspect's voice to identify it, despite his obviously very limited exposure.

In many, if not most cases, this "minimally familiar" approach of Rule 901 does not appear to create significant risk of misidentification. For one thing, the identification often *does* have substantial indicia of reliability. For example, sometimes the person identifying the voice actually participated in the tape-recorded conversation.[75] In other cases, the identifying witness almost certainly

---

71. FED. R. EVID. 901(b)(5).

72. FED. R. EVID. 901 advisory committee's notes. In light of these remarks, the word "opinion" in the rule is best understood as referring to lay opinion.

73. Of course, other rules apply as well. For example, the content of the tape must be relevant to issues in the case. *See* FED. R. EVID. 402.

74. United States v. Plunk, 153 F.3d 1011, 1023 (9th Cir. 1998) (citations omitted).

75. United States v. Carrasco, 887 F.2d 794 (7th Cir. 1989) (person who identified voices was an eyewitness and an active participant in the recorded conversations); People

was familiar enough with the voice to identify it correctly. Experience and research[76] support the intuitions of judges that people typically can identify the voice of a close relative[77] or that someone who has heard a voice fifty or sixty times is likely to recognize it if he hears it on a tape.[78]

Moreover, the circumstances under which the tape was made usually provide ample evidence of reliability to reduce concerns about possible due process violations. When an officer investigating the defendant's conduct has recorded a wiretapped conversation between two people, and one of the speakers used the wiretapped phone at the defendant's residence, responded to being called by the defendant's name, and said the kinds of things the defendant often said, the defendant was most likely one of the speakers.[79] Thus, when the circumstantial evidence is robust, the risk of error is rather low, even if the identifying witness's only experience with the voice was when he heard the defendant speak six months earlier at the defendant's arrest.[80]

Perhaps the most compelling circumstance is the existence of the tape itself. Of course, it is possible to misidentify a voice on a tape. But the tape limits the range of plausible identifications, and its very existence provides the defendant with the opportunity to dispute the identification through witnesses who testify to the contrary. Furthermore, if a police officer identifies the speaker based on minimal exposure to his voice, the identification obviates the need to find and subpoena neighbors, former teachers, or others to do the same thing. If it does not add significantly to the likelihood of misidentification (an issue to which we return below), then the *pro forma* authentication is efficient and not unfair.

In fact, courts have long permitted voice identification solely by circumstantial evidence when the voice is on the telephone at a number that the caller contacted. In one 1924 case decided by the

---

v. Griffin, 592 N.E.2d 930 (Ill. 1992) (participant in conversation testified that the tape recording was accurate and identified the voice of the defendant); Hasley v. State, 786 S.W.2d 733 (Tex. App. 1989) (police officer secretly recorded defendant and then later identified his voice on the tape).

76. *See* Part II *infra.*

77. *See* Mutz v. State, 862 S.W.2d 24 (Tex. Ct. App. 1993) (permitting identification of ex-father-in-law's voice when witness and defendant lived near each other and their relationship had lasted seven years).

78. United States v. Phavong, No. 98-50230, 1999 U.S. App. LEXIS 20882 (9th Cir. July 12, 1999); *see also* United States v. Puentes, 50 F.3d 1567 (11th Cir. 1995) (permitting agent to testify that he knew voice on a particular tape belonged to defendant based on agent's hearing defendant's voice frequently during a two-month wiretap).

79. *See* United States v. Ladd, 527 F.2d 1341, 1342 (5th Cir. 1976).

80. *See id.*

First Circuit, three defendants appealed their mail fraud convictions.[81]
Part of the evidence involved telephone confirmations of dishonest
stock sales.[82]   Regarding the identification of the voices on the
telephone as belonging to the defendants, the court held:

> The gist of the problem is whether there was sufficient
> identification of the persons in the Boston office calling the
> witnesses, or later responding to the calls of the witnesses, so as to
> justify the court in admitting the evidence. Plainly, recognition of
> the voice is not the only means of identification. Circumstantial
> evidence may be as persuasive as testimony that the voice is
> recognized.[83]

Since the adoption of the Federal Rules of Evidence in 1975,[84]
federal courts continue to permit identification of voices on the
telephone by circumstantial evidence alone.[85] Occasionally, courts do
not even require that a taped voice be authenticated by someone with
knowledge of the voice, relying on telephone identification cases, and
not on Rule 901(b)(5).[86]

Professors Risinger, Denbeaux, and Saks point to similar issues
in the law governing the identification of documents.[87] While the
most compelling reason to consider a document as authentic often
involves the circumstances in which it was discovered (e.g., in the
defendant's desk drawer), the system also requires some formal
identification of the defendant as the document's author.[88] The
acceptance of handwriting experts grew out of the need to provide an
identifying witness when there was no other witness available.[89] To
the extent that courts treat the personal identification as a formality,
relying principally on the circumstantial evidence, the issues raised
with respect to speaker identification mimic those that arise in the
authentication of documents.[90]

---

81. Lewis v. United States, 295 F. 441 (1st Cir. 1924).

82. *Id.* at 443.

83. *Id.* at 444.

84. FED. R. EVID. 901(b)(6) (dealing with the admissibility of telephone
conversations). For a discussion of what circumstantial evidence courts require, see 5
JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN'S FEDERAL EVIDENCE §
901.08 (Joseph M. McLaughlin ed., 1998).

85. United States v. Hernandez-Garcia, No. 98-1750, 2000 U.S. App. LEXIS 2258, at
**4–5 (1st Cir. Feb. 15, 2000); United States v. Degaglia, 913 F.2d 372, 376 n.4 (7th Cir.
1990).

86. *See* United States v. Puerta Restrepo, 814 F.2d 1236, 1241 (7th Cir. 1987).

87. D. Michael Risinger, Mark P. Denbeaux & Michael J. Saks, *Exorcism of Ignorance
as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification
"Expertise"*, 137 U. PA. L. REV. 731, 752 (1989).

88. *Id.*

89. *Id.* at 754–59.

90. The advisory committee's note to Rule 901(b)(5) distinguishes voice and
handwriting identifications. Rule 901(b)(2) requires nonexperts to have familiarity with

When defendants raise objections to intuitively unreliable identifications of voices on tape, courts typically admit the evidence anyway, leaving it to the defense to attack its validity at trial. In essence, once a tape is admitted, its contents are considered conditionally relevant,[91] depending on whether the jury ultimately decides that it really is the defendant's voice on the tape. Below is an excerpt from a recent Sixth Circuit case, *United States v. Knox.*[92] The officer identifying the defendant's voice had heard it only once some three years before his testimony at trial:

> Special Agent Collins's testimony that he recognized Sam's voice as that of Knox based on a conversation some three years earlier was viewed with some skepticism by the district court. There is no question, however, that this testimony is adequate to authenticate Sam's voice as that of Knox for purposes of admissibility in conformance with Fed. R. Evid. 901. All that is required under the rule is that the identifier, Collins, have heard the voice of the alleged speaker, Knox, "at any time." If the district court meant to set a stricter standard—about which point we are, concededly, uncertain—it abused its discretion. Of course, it will certainly be open to the defendant to argue to the jury that Collins is simply wrong, and that it is improbable that Collins could remember in 1996 a voice he heard in 1993. This is a question of the weight of the evidence, however, not of its admissibility.[93]

Even on its own terms, the Sixth Circuit's opinion reveals the need for greater linguistic sophistication in voice identification cases. The court invites defendants to argue that the witness identifying the voice "is simply wrong, and that it is improbable" that he could remember a voice that he had heard three years earlier.[94] But what makes it improbable? If people typically remember voices that well, then it is not improbable. If they don't, then it is improbable. This is, in other words, an issue that can and should be informed by scientific research. As an initial matter, it will be up to defense attorneys to educate themselves sufficiently to raise these issues, whether in pretrial motions or at trial.

Likewise, the Fifth Circuit affirmed a conviction that rested in part on the expert testimony of an FBI agent who identified the voice on a tape as the defendant's.[95] It did not matter that the FBI "used regular agents to make the identifications rather than using the voice identification experts it has on staff," or that "the prosecution

---

handwriting "not acquired for purposes of the litigation." In contrast, a voice on tape may be identified by someone who became familiar with the voice after the tape was made.

91. *See* FED. R. EVID. 104(b), 401.

92. No. 97-5492, 1998 U.S. App. LEXIS 27655 (6th Cir. Oct. 22, 1998).

93. *Id.* at **9–10 (citation omitted).

94. *Id.* at *9.

95. United States v. Lampton, 158 F.3d 251 (5th Cir. 1998).

admitted it had misidentified some of the parties on these phone calls."[96]

Applying Rule 901, the Tenth Circuit has also required very little evidence to support voice identifications,[97] and the Second Circuit has specifically rejected what it called a "rigid standard" regarding the admissibility of tape-recorded evidence, noting that authentication merely makes the tapes admissible, leaving the issue of reliability to the jury.[98] Similarly, an appellate court in New York state was recently satisfied when an officer identified a voice on a tape based on "a lengthy conversation he had with defendant on the day of the arrest" some fifteen months earlier.[99]

The Seventh Circuit seems to apply a somewhat stronger standard of reliability. For a tape recording to be admissible, the government must establish by *clear and convincing evidence* that it is a "true, accurate, and authentic recording of the conversation, at a given time, between the parties involved."[100] Yet what appears to be a higher burden of proof largely dissipates in the great deference that appellate courts give to the trial judge's decision. Such a decision is not overturned on appeal absent "extraordinary circumstances."[101]

Consider, for example, the Seventh Circuit's holding that a two-hour conversation that a police officer had in English with a defendant some two years earlier was enough to permit the officer to identify a voice on a tape as being that of the defendant.[102] Although reciting the "clear and convincing" standard,[103] the court held that "questions concerning the length of Officer Johnson's previous contact with Vega or the time between this contact and the identification, simply go to the weight the jury accords this evidence, not to its admissibility."[104] It made no difference that the voice on the tape was speaking Spanish. Thus, even jurisdictions that appear to have a somewhat higher standard turn out, in reality, to apply the rule similarly.

---

96. *Id.* at 259; *see* United States v. Cuesta, 597 F.2d 903, 915 (5th Cir. 1979) ("Once this minimal showing has been made, the jury determines the weight to accord the identification testimony.").

97. United States v. Axselle, 604 F.2d 1330 (10th Cir. 1979) (permitting DEA agent to identify voice on tape of a telephone call as defendant's, having heard defendant's voice only once at a court hearing that occurred thirty days after the tape was made).

98. United States v. Tropeano, 252 F.3d 653, 661 (2d Cir. 2001).

99. People v. Rendon, 709 N.Y.S.2d 698, 701 (N.Y. App. Div. 2000).

100. United States v. Faurote, 749 F.2d 40, 43 (7th Cir. 1984).

101. *Id.*

102. United States v. Vega, 860 F.2d 779 (7th Cir. 1988).

103. *Id.* at 784.

104. *Id.* at 788.

(2) Some Problems with the Minimalist Approach of Rule 901

The Supreme Court has held that the key issue for purposes of due process analysis is the likelihood of misidentification.[105] In many cases involving tapes, the identification procedure is suggestive, and the witness's exposure to the voice is so minimal and, at times, so long ago, that the identification cannot seriously be considered reliable.[106] In these cases, only the circumstances taken as a whole can provide the indicia of reliability necessary to meet due process considerations. When those indicia are absent, courts should take a closer look, notwithstanding Rule 901's minimal standards.

Most opinions do not describe the procedure by which the witness identified the defendant's voice. Yet it is only reasonable to infer from these cases that the procedure is typically highly suggestive. Much of the time, it appears that the prosecutor contacts an officer who heard the defendant's voice, perhaps during an earlier arrest, and asks him to identify it on a tape.[107] If it were otherwise, the many opinions that reject challenges to authentications would highlight additional facts that show the identification to be non-suggestive. Moreover, if courts do not require more, why should the prosecutor risk being unable to authenticate the tape by making the identification procedure harder for the police officer than the law requires?

Most courts have not applied *Biggers/Manson* criteria to the identification of voices on tape. The few that have done so have invariably found no due process violation, even when the suggestiveness of the identification rings out. Consider *United States v. Zambrana*, where a government agent identified the defendant's voice as being on a tape recording and the court admitted it under Rule 901.[108] The evidence of suggestiveness was particularly strong. As the agent listened to a tape, he had a transcript that listed the defendant's name in the margin.[109] Nonetheless, the court held that

---

105. Neil v. Biggers, 409 U.S. 188, 198 (1972).

106. *See, e.g.,* United States v. Knox, No. 97-5492, 1998 U.S. App. LEXIS 27655 (6th Cir. Oct. 22, 1998). See Part II *infra* for a discussion of the factors affecting the reliability of an identification.

107. See *United States v. Pheaster*, 544 F.2d 353, 369 (9th Cir. 1976), in which the court described the relevant facts as follows:

On July 2, 1974, an FBI agent asked Officer Turley whether he knew Pheaster and whether he could identify his voice. Upon receiving affirmative answers, the agent played for Officer Turley a tape-recorded telephone conversation between Mr. Adell and one of the kidnappers. Officer Turley identified the caller as Pheaster.

108. 841 F.2d 1320 (7th Cir. 1988).

109. *Id.* at 1338. The problem in *Zambrana* was compounded by providing the jury with transcripts during the trial, once again with the defendant's name in the margin. The court of appeals stressed that the judge did not admit the transcripts into evidence, warned

this was not overly suggestive because the agent then went on to identify the defendant's voice on additional tapes without a transcript.[110]   Clearly the damage had been done.   An initially suggestive identification can taint later ones, and that was precisely the risk in this case.[111]

Similarly, in *United States v. Degaglia*, the defendant was recorded speaking on the telephone with a government informant.[112] Later, he was arrested by an agent Olson, who spoke with him for around ninety minutes at that time and also on a couple of subsequent occasions.[113]   At trial, Olson identified the defendant's voice as being on the tape.[114]   The court rejected the defendant's contention that the identification was overly suggestive, even though Olson knew he was being called to court to identify the defendant's voice on the tape.[115]   If Degaglia wanted to claim that it was not his voice on the tape, he would have to do so at trial.[116]

The procedure in *Degaglia* was very much like a showup. Rather than being presented with a number of candidates from which to choose, the identifier was just being asked to answer "yes" or "no" to a single proposed candidate. Almost all courts, as we observed in the previous section, would regard this procedure as impermissibly suggestive in the earwitness context and would require a *Biggers* analysis of reliability. In contrast, courts that look at suggestiveness in the context of authentication of tapes by the police have found the procedures to be adequate.[117]

At the same time, it appears that the objections that defendants raise to these identifications are almost always formal ones. The defendant claims that the identification did not meet legal standards

---

the jury that it could not consider the names in the transcripts to constitute substantive evidence, and did not let the jury take the transcripts with them during their deliberation. *Id.* at 1339. Still, this procedure was quite suggestive, and in this day and age of word processors it is quite unnecessary. It would be easy enough to use pseudonyms, or letters like *A* and *B*. In another case in which a jury received transcripts with names in it, the trial lawyers failed to object, and the issue was lost on appeal. United States v. Vega, 860 F.2d 779 (7th Cir. 1988).

110. *Zambrana*, 841 F.2d at 1339.

111. In the context of eyewitness identification, this phenomenon has been noticed in both the scholarly literature and the case law. *See, e.g.,* Foster v. California, 394 U.S. 440, 442–43 (1969). For a recent discussion, see Bruce W. Behrman & Sherrie L. Davey, *Eyewitness Identification in Actual Criminal Cases: An Archival Analysis*, 25 LAW & HUM. BEHAV. 475 (2001).

112. 913 F.2d 372 (7th Cir. 1990).

113. *Id.* at 374.

114. *Id.*

115. *Id.* at 376–77.

116. *Id.* at 378.

117. *See, e.g.,* United States v. Plunk, 153 F.3d 1011 (9th Cir. 1998); United States v. Alvarez, 860 F.2d 801 (7th Cir. 1988).

but provides no reason to believe that the identification is not accurate, by producing alibi witnesses or witnesses who disagree with the identification, for example. The absence of such evidence tends to support the notion that Rule 901's minimal standards regarding the identification of speakers on tape recordings do not generally result in a serious risk of misidentification.

Nonetheless, there are several situations in which we believe the system to be far too casual. First, if a defendant comes forward with facts that raise concerns of a mistaken identification, then courts should apply *Biggers* and evaluate the potential for misidentification before admitting a suggestive identification of a taped voice based on minimal exposure to it. Otherwise, we agree with most courts that such analysis is generally not necessary as a prerequisite for admission of a tape recorded voice into evidence. This balance requires greater indicia of reliability when lax standards for admissibility create a demonstrable risk of error. The suggested procedure, we believe, is both loyal to the streamlined authentication process envisioned in Rule 901 and respectful of the due process concerns that pervade this area of law. On the one hand, courts will not be tempted to say that suggestive identifications are not suggestive, creating precedents that might affect later decisions when due process analysis really matters. On the other hand, when evidence is produced that casts doubt on a suggestive identification, the mechanical nature of Rule 901 will not override due process concerns.

Second, courts should be careful not to admit a suggestive identification of a recorded voice when the circumstantial evidence is not strong. Consider a recorded bomb threat made from a public telephone in a major city. Suppose that police arrest a malcontent whom they suspect to have made the call. Should we allow one of the arresting officers, who heard the suspect say a few words while being booked, to testify that the voice on the tape belongs to the defendant? Not only might the call have been made by any one of millions of people, but telephones transmit only a limited range of acoustic information, which is even further degraded by a tape recorder.

Third, when there is no tape recording, the *Biggers/Manson* analysis should clearly apply. A recent Fifth Circuit case applied only the minimalist approach of Rule 901 in a case where a police officer identified the defendant's voice as the one he had heard in a telephone surveillance.[118] No tape was made. The facts do not show whether the identification was suggestive enough to trigger application of the *Biggers* criteria, but the issue should have been

---

118. United States v. Townsend, No. 97-60491, 1999 U.S. App. LEXIS 13872 (5th Cir. June 24, 1999).

raised and discussed. Courts should not give the impression that due process considerations are suspended when it is the police who make the identification. As noted earlier, the Second Circuit has defended Rule 901's minimalist approach to identifying voices on tape precisely because the existence of the tape reduces the likelihood of misidentification.[119] When no tape exists, the rationale disappears.

In deciding whether an identification meets due process standards, courts must determine the reliability of lay voice identification. Moreover, once the evidence is admitted, jurors must determine how reliable the identification was. As we will see, however, people are not very good at identifying voices, especially under certain circumstances that have been the subject of considerable study. And despite the tendency of many courts to leave the issue to the jury, people do not always have good intuitions about how much to rely on the accuracy of earwitness identifications. We turn to the question of reliability below.

II. Voice Recognition Research and the Reliability of Identifications

People make mistakes identifying voices even under the best of circumstances. Guy Paul Morin's DNA exoneration in Ontario is a startling recent reminder.[120] One of the earliest known cases of speaker misidentification is the trial of William Hulet, who was accused of having executed King Charles I.[121] Once the monarchy was restored under Charles II, one of its first orders of business was to prosecute for treason those involved in the regicide. The evidence against Hulet consisted almost entirely of rumor and innuendo, much of which would be excluded as hearsay today. Especially probative was testimony by Richard Gittens, who not only was a witness to the execution, but also belonged to the same regiment as Hulet did at the time. Gittens testified that he had heard the executioner, whose face was obscured, beg the king's forgiveness and that he knew that it was Hulet "by his speech."[122] Cross-examined later by Hulet himself, who asked him how Gittens knew that he (Hulet) had been on the scaffold at the time, Gittens replied, "By your voice."[123] After deliberating for more than the usual time, the jury returned to declare Hulet guilty of high treason,[124] the punishment for which was normally a quite gruesome death. This case might seem to be of little more than

---

119. *See* United States v. Albergo, 539 F.2d 860 (2d Cir. 1976).
120. *See supra* note 11.
121. Trial of William Hulet, 5 Howell's State Trials 1185 (1660).
122. *Id.* at 1186.
123. *Id.* at 1187.
124. *Id.* at 1195.

passing historical interest but for a footnote inserted by the editors. They report that the actual perpetrator was the ordinary hangman, who later confessed, and that the court, "being sensible of the injury done to [Hulet], procured his reprieve."[125]

Rule 901(b) assumes that people are basically accurate in identifying voices and realistic in their assessments of how likely it is that an identification is correct. In this section, we examine some research that suggests these assumptions are not entirely correct. First, we look at research that addresses how good people generally are at identifying voices. We will see that people are very accurate when it comes to recognizing voices they know well but are much less so with unfamiliar ones. We will also see that issues such as the number of exposures to the voice, the delay in making the identification, the skill of the identifier, and the presence of stress or disguise, all play roles in determining whether an identification is likely to be accurate.

We then ask whether we are realistic in our estimates of how accurate identifiers are likely to be. We will see that there is little or no relationship between confidence and accuracy, but jurors are likely to take the identifier's level of confidence very seriously. We will also see that people tend to overestimate the ability of others to identify voices.

A. Factors Affecting the Reliability of Voice Identification

Researchers have uncovered a number of factors that make voice identification easier or harder. Much of this work has been conducted in Europe, Canada, and Australia and has thus been less accessible to the American legal community.[126] These researchers have found that familiarity with a voice, knowing in advance that one will later have to identify a voice, length of exposure, the language being spoken, foreign accents, length of the delay in performing the identification, and other factors play significant roles in people's ability to identify voices. Most of these factors are completely absent from any discussion in the case law. In contrast, a witness's confidence in the accuracy of the identification, which courts sometimes consider relevant, does not correlate substantially with correctness of identification. In this section, we look at empirical data that teases out many of these factors.

---

125. *Id.* at 1185 n.†.

126. One of the most prolific writers on this subject is Professor Yarmey, a Canadian psychologist. For a good summary of the research in this area, see A. Daniel Yarmey, *Earwitness Speaker Identification*, 1 PSYCHOL. PUB. POL'Y & L. 792 (1995).

(1) Familiarity

Just about everybody would assume that people are better at identifying familiar voices than unfamiliar ones. The assumption is largely correct, yet questions remain. How much difference does familiarity make? Does it matter much *how* familiar the voice is? What is the rate of error despite familiarity?

Some of these issues have recently been studied by Daniel Yarmey and a group of his colleagues.[127] In one study, sixty-eight people agreed to participate as "speakers."[128] Each recorded a sixty-four-word passage, and then two minutes of spontaneous speech.[129] The speakers were asked for the names of friends and associates who might participate in a subsequent voice identification study.[130] The speakers also identified themselves as belonging to one of the following categories with respect to each such friend or associate:

A high familiar speaker: "[A]n immediate family member or best friend."[131]

A moderate familiar speaker: "[A] co-worker, team-mate, club-mate, or general friend."[132]

A low familiar speaker: "[A] casual acquaintance, such as next door neighbor or associate, who would be expected to have talked with the listener for only a few minutes on occasion in any week over the last year."[133]

The speakers were asked not to discuss the experiment with any of the people that they named.[134]

For each listener, the experimenters were able to find at least one speaker who was a high familiar speaker, one who was a moderate familiar speaker, and one who was a low familiar speaker.[135] The listeners were then presented with passages from four different voices: three that varied with degree of familiarity, and also an entirely unfamiliar voice.[136] Listeners were asked to say who the speaker was, if they could, as soon as they recognized the voice.[137] They then listened to the rest of the passage and were permitted to change their minds if they thought they had initially made a

---

127. Yarmey et al., *supra* note 11.
128. *Id.* at 286.
129. *Id.* at 286–87.
130. *Id.* at 287.
131. *Id.*
132. *Id.*
133. *Id.*
134. *Id.*
135. *Id.* at 284.
136. *Id.* at 287.
137. *Id.*

mistake.[138] This experimental format should produce more correct responses than one in which subjects do not have a chance to change their minds.[139]

Some of the results of this study[140] are summarized in Table 1:

TABLE 1
Accuracy (percent) for Identifying Voices of Varying Familiarity[141]

FAMILIARITY	CORRECT RESPONSE RATE	FALSE ALARM RATE
HIGH	89	6
MODERATE	75	12
LOW	66	20
UNFAMILIAR	61	36

These results are striking in several ways. First, as expected, familiarity does matter. We are pretty good at recognizing the voices of people we know well (89% correct), not as good at identifying the voices of people we know casually (66% correct), and even worse at acknowledging that we don't know a voice at all (61% correct).

In addition, many of the errors are false alarms: identifiers say they recognize a voice as belonging to a particular speaker but are wrong. Because listeners had the choice of stating that they did not recognize a voice, one would have expected the total of "don't know" answers to increase as familiarity lessens. That was not the case, however. Instead, the false alarm rate went up as familiarity went down. Moreover, false alarms account for a substantial percentage of the errors for the unfamiliar voices; no less than 36% of subjects claimed to recognize a voice they had, in fact, never heard before.

The Yarmey group's study is not unique in this finding. Earlier work by Harry Hollien and his colleagues had reached a similar conclusion.[142] Hollien's team presented subjects with recordings of familiar and unfamiliar voices and then immediately tested them by asking them whether a series of voices matched the one they had

138. *Id.* at 284.

139. In fact, it does. In another experiment reported in Yarmey et al., *supra* note 11, subjects were not given a chance to change their minds. The correct response rates were lower for three of the four levels of familiarity (high=85%, moderate=79%, low=49%, unfamiliar=55%).

140. There were many other results, some of which we discuss below. For example, listeners had a much harder time identifying the voices of whisperers. *See infra* Part II.A.6.

141. Yarmey et al., *supra* note 11, at 291–92.

142. *See* Harry Hollien, Wojciech Majewski & E. Thomas Doherty, *Perceptual Identification of Voices Under Normal, Stress and Disguise Speaking Conditions*, 10 J. PHONETICS 139 (1982).

heard.[143] The study found that when a normal tone of voice was used in the recording (as opposed to a stressed or disguised voice),[144] subjects identified familiar voices with 98% accuracy whereas accuracy dropped with unfamiliar voices to only around 40%, even with almost no lapse in time between the initial exposure and the identification.[145] Contrary to what most people would expect, fewer than half of the subjects were able to identify a previously unfamiliar voice they had heard only a brief time before.

These results confirm our intuitions that people are generally good at recognizing familiar voices. Yet they show remarkably high rates of error in identifying unfamiliar voices. The assumption made by many judges, that someone familiar with a voice can correctly identify it, thus appears to be partially correct. It is true that someone who is highly familiar with a voice can correctly select it from a limited range of alternatives. However, the presumption made by many courts that a policeman who briefly hears a voice once can later identify it on tape seems quite questionable and becomes more questionable as the number of potential target voices increases. The high rate of mistaken identification of unfamiliar voices, which parallels findings regarding eyewitness identification, is especially troubling because of its potential to lead to false convictions.

(2) Amount of Exposure

From the earliest days of voice identification research, experimenters have asked how much exposure to a previously unfamiliar voice is sufficient.[146] Whether we are concerned about a rape victim identifying the voice of her attacker, or a police officer identifying the voice of the defendant as the one on the tape, the legal system routinely deals with situations in which the witness identifying a voice had only brief exposure to it.

In another set of experiments performed by Professor Yarmey, subjects participated in a telephone conversation with the experimenter.[147] The length of the conversation was either short (average 3.2 minutes), medium (average 4.3 minutes), or long (average 7.8 minutes).[148] Subjects then received a second phone call,

---

143. *Id.* at 141.

144. We discuss stressed and disguised voices *infra* Part II.A.6.

145. Hollien et al., *supra* note 142, at 142.

146. *See* I. Pollack, J.M. Pickett & W.H. Sumby, *On the Identification of Speakers by Voice*, 26 J. ACOUSTICAL SOC'Y AM. 403 (1954) (finding that duration of the sample is important to the extent that it gives the hearer exposure to a larger array of different sounds articulated by the speaker).

147. A. Daniel Yarmey, *Voice Identification Over the Telephone*, 21 J. APPLIED SOC. PSYCHOL. 1868 (1991).

148. *Id.* at 1870.

and were asked if they could identify the voice they heard in the first call out of a lineup of six voices presented in the second call.[149] Half the subjects heard a lineup that did not contain the first voice at all.[150] The other half heard a lineup that did contain the first voice.[151] Some of the subjects received this second call immediately after the first one (immediate test), some received it two hours later (two-hour delay), and others received it two or three days later (two/three-day delay).[152]

The results, once again, are not intuitively evident. First, the interval between the two calls did not produce statistically significant results, suggesting that vocal memory does not decay for two or three days, given adequate exposure.[153] Second, the length of the original exposure did matter.[154] For subjects receiving a lineup that actually contained the target voice, 24% who had a short original conversation identified it, while 48% who had a long conversation identified it.[155] Third, the rate of false alarms went up among those receiving a lineup containing the target voice when the exposure to the voice was longer (14% versus 35%), and was even higher (48%, 51%, and 44%) for all three lengths of exposure when the target was not present.[156] Consistent with what we saw in the previous section, people asked to participate in a voice identification procedure seem predisposed to identifying someone, even if that means making a mistake.[157]

Other researchers have found that the *number* of initial exposures to a voice (not just the *length*) is of critical importance. Defenbacher and his team report a study in which one group of listeners heard a sixty-second passage to which they were told to pay close attention.[158] When asked to identify that voice two weeks later out of a voice lineup containing nine voices, they were correct 29% of the time, made false alarms 14% of the time, and the rest of the time did not know.[159] A second group of subjects heard the same sixty seconds of speech, but it was divided into fifteen to twenty second segments and presented to them over the course of three consecutive

---

149. *Id.* at 1871.
150. *Id.*
151. *Id.*
152. *Id.*
153. *Id.*
154. *Id.*
155. *Id.* at 1872.
156. *Id.*
157. *Id.* at 1874–75.
158. Kenneth A. Defenbacher, John F. Cross, Robert E. Handkins, June E. Chance, Alvin G. Goldstein, Richard Hammersley & J. Don Read, *Relevance of Voice Identification Research to Criteria for Evaluating Reliability of an Identification*, 123 J. PSYCHOL. 109 (1989). The reported study was conducted by two co-authors of the article, A.G. Goldstein and J.E. Chance.
159. *Id.* at 111–12.

days.[160] Their hit rate was a perfect 100%.[161] The authors concluded that a witness who hears sixty seconds of speech on one occasion is less likely to recognize the suspect's voice later than is someone who hears fifteen to twenty segments on three or four separate occasions.[162]

Interestingly, when subjects heard a passage only half as long, even the three-day distribution did not rescue them from poor performance. Apparently, exposure to thirty seconds is not enough to support recollection two weeks later, whether the passage is heard in its entirety or in separate segments.[163]

The research thus shows that both amount and frequency of exposure are significant in identifying a previously unfamiliar voice. Consistent with our intuitions, a longer initial exposure will lead to a more reliable identification later. Less intuitively obvious are the findings that exposure of half a minute is generally too little and that frequency of exposure is also relevant. As we have seen, the law takes little account of such results. Hauptmann was convicted of murdering the Lindbergh baby based largely on Lindbergh's being exposed to the speech of the perpetrator for perhaps two or three seconds. Even today, courts sometimes allow an identification based on a very brief exposure, including in one case the single word "yes."[164]

To date, there is not much research on the amount of exposure it takes to recognize familiar voices. We have all had the experience of making mistakes in recognizing familiar voices, especially on the telephone. Peter Ladefoged, an eminent phonetician who has studied the voice identification issue, and whom one would expect to be quite good at the task, has admitted that he could not even identify his own mother's voice saying "hello."[165] In fact, he also did not recognize her voice when the input was a full sentence.[166] A recent experiment by Australian researchers confirms Ladefoged's experience as typical. Based on the word "hello" alone, subjects were able to identify highly familiar voices a mere 47% to 60% of the time.[167] Increasing the length of the utterance to eight syllables resulted in 70% to 100% accuracy.[168] This area is one that is ripe for additional research.

---

160. *Id.* at 112.

161. *Id.*

162. *Id.* at 117.

163. *See id.*

164. United States v. Infelice, 506 F.2d 1358, 1365 (7th Cir. 1974).

165. Reported in Yarmey, *supra* note 126, at 797.

166. *Id.*

167. Phil Rose & Sally Duncan, *Naïve Auditory Identification and Discrimination of Similar Voices by Familiar Listeners*, 2 FORENSIC LINGUISTICS 1, 8 (1995).

168. *Id.* at 10.

(3) Delay

We all know that memory deteriorates over time, but research shows that it doesn't happen linearly. It seems that we remember voices quite well for some period of time, perhaps as long as a few weeks, and then our memories fade significantly.

A simple, but elegant, experiment was published in 1937 by Frances McGehee.[169] The experiment, inspired by the Hauptmann trial, aimed to determine how well people can identify unfamiliar voices after extended periods of time.[170] In the study, students listened to a person reading a fifty-six-word passage from behind a screen.[171] The students were asked at various subsequent times whether they recognized any of five voices presented to them at the testing session.[172] The results are presented in Table 2:

TABLE 2
Effect of Delay on Accuracy of Identification[173]

DELAY	CORRECT RESPONSES
1 DAY	83%
2 DAYS	83%
3 DAYS	81%
7 DAYS	81%
2 WEEKS	69%
3 WEEKS	51%
3 MONTHS	35%
5 MONTHS	13%

In a follow-up study conducted under somewhat different conditions, McGehee found that performance deteriorated to 48% after a two-week delay, but stayed more or less steady after that.[174]

---

169. McGehee, *supra* note 10. The results are reported in a useful survey article, Ray Bull & Brian R. Clifford, *Earwitness Voice Recognition Accuracy, in* EYEWITNESS TESTIMONY: PSYCHOLOGICAL PERSPECTIVES, *supra* note 13, at 92, 116. Bull and Clifford also report two of their own experiments. In one, they tested people after delays of 10, 40, 100 and 130 minutes. They found that people were significantly more accurate at ten minutes than at any of the longer times, but that the longer periods did not differ from each other significantly. *Id.* at 118. In the second study, they looked at delays of ten minutes, one day, one week and two weeks. Again, the only difference was between ten minutes and all the others. *Id.* at 118–19.

170. McGehee, *supra* note 10, at 249–50.

171. *Id.* at 255.

172. *Id.* at 256.

173. *Id.* at 262.

174. Frances McGehee, *An Experimental Study of Voice Recognition*, 31 J. GEN. PSYCHOL. 53 (1944).

Similarly, Defenbacher and his team found significant decay in recollection after two weeks, especially when the listener had only a single exposure to the voice.[175] They concluded that "[i]f the initial memory strength of the voice trace is weakly enough established, then, voice identification accuracy will not be very impressive even at delay intervals briefer than those possible in forensic situations."[176]

While the numbers differ somewhat from one study to another, perhaps depending on the amount and frequency of the initial exposure, the overall picture is fairly clear. In identifying unfamiliar voices, we perform much better if asked to do so immediately after hearing the voice. If there is a delay beyond that, our memories seem to remain fairly stable for a few weeks, after which performance drops off significantly. Moreover, at least after a rather brief initial period, the amount of exposure to the voice interacts with the length of the delay.

Once again, we have seen little indication that courts making evidentiary rulings take these findings into account, or that jurors do so in evaluating evidence that has been admitted. The twenty-nine-month delay in the Hauptmann case might still be acceptable today in some courts. This is not to say that courts do not consider the issue of time lapse; in fact, they almost always mention it. But they greatly underestimate the extent to which memory for voices decays over time. Consider the New York case in which a court allowed a policeman to identify a voice on tape after a time lapse of fifteen months,[177] or the *Knox* case, where the delay was around three years.[178] We are not claiming that it is impossible to remember a voice for that period of time, but we do believe that the legal system should take this cognitive frailty into account far more than it does.

(4) Individual Variation

Some people seem to be born musicians. They can hear a tune once and sing it exactly on key. Others are virtually tone deaf. Do we vary similarly in our ability to identify voices? The research is clear: Some people are quite good at identifying speakers from their voices, and other people are terrible at it.[179] This should not be surprising. We know from both personal experience and from experimental testing that people differ enormously in their abilities to

---

175. Defenbacher et al., *supra* note 158, at 116.
176. *Id.*
177. People v. Rendon, 709 N.Y.S.2d 698, 701 (N.Y. App. Div. 2000).
178. United States v. Knox, No. 97-5492, 1998 U.S. App. LEXIS 27655, at **9–10 (6th Cir. Oct. 22, 1998) (citation omitted).
179. See HARRY HOLLIEN, THE ACOUSTICS OF CRIME 197 (1990), for many citations to the literature.

recognize faces.[180] Why should voice recognition be different? The
legal system does not recognize such differences in skill. The rules of
evidence certainly do not, and we have never seen a published
opinion in which this issue was raised.

Experimenters have investigated the extent of individual
variation in identifying voices and have tried to determine whether it
is possible to predict ability to recognize voices from other cognitive
skills. In one recent study, Olaf Köster and some colleagues gave
thirty subjects (twenty-two non-expert subjects and eight expert
subjects) a test in auditory speech sensitivity.[181] The test required
subjects to imitate the relative pitch of two sounds, the voice onset of
a syllable (voicing is turned on earlier in a syllable beginning with d
than with one that begins with t), rhythm, nonsense syllables, and
other linguistically-relevant sounds.[182] The maximum possible score
on this test was 108.[183] Scores varied from a low of fifty-five to a
perfect score of 108.[184]

These same subjects were also given a test of their speaker
recognition ability.[185] First, they listened to a five-minute sample of a
male speaking German (the entire experiment was carried out in
Germany in German).[186] After a five-minute break, subjects were
presented with eighteen samples of speech from each of six different
male speakers with similar voices, for a total of 108 samples.[187] One of
the six was the target voice, to which they had just been exposed. The
only task was to indicate, for each of these 108 speech samples,
whether it was uttered by the target.[188] In a perfect performance, the
subject would identify all eighteen of the target's samples as the
target's ("hits"), and identify none of the other ninety samples as the
target's ("false alarms").[189]

The results showed great variation in ability. While eight
subjects got perfect scores, one subject made just about as many false
alarms as hits.[190] Importantly, there were statistical correlations
indicating that people who perform better on the auditory speech

180. *See* Hadyn D. Ellis, *Practical Aspects of Face Memory*, *in* EYEWITNESS
TESTIMONY: PSYCHOLOGICAL PERSPECTIVES, *supra* note 13, at 12, 36.

181. Olaf Köster, Markus M. Hess, Niels O. Schiller, & Hermann J. Künzel, *The
Correlation Between Auditory Speech Sensitivity and Speaker Recognition Ability*, 5
FORENSIC LINGUISTICS 22 (1998).

182. *Id.* at 23–24.
183. *Id.* at 25.
184. *Id.* at 27.
185. *Id.* at 25.
186. *Id.* at 25–26.
187. *Id.* at 26.
188. *Id.*
189. *Id.*
190. *Id.* at 27–28.

sensitivity test are likely to be better at speaker identification.[191] Yet much of the correlation comes from the fact that poor performers on the speech sensitivity test are typically not very good at speaker identification. This suggests that in future work it will be easier to predict poor identifiers than it will be to predict good identifiers. What is clear right now is that in some cases the legal system should permit defendants to inform the jury that some people are good at identifying speakers and some are bad at it, even under the optimal experimental conditions of this study.[192]

When it comes to admitting tape-recorded evidence, judges sometimes seem to assume that law enforcement officers will be particularly good at this task, but little evidence supports this assumption. Interestingly, experimental evidence suggests that police officers are no better at eyewitness identification than lay witnesses.[193] Thus, the law's main assumption regarding variation in voice identification abilities is at best unproven and quite possibly wrong.

(5) Emotional State and Tone of Voice

Many crimes requiring voice identification as part of their solution happen suddenly. This is especially true of violent crimes such as rape, burglary, and robbery. The victim or other witness did not see the perpetrator but is later asked if she can identify his voice. One question we might ask is whether the stress of these experiences heightens one's perceptiveness, making it easier to identify a voice later, or whether stress has the opposite effect. Research on this issue concerning eyewitness identification shows that stress makes us worse at identifying faces, despite our intuitions to the contrary.[194] Does the same hold true for the identification of people by their voices?

In an interesting study, Saslove and Yarmey had 120 experimental subjects engage in what they were told was an experiment on clairvoyance.[195] While an experimenter was conversing with a subject, an angry, hostile voice was heard from a tape recorder in the next room for about twelve seconds.[196] The experimental

---

191. *Id.* at 29–30.

192. Other researchers have reached similar conclusions. *See* Gea DeJong, Earwitness Characteristics and Speaker Identification Accuracy (1998) (unpublished Ph.D. dissertation, University of Florida) (on file with Lawrence Solan).

193. A. DANIEL YARMEY, THE PSYCHOLOGY OF EYEWITNESS TESTIMONY 159 (1979).

194. ELIZABETH F. LOFTUS, EYEWITNESS TESTIMONY 33–36 (1979); Ellis, *supra* note 180, at 20.

195. Howard Saslove & A. Daniel Yarmey, *Long-Term Auditory Memory: Speaker Identification*, 65 J. APPLIED PSYCHOL. 111 (1980).

196. *Id.* at 112.

subjects were subsequently[197] asked to pick the voice out of a voice lineup of five speakers.[198] All five speakers uttered the same words as the original angry voice.[199] For half of the subjects, the target voice used the same hostile tone.[200] For the other half, she used a calm voice.[201] In addition, half the subjects were told in advance that they would be asked to identify a voice, while the other half were uninformed.[202] Thus, there were thirty subjects in each of four conditions.

The results, summarized in Table 3, are dramatic:

Table 3
Number of Subjects Out of Thirty Correctly
Identifying Speaker[203]

	TONE OF TARGET VOICE	
INSTRUCTIONS	NO CHANGE (ANGRY)	CHANGE (CALM)
INFORMED	29	9
UNINFORMED	22	4

Since there were only five voices, one would expect six of the thirty subjects responding to this condition to identify the target voice even if everyone were guessing. In the hardest condition, where subjects were uninformed and where the target tone of voice was different from the original, subjects performed worse than chance. This study suggests that voice identification based on short exposure under stressful conditions is likely to be inaccurate, although that issue remains controversial.[204]

The Saslove and Yarmey experiment also suggests that certain voice qualities vary with the emotional state of the speaker. Research shows this to be the case. For instance, a voice's *fundamental frequency*, which relates largely to pitch, increases when we speak under stress.[205] Unfortunately, such changes are not always

---

197. By "subsequently" we mean either immediately or one day later. This lapse in time was not significant, so we have combined the results for purposes of our presentation. Below we will discuss the issue of the time it takes for memory of voices to decay.

198. Saslove & Yarmey, *supra* note 195, at 112.

199. *Id.*

200. *Id.* at 113.

201. *Id.*

202. *Id.* at 112–13.

203. *Id.* at 113.

204. For example, Hollien reports an experiment that showed that stress or arousal *improved* the accuracy of earwitness identifications. HOLLIEN, *supra* note 179, at 201–02.

205. HOLLIEN, *supra* note 179, at 258–63; Marianne Jessen, *Phonetic Manifestations of Cognitive and Physical Stress in Trained and Untrained Police Officers*, 4 FORENSIC LINGUISTICS 125–47 (1997); *see also* Hermann J. Künzel, *Effects of Voice Disguise on*

predictable, which means that to the extent that emotional states lead to changes in voice quality, they complicate the process of voice identification.[206] Note that in the Saslove and Yarmey experiment, the rate of correct identifications was relatively good when subjects were able to compare an angry target voice with other angry voices. In contrast, their identification levels were quite low when they were later asked to compare the originally angry voice with calm voices.

Because perpetrators of a crime are likely to be excited or angry, and the victims under stress, voice identification in these circumstances may be difficult. Yet that is precisely the condition, as Saslove and Yarmey state, that "might be considered most similar to the legal setting."[207] One case where the suspect's tone of voice made a difference was *State v. Johnson*, where a man was very calm and soft-spoken while raping a woman.[208] When later confronted with his voice through an open door at the police station, where he was speaking in an angry and abusive tone, she could not positively identify him.[209] When he calmed down and spoke more normally, however, she claimed to recognize his voice immediately.[210]

Although the emotional state and tone of voice of the speaker are important in predicting the reliability of an identification, courts do not take these factors into consideration as such. Part of the reason may be that the *Biggers* factors were formulated primarily to deal with eyewitness identification and therefore do not take into account some of the specific factors relating to voice identification. However, the *Biggers* criteria do include the degree of attention that the witness was paying, which relates to the emotional state of the witness. Courts seem to assume that stress increases the witness's attention and thus the reliability of the identification. The extent to which a witness was paying attention certainly is relevant, but the evidence so far indicates that stress in itself undermines reliability.

(6) The Problem of Disguise

Even more troublesome for voice identification are attempts to disguise one's voice or imitate the voice of someone else. The easiest way to disguise a voice is to whisper. Many of the acoustic features

---

Speaking Fundamental Frequency, 7 FORENSIC LINGUISTICS 149 (2000) (investigating more closely the phonetic events that occur during voice disguise); Dagmar Boss, *The Problem of F0 and Real-Life Speaker Identification: A Case Study*, 3 FORENSIC LINGUISTICS 155 (1996).

206. Gudrun Klasmeyer & Walter F. Sendlmeier, *The Classification of Different Phonation Types in Emotional and Neutral Speech*, 4 FORENSIC LINGUISTICS 104 (1997).

207. Saslove & Yarmey, *supra* note 195, at 115.

208. 351 A.2d 787, 789 (N.J. Super. Ct. App. Div. 1976).

209. *Id.*

210. *Id.*

that permit us to identify a speaker (like voicing) are absent when people whisper. Thus, the distinction between voiced consonants (like *z*) and voiceless ones (like *s*) largely disappears when a person whispers. As a result, the words *zap* and *sap* are difficult to distinguish when whispered.

Yarmey and his colleagues, in the same set of studies discussed earlier in connection with familiarity, had speakers record a speech sample in a whisper.[211] Recall that the experiments compared people's ability to identify voices based on their familiarity with the speaker. When the passage was whispered, highly familiar voices were identified correctly 77% of the time (versus 89% in a normal tone), moderately familiar voices 35% (versus 75%), voices with low familiarity 22% (versus 66%), and unfamiliar voices were acknowledged as such 20% (versus 61%).[212] False alarm rates were also significantly higher.[213] In short, a speaker who wishes to mask his voice by whispering has a good chance of succeeding—especially if he is not a very close friend or family member of the hearer, and even then he might succeed. Perhaps more disturbing is that independent panelists, when asked how often listeners were likely to be correct in identifying whispered voices, wildly overestimated their capacity to do so, guessing 91% for highly familiar voices, down to 74% for unfamiliar ones (versus actual success rates of 77% and 20%).[214] If jurors have similar misconceptions about this skill, it is not good news for defendants accused of having whispered an incriminating or illegal statement.

Studies have reached similar conclusions regarding other types of phonetic disguises. The Hollien group instructed speakers to mask their voices however they wished.[215] Experimental subjects were able to identify disguised familiar voices 79% of the time but could do no better than 20.7% with disguised unfamiliar voices.[216] Tactics used in some criminal contexts can lead to complex phonetic changes that make speech significantly more difficult to identify. For example, Brazilian kidnappers have been reported to place a pencil between their front teeth, under the tongue, to disguise their ransom demands. This leads to complex phonetic changes in speech that make the speaker significantly more difficult to identify.[217]

---

211. Yarmey et al., *supra* note 11, at 291.
212. *Id.* at 292.
213. *Id.*
214. *Id.* at 294.
215. Hollien et al., *supra* note 142, at 141.
216. *Id.* at 142.
217. Ricardo Molina de Figueiredo & Helena de Souza Britto, *A Report on the Acoustic Effects of One Type of Disguise,* 3 FORENSIC LINGUISTICS 168 (1996).

Imitation is an especially pernicious form of disguise. People who are good at imitating the voices of others have the power to cast suspicion on the innocent. How good are people at detecting imitators? A study conducted in Sweden examined how well people could identify in a voice lineup the voice of Carl Bildt, the former Prime Minister of Sweden, who was well known to subjects.[218] In one set of conditions, Bildt's voice was present in the lineup along with that of a good political impersonator, who imitated Bildt's voice.[219] Encouragingly, subjects almost always knew the real Bildt.[220] But when Bildt's voice was not among the choices, almost all subjects mistook the impersonator's voice for Bildt's.[221] These results suggest that a good imitation can fool people, especially when the actual voice is not present for comparison.

The emerging field of forensic phonetics is making progress in characterizing various ways in which people can mask their voices but still has not produced a systematic approach to the problem.[222] Researchers have begun to determine what features of a speaker's normal voice are likely to remain intact even when he tries to disguise it.[223] Yet disguise remains a problem, both for lay and expert identification of voices.

(7) Foreign Languages, Accents, and Other Linguistic Variables

Research has shown that eyewitnesses are generally better at identifying someone of the same race.[224] Are people similarly better at identifying speakers of their own language? In one experiment, Köster and Schiller investigated how well native speakers of Spanish and Chinese can identify a German speaker by his voice.[225] The experimenters contrasted subjects who knew German as a second language with those who knew no German at all.[226]

218. Frank Schlichting & Kirk P.H. Sullivan, *The Imitated Voice—A Problem for Voice Lineups?*, 4 FORENSIC LINGUISTICS 148, 149 (1997).

219. *Id.* at 151.

220. *Id.* at 155.

221. *Id.*

222. The International Association for Forensic Phonetics ("IAFP") has regular conferences to discuss advances in the field, and a code of conduct governing expert testimony by its members. *See* IAFP website, *at* http://www.iafp.net. The journal *Forensic Linguistics* publishes a great deal of the relevant literature. Most of this work originates in Europe.

223. *See, e.g.*, Geoff Lindsey & Allen Hirson, *Variable Robustness of Nonstandard /r/ in English: Evidence from Accent Disguise*, 6 FORENSIC LINGUISTICS 278 (1999); Herbert Masthoff, *A Report on a Voice Disguise Experiment*, 3 FORENSIC LINGUISTICS 160 (1996).

224. *See* YARMEY, *supra* note 193, at 130–31.

225. Olaf Köster & Niels O. Schiller, *Different Influences of the Native Language of a Listener on Speaker Recognition*, 4 FORENSIC LINGUISTICS 18 (1997).

226. *Id.* at 19.

Subjects were presented with a five-minute sample of a native speaker of German speaking in that language.[227] They were then asked to identify the speaker from a voice lineup consisting of six native speakers of German.[228] The results show that Spanish speakers who know some German performed significantly better than Spanish speakers who do not know German.[229] The same result held for the Chinese speakers.[230] These findings suggest that results of voice lineups involving speech samples in a language that the witness does not understand should, as the authors note, "be handled with caution."[231]

A study by Charles P. Thompson came to a similar conclusion, while providing some further details.[232] The voices were produced by Spanish-English bilingual speakers.[233] Some samples were in Spanish, others in English, and yet others in English with a Spanish accent.[234] Each sample was presented, along with five others of the same type, to monolingual English speakers who had been familiarized with the target voice one week before.[235] Results confirmed that English speakers are much worse at identifying someone speaking Spanish.[236] When confronted with speakers using English with a Spanish accent, the accented voices were recognized better than Spanish voices, but worse than English ones.[237]

A variation on this experiment was conducted by Goggin and her colleagues.[238] This research team employed a similar methodology but used listeners who were bilingual in English and Spanish.[239] As opposed to monolingual English speakers, the English-Spanish bilinguals did not differ significantly in their ability to identify targets speaking English, Spanish, or English with a Spanish accent.[240]

The fact that it is significantly more difficult to identify someone speaking an unknown language reveals that the term *voice identification* is actually a misnomer. If the task were simply to

---

227. *Id.* at 22.
228. *Id.*
229. *Id.* at 23.
230. *Id.*
231. *Id.* at 25.
232. Charles P. Thompson, *A Language Effect in Voice Identification*, 1 APPLIED COGNITIVE PSYCHOL. 121 (1987).
233. *Id.* at 122.
234. *Id.*
235. *Id.* at 123.
236. *Id.* at 124, 126.
237. *Id.* at 125.
238. J. P. Goggin, C. P. Thompson, G. Strube & L.R. Simenthal, *The Role of Language Familiarity in Voice Identification*, 19 MEMORY & COGNITION 448 (1991).
239. *Id.* at 453, 455.
240. *Id.* at 456.

identify a voice, it would logically make no difference at all whether we understand the language. In fact, it does matter. The reason is that the ultimate task is to identify the *speaker*. The quality of the speaker's voice may be an important clue in this endeavor, but it is not the only one. We also use other linguistic variables that depend on our ability to understand what is said and how it is said. Hollien's team, for example, lists a number of speech characteristics that listeners use to identify the speaker, including dialect, unusual use of linguistic stress or affect, idiosyncratic language patterns, speech impediments, and idiosyncratic pronunciations.[241]

We have the most nonacoustic (or "non-voice") information at our disposal when we hear someone speak our native language, less when the speaker is not a native speaker of our language, and none at all when the language we hear strikes us as merely a babble of meaningless sounds. What if subjects are confronted with an actual babble of meaningless sounds? Niels Schiller and colleagues presented what was essentially a long series of *mamamamama* to native English and German speakers, along with some English speakers who had studied German.[242] Correct identification of the speaker was low for all three groups, especially when compared to the success rate of a native speaker of a language who is asked to identify speech produced by other native speakers of that language.[243]

The effect of general speech characteristics on voice identification has not been extensively studied.[244] Fortunately, some courts seem to have an intuitive notion that foreign accents may present a problem and have sometimes required expert testimony if a witness is to identify a voice by its accent.[245] Others, as we saw above, seem to consider it largely irrelevant that a speaker was speaking Spanish on one occasion and English on another.[246]

---

241. HOLLIEN, *supra* note 179, at 198.

242. Niels O. Schiller, Olaf Köster & Martin Duckworth, *The Effect of Removing Linguistic Information Upon Identifying Speakers of a Foreign Language*, 4 FORENSIC LINGUISTICS 1 (1997).

243. *Id.* at 10, 11.

244. For a study on how foreign accent can help identify a speaker—or eliminate a suspect—see Henry Rogers, *Foreign Accent in Voice Discrimination: A Case Study*, 5 FORENSIC LINGUISTICS 203 (1998). On the difficulty that many Swedish speakers have in successfully imitating regional accents, see Duncan Markham, *Listeners and Disguised Voices: The Imitation and Perception of Dialectal Accent*, 6 FORENSIC LINGUISTICS 289 (1999).

245. People v. King, 584 N.Y.S.2d 153 (N.Y. App. Div. 1992) (not permitting lay witness to testify about whether defendant speaks with a Jamaican accent). *But see* People v. Sanchez, 492 N.Y.S.2d 683 (N.Y. Sup. Ct. 1985) (permitting lay witness to testify about perpetrator's accent, but acknowledging that expert linguistic testimony might sometimes be necessary).

246. United States v. Vega, 860 F.2d 779, 790 (7th Cir. 1988).

(8) Other Factors

This discussion of factors bearing on our ability to recognize voices is by no means exhaustive. For example, age plays a role in the reliability of voice identification. Speakers between twenty-one and forty are better voice identifiers than are adults over forty.[247] Moreover, people's voices change as they get older.[248] In fact, that issue arose in a case decided by the United States Court of Appeals for the First Circuit.[249] In that case, the defendant's voice was recorded in 1971, but it wasn't until 1975 that an agent met the defendant and compared the voice on the tape to the voice he had heard.[250] The defendant argued that this four-year delay made the identification improper. The court disagreed, applied Rule 901, and stated that the delay should go to the weight that the jury gives to the evidence.[251]

Nonetheless, we have attempted to touch on the most legally-relevant research on how good people are at identifying voices. The legal system relies almost entirely on its own notions of common sense and intuition and has never systematically taken this knowledge into account. Moreover, to the extent that the law purports to require some investigation into the reliability of an identification, the research reported here suggests that the *Biggers* criteria fall seriously short of the mark.

B. Witness Confidence

Among the criteria for predicting reliability that the Supreme Court endorsed in *Biggers* is "the level of certainty demonstrated by the witness at the confrontation."[252] Research shows that jurors, like judges, take statements of confidence seriously. Unfortunately, research also indicates that there is at best a limited relationship between the probability of accuracy and the degree of confidence that the witness has in the identification. If people react positively to the confidence of the identifier and confidence fails to predict accuracy,

---

247. Ray Bull & Brian R. Clifford, *Earwitness Voice Recognition Accuracy, in* EYEWITNESS TESTIMONY: PSYCHOLOGICAL PERSPECTIVES, *supra* note 13, at 92, 95–97. Similarly, whether because of attention and motivation, or because of cognitive development, fifteen- to sixteen-year-olds perform better than twelve-year-olds at identifying voices. *Id.*

248. *Id.* See discussion of how the relative ages of the speaker and listener might affect identification, *infra* note 251.

249. United States v. DiMuro, 540 F.2d 503, 513 (1st Cir. 1976).

250. *Id.*

251. Recent research suggests that the court was right in this decision. *See* Harry Hollien & Reva Schwartz, *Aural-Perceptual Speaker Identification: Problems with Noncontemporary Samples*, 7 FORENSIC LINGUISTICS 199 (2000).

252. Neil v. Biggers, 409 U.S. 188, 199 (1972).

then we might expect people to overestimate the likelihood that an identification will be accurate. That is just what seems to happen.

Several researchers have studied the relationship between accuracy and confidence in connection with speaker identification.[253] For the most part, the research indicates little positive correlation. Defenbacher et al., who set out to study the significance of the factors suggested in *Biggers*, conclude from their studies that the Supreme Court was probably wrong. "The safest generalization to make is that earwitness as well as eyewitness confidence are not very reliable indices of identification accuracy."[254] Yarmey's review of the literature led him to reach the same bottom line.[255]

Moreover, jurors are likely to be swayed by confidence levels. A recent study by Amy Bradfield and Gary Wells shows that people pay a great deal of attention to how confident a witness is in his identification in deciding how much weight to give it.[256] This bias can lead to insufficient skepticism on the part of jurors whose job it is to assess the reliability of a witness's identification.

Finally, people seem to have an inflated sense of how good we are, as human beings, at identifying voices. Recall the study by the Yarmey team demonstrating that people are much more successful at identifying highly familiar voices than they are at identifying voices of moderate or low familiarity, or identifying unfamiliar voices.[257] In a related study discussed in the same article, another set of experimental subjects was asked to estimate how good listeners would be at identifying voices from each of the four levels of familiarity.[258] For every level of familiarity, people assume that identifiers will be more accurate than is the case.[259] This gap between perception and reality suggests that jurors may be predisposed to give too much weight to identification by voice.

These results have serious ramifications. Prosecutors must prove their cases beyond a reasonable doubt. Although jury instructions do

---

253. This issue has also been widely studied in the context of eyewitness identification. *See, e.g.*, CUTLER & PENROD, *supra* note 13, at 94–95.

254. Defenbacher et al., *supra* note 158, at 115.

255. Yarmey, *supra* note 126, at 803 ("The above research has consistently failed to reach .50 correlation coefficients, which reinforces the conclusion that earwitness confidence is an unreliable criterion to judge speaker identification accuracy.").

256. Amy J. Bradfield & Gary L. Wells, *The Perceived Validity of Eyewitness Identification Testimony: A Test of the Five* Biggers *Criteria*, 24 LAW & HUM. BEHAV. 581 (2000).

257. Yarmey et al., *supra* note 11.

258. *Id.* at 287–88.

259. For highly familiar voices, subjects predicted 97% accuracy, while identifiers were 89% accurate. For moderately familiar voices, subjects predicted 89% accuracy, while identifiers were 75% accurate. For low familiarity, the predicted and actual numbers were 90% and 66%, respectively, and for unfamiliar voices 85% and 61%.

not use numerical certainty thresholds, most people within the system, when asked, say that proof beyond a reasonable doubt requires about a 90% level of certainty.[260] Based on this experimental data, people appear to assume that under most circumstances voice identification is correct about 90% of the time; but in reality, it is significantly less reliable, especially when we are not very familiar with the voice being identified. The legal system's failure to correct this overestimation may result in some jurors wrongly concluding that the government has met its burden of proof.

Taken together, these facts tell a disturbing story. People rely on an identifier's level of confidence in judging how accurate the identification is likely to be. But that level of confidence correlates only slightly with the likelihood of accuracy. The result is that people tend to place too much credence in an identification. Again, this situation cries out for judicial safeguards. We present some possible solutions to these problems at the end of this Article.

III. Expert Voice Identification

When the issue is the admission of a tape-recorded voice, someone must determine whether the defendant's voice matches the voice on a tape. We have seen that in the typical case a witness, usually a police officer, is called to identify the voice as that of the defendant. The question we ask here is whether training in voice identification, or phonetics in general, is helpful in this task. If so, it calls into question the statement in the Advisory Committee notes to Rule 901 that "aural voice identification is not a subject of expert testimony."[261]

We then turn to the question of sound spectrography, or voiceprints, a largely mechanical method that is claimed to be able to distinguish voices. If such a method is reliable, it could prove quite useful, not only in laying a foundation to admit tape recorded evidence, but also to help prove the ultimate issue of the identity of the speaker. As we will see, voiceprints are more reliable than is sometimes suggested in the literature, but questions still remain, especially when recordings are made in settings that are less than ideal.

A final role that experts might perform is to inform courts and juries about the ability of people to make lay voice identifications. For example, it might have been helpful to Hauptmann's defense to have had an expert testify about some of the problems with

---

260. For a summary of this work, see Lawrence M. Solan, *Refocusing the Burden of Proof in Criminal Cases: Some Doubt About Reasonable Doubt*, 78 TEX. L. REV. 105, 125–29 (1999).

261. FED. R. EVID. 901 advisory committee's note, subd. (b), ex. 5.

Lindbergh's identification of Hauptmann's voice. Whether conveyed by an "educational expert" or by jury instructions, the type of learning discussed in this Article should make its way to the jury when a case rests in part on a questionable identification of a speaker.

A. Aural Identification by Experts

Consider a case in which the issue is whether the voice on a tape recording is the defendant's voice. The question arises in many drug and organized crime cases. As we have seen, some people are better at making this comparison than are others.[262] Here, we ask whether linguistic experts specializing in phonetics are typically better than lay people at aural identification (i.e., identification by ear as opposed to using machines). The answer seems to be yes.

Hollien and Schwartz tested people's abilities to identify voices by comparing contemporary samples with non-contemporary ones.[263] All samples were on tape.[264] There were three groups of subjects: people with no background in phonetics, experienced phoneticians, and students with some background in phonetics.[265] The results are presented below:

TABLE 4

Effect of Expertise on Voice Identification Skill[266]

LENGTH OF DELAY	LAY SUBJECTS	PHONETICIANS	STUDENTS
4 WEEKS	74%	89%	76%
20 YEARS	32%	74%	33%

From these results, it appears that training in phonetics increases performance on identification tasks.

In another study, Schiller and Köster had six male native speakers of German record a passage that lasted about one minute.[267] For each speaker, three pieces were spliced out of the original tape, and each of those three pieces was recorded six separate times, making a total of 108 speech samples (6 x 3 x 6 = 108).[268] Twenty-seven subjects participated in the experiment.[269] Seventeen were

---

262. *See* Bull & Clifford, *supra* note 247, at 94–105.
263. Hollien & Schwartz, *supra* note 251.
264. *Id.* at 203.
265. *Id.* at 204.
266. *Id.* at 206.
267. Niels O. Schiller & Olaf Köster, *The Ability of Expert Witnesses to Identify Voices: A Comparison Between Trained and Untrained Listeners*, 5 FORENSIC LINGUISTICS 1 (1998).
268. *Id.*
269. *Id.* at 2.

college students, and ten were experts in phonetics.[270] For each subject, one of the original six speakers was designated as the target.[271] The subject listened five times to the target reading the entire passage.[272] Then, after a five minute break, subjects were presented with all 108 short segments, and instructed to indicate whether the segment was uttered by the target.[273]

Both expert and lay subjects did very well on this test, which was designed around ideal conditions.[274] Still, the experts performed significantly better (98% hits, 1% false alarms) than the lay subjects (92% hits, 2% false alarms).[275] The differences may not look dramatic, but the reduced error rate would certainly be important in a trial setting.[276]

A second study by Köster et al. resulted in a similar finding.[277] Experts in speech or singing and non-experts participated in a voice identification task.[278] While half of the non-experts (eleven of twenty-two) were poor identifiers, only one quarter of the experts (two of eight) were poor identifiers.[279] Although the number of subjects is low, the study suggests that the use of experts may improve the accuracy of voice identification under certain forensic circumstances.[280] Significantly, these studies highlight the fact that phonetics is an independent field of scientific research, which takes seriously the need to investigate its own strengths and limitations. This internal scrutiny distinguishes it from other areas of forensic identification, such as handwriting analysis and microscopic hair

---

270. *Id.* at 3.

271. *Id.*

272. *Id.*

273. *Id.* at 4.

274. *Id.*

275. *Id.*

276. Here, as in any forensic identification testimony, one must take into account that the experimental studies, conducted under controlled conditions, are relatively free of "observer biases." In the actual setting, both experts and lay witness understand the consequences of their answers to the possible outcome of the case. For further discussion, see D. Michael Risinger Michael J. Saks, William C. Thompson & Robert Rosenthal, *The* Daubert/Kumho *Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion*, 90 CAL. L. REV. 1 (2002).

277. *See* Köster et al., *supra* note 181. As discussed above, the study showed correlation between voice identification ability and performance on a speech sensitivity test.

278. *Id.* at 27.

279. *Id.* at 27–29.

280. *See* Ricci v. Urso, 974 F.2d 5, 7 (1st Cir. 1992) (holding that detective did not have expert training in voice identification).

analysis, which have received substantial criticism in the legal literature.[281]

These results do not mean that the system should *require* expert identification of voices on tapes. But they do suggest that the courts should be receptive to such experts in cases where voice identification is critical, especially when the admission of the tape into evidence is based on little exposure to the voice.[282] Expert phoneticians may also be appropriate when a police officer or other witness becomes familiar with a voice specifically in order to become eligible as an authenticating witness.[283]

Consider *United States v. Drones*, in which the Fifth Circuit recently reversed the district court's grant of the defendant's petition for *habeas corpus* based on ineffective assistance of counsel.[284] The case against Drones, who was convicted of various drug crimes in a Texas state court, depended heavily on a police officer's identification of his voice on a tape.[285] His attorney, however, did nothing to challenge the identification.[286] Later, after new counsel was retained, witnesses who knew Drones listened to the tape, and concluded that it was not his voice.[287] In addition, a voice identification expert opined from both aural and spectrographic analysis that the voice on the tape was not that of Drones.[288] The government also had an expert who debunked spectrographic analysis but agreed that the voice on the tape did not sound like that of the defendant.[289]

The Fifth Circuit held that the original lawyer did not act irresponsibly in not pursuing the lay testimony, which may have opened up other questions about the defendant's background.[290] It also held that it was not irresponsible of the original lawyer not to pursue spectrographic comparison of Drones's voice with that on the tape, since courts have not been receptive to such analysis in recent years.[291] But the court never said why the defendant's lawyer should not have had an expert study both the tape and exemplars of the defendant's voice and offer his opinion based on aural analysis.[292] At

---

281. *See, e.g.*, Michael J. Saks, *Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science*, 49 HASTINGS L.J. 1069 (1998).

282. *See supra* Part I.B.

283. *See, e.g.*, United States v. Alvarez, 860 F.2d 801, 807–09 (7th Cir. 1988).

284. 218 F.3d 496, 505 (5th Cir. 2000).

285. *Id.* at 498.

286. *Id.*

287. *Id.* at 499.

288. *Id.* at 498.

289. *Id.* at 499.

290. *Id.* at 501.

291. *Id.* at 504. We discuss, and agree with this aspect of the holding, *infra* note 356.

292. See *infra* note 346 for a discussion of how current doctrine concerning the admissibility of expert testimony impacts on voice identification expertise.

the very least, an expert could have pointed out to the jury ways in which particular sounds differed from one sample to the other and left it to them to agree or disagree. The case is especially compelling because the government's own expert shared the defense expert's opinion[293] and because the standards for admitting the original identification are so relaxed.

In fact, courts have permitted expert phoneticians to present opinion testimony on speaker identification when the voice is on tape.[294] Their expertise is parallel to that of other experts who are permitted to assist the jury on questions of identification. For example, experts are permitted to interpret surveillance photos to point out similarities between the facial features of the defendant and the individual in the surveillance photo.[295] In one case, it was held to be reversible error not to permit the defendant to offer such testimony to the jury.[296] While these experts are not always permitted to offer opinions as to identity, they are routinely allowed to share with the jury detailed observations regarding facial shape and measurement.

At the very least, phoneticians should be permitted to point out similarities and differences between the defendant's voice and that of the person on a tape in order to make them salient to the jury. To the extent that such analysis can be enhanced by comparison of the relevant features of spectrograms, we see no reason why the experts should not be permitted to use that information as well. For example, to enhance her testimony based on aural comparison of two voices, a phonetician may want to show a jury how one speaker's [a] sound routinely appears in one area of the spectrogram, while the [a] sound on the tape that is in evidence appears elsewhere. Most phoneticians use both types of information.[297] This use of acoustic information is quite different from that used by so-called voiceprint experts, whose claims have been a matter of controversy for several decades, an issue to which we now turn.

B. Spectrographic Evidence

DNA evidence has become an important forensic tool, both for law enforcement agencies and for those who have been wrongly

293. *Drones*, 218 F.3d at 499.
294. See *infra* note 388 describing Professor Hollien's testimony.
295. *See, e.g.*, United States v. Everett, 825 F.2d 658, 662 (2d Cir. 1987); United States v. Barrett, 703 F.2d 1076, 1084 n.14 (9th Cir. 1983); *see also* 4 WEINSTEIN & BERGER, *supra* note 84, § 702.02[2].
296. United States v. Alexander, 816 F.2d 164, 168–69 (5th Cir. 1987).
297. *See* HOLLIEN, *supra* note 179; Francis Nolan, Speaker Identification Evidence: Its Forms, Limitations, and Roles (2001) (unpublished manuscript, on file with *Hastings Law Journal*).

accused. A technology able to compare voices with such accuracy would obviously be a welcome addition to the criminal justice system. For a time, at least, there was hope that "voiceprints" could play that role. Voiceprints, or technically "sound spectrograms," are visual representations of the frequencies and amplitude of sounds as represented on a time line. In the forensic setting, spectrographic analysis involves visual comparison of the spectrogram of the questioned voice with one from a known voice, typically the voice of the defendant in a criminal trial. Most of those who conduct this kind of analysis are not phoneticians, but rather police officers and technicians who have been trained for this specific task and typically have limited backgrounds in acoustics or phonetics. The main issue is whether the methodology produces sufficiently reliable results.

The early history of this debate is both legally significant and interesting in its own right. Sound spectrography was developed in the 1940s by Bell Laboratories for teaching deaf people how to speak[298] and was quickly pursued for use in military operations during World War II.[299] Then, in 1962 Lawrence G. Kersta, of Bell Labs, published an article in *Nature*, making some extravagant claims about the ability to identify speakers by their voiceprints.[300] He likened this technology to fingerprints and asserted that people's voices are also unique and identifiable through visual inspection of their voiceprints.[301] The scientific community reacted skeptically. For example, the Committee on Speech Communication of the Acoustical Society of America had the following reaction:

> We conclude that the available results are inadequate to establish the reliability of voice identification by spectrograms.... Procedures exist, as we have suggested, by which the reliability of voice identification methods can be evaluated. We believe that such validation is urgently needed.[302]

---

298. *See* HOLLIEN, *supra* note 179, at 90–91.

299. *See* Oscar Tosi, Herbert Oyer, William Lashbrook, Charles Pedrey, Julie Nicol & Ernest Nash, *Experiment on Voice Identification*, 51 J. ACOUSTICAL SOC'Y AM. 2030, 2031 (1972).

300. L. G. Kersta, *Voiceprint Identification*, 196 NATURE 1253 (1962). Toward the conclusion, Kersta writes: "It is my opinion, however, that identifiable uniqueness does exist in each voice, and that masking, disguising, or distorting the voice will not defeat identification if the speech is intelligible." *Id.* at 1257. For a discussion of Kersta's scholarship, see PAUL C. GIANNELLI & EDWARD J. IMWINKELREID, SCIENTIFIC EVIDENCE § 10-2, at 299 (2d ed. 1993).

301. Kersta, *supra* note 300.

302. Richard H. Bolt, Franklin S. Cooper, Edward E. David, Jr., Peter B. Denes, James M. Pickett & Kenneth N. Stevensal, *Speaker Identification by Speech Spectrograms: A Scientist's View of its Reliability for Legal Purposes*, 47 J. ACOUSTICAL SOC'Y AM., 597, 603 (1970).

Prominent phoneticians, including Dr. Peter Ladefoged of UCLA, went on record as opposing the use of spectrography in the courtroom as inadequately tested.[303]

During the late 1960s and early 1970s, American jurisdictions were divided on the issue of spectrographic evidence. Some courts rejected the methodology as not widely enough accepted within the scientific community.[304] Typically, these jurisdictions applied the *Frye* test, which stated that for expert opinion testimony to be admitted, "the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."[305] Other appellate courts reached the opposite conclusion, holding that it was within the discretion of the trial judge to admit the testimony because the technology showed indications of being sufficiently reliable.[306]

At this same time, Oscar Tosi, a Michigan State University professor, published studies that purportedly demonstrated great accuracy in the identification of speakers by visual inspection of voiceprints.[307] Tosi's design corrected the worst problems with Kersta's study. While Kersta used a closed set of possible voices and contemporaneous recording under laboratory conditions, Tosi's study compared non-contemporaneous recordings in both fixed and random contexts, which more closely simulated forensic settings.[308] Tosi's results were impressive. There were approximately 6% false alarm errors and approximately 13% false elimination errors.[309] He conjectured that the error rate could be reduced further if the examiner is permitted to answer "I don't know" if he did not reach a high level of certainty.[310]

The Tosi study led to modification of Professor Ladefoged's opposition to the use of voiceprints, but his softening was hardly a ringing endorsement. Ladefoged expressed his position in a 1971

---

303. *See* Peter Ladefoged & R. Vanderslice, *The Voiceprint Mystique, in* WORKING PAPERS IN PHONETICS No. 7, at 126 (UCLA Dept. of Linguistics Nov. 1967).

304. *See, e.g.*, State v. Cary, 239 A.2d 680 (N.J. Super. Ct. Law Div. 1968).

305. Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

306. *See, e.g.*, State *ex rel.* Trimble v. Hedman, 192 N.W.2d 432, 440–41 (Minn. 1971). For a catalogue of these early cases, see DAVID L. FAIGMAN, DAVID H. KAY, MICHAEL J. SAKS & JOSEPH SANDERS, MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY § 25-1.0, at 190 (1997). *See also* Lisa Rafferty, Note, *Anything You Say Can and Will be Used Against You: Spectrographic Evidence in Criminal Cases*, 36 AM. CRIM. L. REV. 291 (1999).

307. *See* Tosi et al., *supra* note 299. This was not Tosi's first study, but it was the seminal one.

308. *Id.* at 2033.

309. *Id.* at 2041.

310. *Id.*

letter to the President's Science Advisor.[311] In the letter, he expressed concern that the Tosi study did not deal with the problem of people in the same community, say a gang of high school dropouts, who might have very similar speaking styles and mutually confusable voices.[312] He also expressed concern about the lack of standards governing voiceprint experts and their work.[313]   The United States Court of Appeals for the D.C. Circuit summarized his position:

> As Dr. Ladefoged's cited letter to the President's Science Advisor itself indicates, however, his conversion to the voiceprint identification technique has been a limited one.   While Dr. Ladefoged stated that new evidence had moved him to 'cautiously reconsider the possibility' of the use of spectrogram analysis in criminal trials, he went on to express a number of continuing reservations. He pointed out, for instance, that the Tosi studies did not necessarily indicate that spectrogram analysis would enjoy a comparable success rate when applied to the general populace, and indicated that voiceprint identification of females would probably be more difficult than identification of males.   Dr. Ladefoged further identified problems arising from voice mimicry and from the possibility of "confusable voices," and concluded that "we do not at the moment know the probable error rate" of a spectrogram analysis technique applied to the broad populace. Thus, viewed in its entirety, Dr. Ladefoged's letter, as he himself characterized it in his testimony, simply reflects a position "of abatement of skepticism towards voiceprint," not one of complete acceptance. This position, according to his testimony, was shared by the community of scientists.[314]

---

311. The letter was published as Peter Ladefoged, *An Opinion on Voiceprints, in* WORKING PAPERS IN PHONETICS No. 19, at 84 (UCLA Dept. of Linguistics June 1971).

312. *Id.* at 85.

313. *Id.* at 87.

314. United States v. Addison, 498 F.2d 741, 744–45 (D.C. Cir. 1974) (citations omitted) (rejecting voiceprint testimony under the *Frye* standard).   Subsequently, Dr. Ladefoged actually testified in favor of the use of spectrograms in the courts.   But his testimony remained guarded:
> Q. Is it then, sir, your testimony that at this time you feel that the use of the voice spectrograph in the identification of or comparison of known voices with unknown voices is a scientifically acceptable method for identification?
> A. Given certain safeguards so that it's performed by a process or that the recordings that the investigator has truely [sic] identified and of no noises or background, undue degree of background noise, given also the fact that there is no claim that these are women's voices of high pitch which make it difficult, also that there is no claim that this is somebody mimicing [sic] somebody else's voice, the defendants, that is somebody mimicing and you can say there is a skilled mimic who is able to mimic very well, then I think that's most of the circumstances, yes, and, oh, and also given the fact that the court is advised of the fact that this is not 100 percent, yes, I would agree that this is a technique that is useable [sic] in a court of law.

People v. Law, 114 Cal. Rptr. 708, 716 n.12 (Cal. Ct. App. 1974) (alterations in original).

In an introductory text to phonetics dating from that same period, Ladefoged explained that it was his "best guess" that experts using spectrograms were wrong in about one case out of twenty, which means that it is a useful—but limited—law enforcement tool.[315] Ladefoged went on to characterize as "completely irresponsible" the assertions of witnesses in court that "[t]he voice on the recording is that of the accused and could be that of no other speaker."[316]

While the scholarly community gave the Tosi study mixed reviews,[317] it was good enough to convince some courts that voiceprint analysis was sufficiently valid for courtroom purposes. The federal and state law reports contain a number of cases in which voiceprint analysts were permitted to testify over the objection of the opposing party, typically for the prosecution in criminal trials.[318] Other courts continued to reject spectrographic evidence. The standard for admissibility under *Frye* was acceptance in the scientific community, and the debate under *Frye* was often, "whose community?"[319] In Tosi's community of supporters, voiceprint analysis was widely accepted. In the broader community of acoustic phoneticians, it was not. This difference explains, at least in part, the divergent court rulings.

Then, in 1975, the Federal Rules of Evidence were adopted. Its standard for admissibility of expert testimony, that the expert "will assist the trier of fact,"[320] would seem to leave more opportunity for a court to admit spectrographic analysis through experts. In 1979, however, an influential report by the National Research Council, usually called the "Bolt Report," questioned the ability of voiceprints to produce accurate results under forensic conditions with sufficiently

---

315. PETER LADEFOGED, A COURSE IN PHONETICS 194 (2d ed. 1982).

316. *Id.*

317. For example, a 1973 letter to the editor of *The Journal of the Acoustical Society of America* by prominent critics concluded:

As scientists rather than lawyers, we offer no judgment as to whether or to what extent speech spectrograms should be used for identification in the courts. We wish only to point out that present methods for such use lack an adequate scientific basis for estimating reliability in many practical situations and that laboratory evaluations of these methods show increasing errors as the conditions for evaluation move toward real-life situations. We hope that our explanations of some of the factors that affect speaker identification will provide the legal profession with helpful information on which to base its own judgments concerning the admissibility of the spectrographic method.

Richard H. Bolt, Franklin S. Cooper, Edward E. David, Jr. Peter B. Denes, James M. Pickett & Kenneth N. Stevens, *Speaker Identification by Speech Spectrograms: Some Further Observations*, 54 J. ACOUSTICAL SOC'Y AM. 531, 534 (1973).

318. *See, e.g.*, Commonwealth v. Lykus, 327 N.E.2d 671 (Mass. 1975). See FAIGMAN ET AL., *supra* note 306, and Rafferty, *supra* note 306, for lists of cases.

319. Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

320. FED. R. EVID. 702.

low rates of error.[321]   The report summarized its findings in the introduction:

> The Committee concludes that the technical uncertainties concerning the present practice of voice identification are so great as to require that forensic applications be approached with great caution. The Committee takes no position for or against the forensic use of the aural-visual method of voice identification, but recommends that if it is used in testimony, then the limitations of the method should be clearly and thoroughly explained to the fact finder, whether judge or jury.[322]

 The Committee later explained:

> The degree of accuracy, and the corresponding error rates, of aural-visual voice identification vary widely from case to case, depending upon several conditions including the properties of the voices involved, the conditions under which the voice samples were made, the characteristics of the equipment used, the skill of the examiner making the judgments, and the examiner's knowledge about the case. Estimates of error rates now available pertain to only a few of the many combinations of conditions in real-life situations. These estimates do not constitute a generally adequate basis for a judicial or legislative body to use in making judgments concerning the reliability and acceptability of aural-visual voice identification in forensic applications.[323]

It is important to note that the Committee did not dispute Dr. Tosi's findings. In fact, Tosi was on the Committee.[324] Rather, the report complained that findings supporting the use of voiceprints were too limited.[325] They failed to consider important real-life variables that would be necessary to draw valid conclusions about forensic use of voiceprints.[326] We make this point not to advocate for the acceptance of voiceprint analysts in the courts (we do not), but to point out that the scientific community has generally been straightforward about the abilities and limitations of voiceprint analysis.

Subsequently, Dr. Ladefoged reached similar conclusions. A Hawaii court quoted him as making the following points in 1985:

> Dr. Ladefoged proposes the following safeguards: (1) two plus minutes of each speech sample; (2) a signal to noise ratio where the signal is higher by 20 decibels; (3) a frequency of 3,000 hertz or better; (4) an exemplar in the same words, the same rate, in the same way, spoken naturally and fluently; and (5) a responsible

---

321. COMMITTEE ON EVALUATION OF SOUND SPECTROGRAMS, *supra* note 17. The chair of this committee was Richard H. Bolt, thus giving the report its name.

322. *Id.* at 2.

323. *Id.* at 60.

324. *Id.* at iii.

325. *Id.* at 63–64.

326. *Id.* at 60.

examiner. Dr. Ladefoged believes there is general acceptance given his safeguards, and he believes there is now more agreement.[327]

Rarely will all these safeguards be met, making visual voiceprint analysis of limited evidentiary value. For this reason, some linguists continue to express serious doubts about the reliability of this technology in a forensic setting. Indeed, one phonetician has called it "a fraud being perpetrated upon the American public and the Courts of the United States."[328]

Surprisingly enough, throughout the 1980s and early 1990s the published opinions, albeit in smaller numbers overall than before, continued to be split on the issue. Decisions to admit voiceprint evidence were reached during that period by the United States Courts of Appeals for the Sixth and Seventh Circuits, a federal district court in Hawaii, the supreme courts of Ohio, Maine, and Rhode Island, and a lower court in New York.[329] But during roughly the same time, voiceprints were held inadmissible by the high courts of Arizona, Colorado, Indiana, Louisiana, and New Jersey.[330] Clearly, the courts are seriously divided.

This disagreement has not abated, despite significant legal developments over the past decade. In 1993, the Supreme Court clarified the standard for admissibility of expert testimony under Rule 702 of the Federal Rules of Evidence when it decided *Daubert v. Merrell Dow Pharmaceuticals, Inc.*[331] The issue in *Daubert*, a products liability case, was whether Bendectin, an anti-nausea drug taken during pregnancy, had caused the plaintiff's birth defects.[332] The epidemiological literature suggested that it did not.[333] The plaintiff in *Daubert* wanted to call experts who would attack the inferences drawn from the data in the published literature and bring to bear the results of animal studies.[334] The trial court had rejected the experts on the grounds that their work had not been published and therefore failed to meet the standards of scientific reliability that the courts had developed under *Frye*.[335] It thus granted summary judgment to the

---

327. United States v. Maivia, 728 F. Supp. 1471, 1477 (D. Haw. 1990) (citation omitted).
328. HOLLIEN, *supra* note 179, at 210. For Hollien's more recent assessment, also negative, see HARRY HOLLIEN, FORENSIC VOICE IDENTIFICATION (2001); FRANCIS NOLAN, THE PHONETIC BASES OF SPEAKER RECOGNITION 25 (1983).
329. FAIGMAN ET AL., *supra* note 306; Rafferty, *supra* note 306.
330. FAIGMAN ET AL., *supra* note 306; Rafferty, *supra* note 306.
331. 509 U.S. 579 (1993).
332. *Id.* at 582.
333. *Id.*
334. *Id.* at 583.
335. *Id.*

defendant, Merrell Dow.[336]   The court of appeals affirmed the trial court's decision.[337]

The Supreme Court reversed, holding that the Federal Rules of Evidence had replaced the *Frye* standard.[338]   It interpreted Rule 702 as requiring courts to engage in a "preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue."[339]   To be "scientifically valid" the proffered evidence need not be uncontroversially accepted in the scientific community.[340]   Rather, "[t]he adjective 'scientific' implies a grounding in the methods and procedures of science."[341]

The Court did not attempt to state conditions that are both necessary and sufficient for evidence to be "scientifically valid."[342]   It did suggest, however, four non-exclusive indicia: whether the theory offered has been tested; whether it has been subjected to peer review and publication; the known rate of error; and whether the theory is generally accepted in the scientific community.[343]   In a subsequent case, *Kumho Tire Co. v. Carmichael*, the Supreme Court made it clear that the *Daubert* criteria are to be applied to experts purporting to testify based on their experience, as well as to experts purporting to rely on scientific advancements.[344]   Since then, Rule 702 has been amended to incorporate these holdings.[345]   Undeniably, these principles govern the admissibility of voiceprint testimony.[346]

No federal courts have ruled on the admissibility of voiceprints since *Daubert*. However, one state court, the Supreme Court of Alaska, ruled voiceprint testimony admissible under *Daubert* in 1999, while the United States Court of Appeals for the Fifth Circuit

---

336. Daubert v. Merrell Dow Pharmaceuticals, Inc., 727 F. Supp. 570, 576 (S.D. Cal. 1989).

337. Daubert v. Merrell Dow Pharmaceuticals, Inc., 951 F.2d 1128 (9th Cir. 1991).

338. *Daubert*, 509 U.S. at 587.

339. *Id.* at 592–93.

340. *Id.* at 590.

341. *Id.*

342. *Id.* at 594.

343. *Id.* at 593–94.

344. Kumho Tire Co. v. Carmichael, 526 U.S. 137, 150–51 (1999).

345. FED. R. EVID. 702 (amended Dec. 1, 2000). The Rule now permits expert testimony if "(1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case." *Id.*

346. There has been a great deal written about the Supreme Court's rulings concerning the admissibility of expert evidence. For an introduction to the issues, see Margaret A. Berger, *The Supreme Court's Trilogy on the Admissibility of Expert Testimony, in* FED. JUDICIAL CTR. 2000, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 9 (Stephen Breyer ed., 2000).

expressed a great deal of skepticism about the technology in 2000. It is instructive to compare the two cases.

As the Alaskan court noted in *State v. Coon*, some published reports support the use of voiceprints in court.[347] Because the philosophy of the Federal Rules, under which *Daubert* was decided, is to deal with controversy by presenting both sides of an argument (rather than by excluding evidence altogether), the court placed as much or more emphasis on studies sponsored by police officials that advocate for the use of voiceprints as it did on publications from the independent scientific community.[348]

Compare *Coon* to *United States v. Drones*,[349] the Fifth Circuit *habeas corpus* case discussed earlier in connection with expert testimony of aural voice comparison. In that case, a defendant convicted of drug offenses in Texas argued that his lawyer had not effectively represented him at trial because he failed to hire a forensic phonetician to compare the voice on a tape to his voice.[350] In support of his motion, Drones enlisted the help of an expert named Steve Cain, the voiceprint expert whose testimony was allowed by the Supreme Court of Alaska in *Coon*.[351] Cain "reached a finding of 'probable elimination,' meaning that at least 80% of the comparable words in the samples were dissimilar aurally and spectrographically."[352] In response, the government called Bruce Koenig, a former FBI employee, who had been one of the early developers of sound spectrography.[353] Koenig testified that "almost nobody" in the relevant scientific community uses spectrographic voice identification because there is no theoretical basis for the proposition that an individual's voice is truly unique and identifiable.[354]

In reversing the lower court's granting of the *habeas corpus* petition, the appellate court characterized spectrography as "a dwindling science," not widely accepted in the scientific community.[355] It quoted Koenig's testimony to the effect that the number of practitioners of forensic voiceprint analysis had dwindled from about

---

347. 974 P.2d 386, 400–01 (Alaska 1999).
348. *Id.* at 402.
349. United States v. Drones, 218 F.3d 496 (5th Cir. 2000).
350. *Id.* at 498.
351. *Id.*; *Coon*, 974 P.2d at 388.
352. *Drones*, 218 F.3d at 498–99.
353. *Id.* at 499.
354. Based on his hearing the voices, however, Koenig agreed with Cain that the voice on the tape was probably not the defendant's. *Id.* *See* discussion *supra* note 300.
355. *Drones*, 218 F.3d at 504.

fifty or sixty to roughly a dozen, as a result of judicial skepticism of the methodology's scientific basis.[356]

While the Court mentioned the Alaskan decision in *Coon*, it left out the most significant irony: The scientific evidence upon which the court permitted Cain to testify in Alaska was a 1986 article that Koenig had published in the *FBI Crime Lab Digest*.[357] In fact, a close reading of Koenig's article suggests that the Alaska court misstated Koenig's position. In the article, Koenig indeed said that the rate of error in the FBI's use of voiceprints was extraordinarily low (0.31% for false identifications and 0.53% for false eliminations).[358] However, he also said that "meaningful decisions were only made in 34.8% of the requested comparisons."[359] Koenig concluded:

> Spectrographic voice comparison is a relatively accurate, but not positive technique for comparing an unknown voice sample with known verbatim voice exemplars. Present use of the technique is limited to a relatively small number of examiners who confront legal barriers to acceptance, limitations in accuracy and no universally recognized examiner qualifications and examination criteria. Its forensic future may shift to testimony where the judge and jury are advised of the technique's probable accuracy or to nontestimonial use as an investigative aid for law enforcement.[360]

It now appears that just about everyone has jumped ship. Voiceprint analysis can be reliable in a limited number of cases, but the scientific community has not adequately established criteria that define those cases in advance. And the underlying philosophical question—whether each individual's voice(print) is indeed unique— has never been answered.

It is true, of course, that voiceprint analysis may well be as reliable—perhaps even more reliable—than lay witness identification of voices on tape, the limitations of which we discussed earlier. Voiceprints might appear useful to bolster an unconvincing aural identification, or to counter such an identification. The problem with admitting such evidence is that scientific (or quasi-scientific) jargon and data tend to have a very strong impact on the jury. In this sense voiceprint analysis is not unlike the use of lie detectors. Whatever usefulness they have is generally outweighed by the misleading impression of infallibility that they convey.

Consequently, we do not currently advocate the use of voiceprint analysis to identify speakers, although we do not object to

---

356. *Id.* at 503–04.
357. Bruce E. Koenig, *Spectrographic Voice Identification*, 13 CRIME LABORATORY DIG. 105 (1986).
358. *Id.* at 115.
359. *Id.*
360. *Id.* at 117.

phoneticians using acoustic information to enhance aural comparison of voices.[361] Forensic phonetics is a vibrant and productive field that has much to contribute to law enforcement and the judicial process.[362] But, as most of the phonetics community recognizes, voiceprint evidence alone, in which speakers are identified by looking at pictures of their voice signals, is presently too unreliable to allow people's freedom to depend on it, at least if not bolstered by considerable confirming evidence.

In the near future we expect to see improved technology in this area, including the introduction of forensic applications of automatic voice recognition devices currently being developed for academic, commercial, and military purposes.[363] A recent article by a group of researchers from North Carolina reports promising preliminary results using a computer to generate probability functions based on ten acoustic features.[364] A.P.A. Broeders, a prominent Dutch researcher, reports a growing use of automatic and semi-automatic voice recognition technologies by law enforcement agencies around the world.[365] Broeders cautions, however, that considerably more work must be done to make any of these technologies reliable enough to meet evidentiary standards.[366]

In the meantime, it is of crucial importance that both lay and expert identification of voices be admitted in court only with appropriate safeguards. We therefore turn now to some recommendations for improving the treatment of speaker identification in the courts.

IV. Recommendations

Having reviewed in detail some of the problems arising from the present voice identification process in American criminal trials, we make the following recommendations:

---

361. *See supra* note 297.

362. See HOLLIEN, *supra* note 179 and many other references cited earlier in this chapter.

363. For a discussion of avenues of research that have yielded impressive results, see Joseph P. Campbell, Jr., *Speaker Recognition: A Tutorial*, 85 INST. ELECTRICAL ELECTRONICS ENGINEERS PROC. 1437 (1997); J.C.L. Ingram et al., *Formant Trajectories as Indices of Phonetic Variation for Speaker Identification*, 3 FORENSIC LINGUISTICS 129 (1996); Wojciech Majewski & Czeslaw Basztura, *Integrated Approach to Speaker Recognition in Forensic Applications*, 3 FORENSIC LINGUISTICS 50 (1996).

364. R. Rodman, D. McAllister, D. Bitzer, L. Cepeda & P. Abbitt, *Forensic Speaker Identification Based on Spectral Moments*, 9 FORENSIC LINGUISTICS 22 (2002).

365. A.P.A. Broeders, Forensic Speech and Audio Analysis: Forensic Linguistics 1998–2001—A Review 5–6 (Oct. 16–19, 2001) (paper presented at Thirteenth INTERPOL Forensic Science Symposium, Lyon, France) (on file with Hastings Law Journal).

366. Id. at 12–13.

1.   Avoid the Problem of Suggestiveness Altogether

The concern with voice identification that rises to a constitutional level is the danger that a suggestive procedure before trial can taint or contaminate the memory of a witness and lead to a misidentification at trial. Because jurors tend to give great credence to witnesses who were at the scene, a faulty identification can lead to a wrongful conviction. This is surely an issue of utmost importance.

The obvious way to avoid the problem is to ensure that if a witness is to identify the perpetrator at trial, the witness's memory should not be compromised beforehand. The solution, whenever feasible, is to present the witness with not just a single voice but with a voice lineup analogous to a visual lineup. Appropriate procedure governing eyewitness lineups has received a great deal of attention[367] and many of the same considerations apply to voice lineups.[368] We will not detail lineup procedures here. But we do note that the National Institute of Justice has published a booklet on eyewitness testimony for law enforcement agencies that proposes procedures that should be effective for law enforcement without being unduly suggestive.[369] We recommend that similar procedures be adopted for cases involving earwitness identification.

2.   Admit Tape Recordings into Evidence Only with Appropriate Safeguards

With tape-recorded evidence subject to Rule 901, the issue is somewhat less clear. Nonetheless, several adjustments, some in the alternative, seem appropriate.

First, the cases contain remarkably little detailed discussion of what an officer was told when asked to identify a voice on tape. Courts should start making serious inquiries into suggestiveness when the authentication of a tape is challenged. Otherwise, no analysis is possible. In some cases, an officer's identification will be suggestive

---

367. *See, e.g.,* Roy S. Malpass & Patricia G. Devine, *Research on Suggestion in Lineups and Photospreads, in* EYEWITNESS TESTIMONY: PSYCHOLOGICAL PERSPECTIVES, *supra* note 13, at 64.

368. For some suggestions, see A.P.A. Broeders, *Earwitness Identification: Common Ground, Disputed Territory and Uncharted Areas,* 3 FORENSIC LINGUISTICS 3 (1996); Francis Nolan & Esther Grabe, *Preparing a Voice Lineup,* 3 FORENSIC LINGUISTICS 74 (1996). One way in which eyewitness and voice lineups differ is that by their very nature, voice lineups can only be presented sequentially. The literature shows that sequential presentation of photographs in the eyewitness identification context significantly reduces the rate of false alarms. *See, e.g.,* Nancy Steblay, Jennifer Dysart, Solomon Fulero & R.C.L. Lindsay, *Eyewitness Accuracy Rates in Sequential and Simultaneous Lineup Presentations: A Meta-Analytic Comparison,* 25 LAW & HUM. BEHAV. 459 (2001).

369. NAT'L INST. OF JUSTICE, U.S. DEP'T OF JUSTICE, EYEWITNESS EVIDENCE: A GUIDE FOR LAW ENFORCEMENT (1999).

because she was presented with a tape recording and asked whether she can recognize Jones's voice on it. The same problem arises when the officer is given a transcript with Jones's name on it or is otherwise informed that Jones's voice is thought to be on the tape. The same holds true when an officer is presented with a recording of a wiretap and told that the telephone being tapped belonged to Jones. On the other hand, if an officer is presented with a tape recording and asked whether she can identify any of the speakers, the procedure is not suggestive.

Second, as we pointed out earlier,[370] a few courts apply *Biggers* analysis to authentications of recorded speech by the police and virtually always find them not to be suggestive, even when they obviously are. In these cases, it seems that courts are actually relying on strong circumstantial evidence of reliability and paying lip service to the quality of the voice identification. One solution is for courts to begin taking the problems of voice identification more seriously and to look carefully for indicia of reliability, even when the identification is by a police officer and a tape exists. No doubt, courts will find that the circumstances, taken as a whole, support admitting the tape into evidence most of the time. But at least the inquiry will be focused on the relevant issues. Alternatively, courts may wish to apply *Biggers* analysis when the defendant has raised a legitimate question about reliability or when the circumstantial evidence is weak.[371] Either alternative should do more to ensure fairness without unduly burdening legitimate law enforcement techniques.

Third, when there is no tape, *Biggers/Manson* must apply. When a court applies the minimalist approach of Rule 901 without regard to constitutional safeguards in such cases, it has erred.[372]

Fourth, we strongly urge that jurors not receive transcripts with names identifying speakers, unless the issue of identity is conceded or for some other reason is incontrovertible. Such transcripts should also not be provided to officers who are asked to identify the voices on a tape. Transcripts with names are very suggestive and there is no good reason for their use. Speakers can be identified as *speaker A* and *speaker B*, or by means of some similar system of nomenclature.

We also recommend that transcripts not be admitted into evidence. Presently there is a split both in the federal circuits and among state courts on this issue.[373] It is probably unavoidable that

---

370. *See supra* Part I.A.

371. See *supra* note 118 for a discussion of this alternative.

372. See discussion of *United States v. Townsend*, No. 97-60491, 1999 U.S. App. LEXIS 13872, at **29–31 (5th Cir. June 24, 1999), *supra* note 118.

373. Many jurisdictions allow a transcript to be admitted into evidence and used by the jury (including the First, Sixth, Ninth, and Eleventh Circuits and several states). 2 CLIFFORD S. FISHMAN & ANNE T. MCKENNA, WIRETAPPING AND EAVESDROPPING

transcripts of tape recordings will be used at trials. In a relatively literate society such as our own, we tend to place great emphasis on written words. This is problematic where the underlying event is oral. There is, first of all, the likelihood that the written transcript is not accurate in places and is thus subject to the interpretation of the transcriber. Perhaps more importantly, we tend to ascribe a highly literal interpretation to written texts.[374] For both reasons, courts should admit only the tape recording itself and make it clear to the jury that the transcript, if provided at all, is merely an aid to understanding the tape.[375]

3. Modify the *Biggers* Factors in Light of Recent Research into Voice Identification

When a voice identification is found to be suggestive, courts apply the *Biggers* factors to determine whether it is nonetheless reliable enough to let the issue go to the jury. Recall that *Biggers* employed the following criteria:

(1) the opportunity of the witness to view the criminal at the time of the crime;

(2) the witness's degree of attention;

(3) the accuracy of the witness's prior description of the criminal;

(4) the level of certainty demonstrated by the witness at the confrontation; and

(5) the length of time between the crime and the confrontation.[376]

---

§ 24:7, at 24-22 to 24-23 (2d ed. 1995). On the other hand, the D.C. Circuit and several other states do not allow the transcript into evidence. *Id.* at 24-24.

374. Peter M. Tiersma, *A Message in a Bottle: Text, Autonomy, and Statutory Interpretation*, 76 TUL. L. REV. 431, 449 (2001).

375. For a sample instruction, see *Dismuke v. State*, 262 S.E.2d 490, 492 (Ga. Ct. App. 1979), where the court instructed the jury as follows:

> The court states on the bench that this tape might be difficult for you to understand. Of course, you are the judges of the evidence in this case, and it is up to you to determine what if any thing [sic], is said on these tapes. Now, the bailiff is passing out a transcript that was admittedly made by Mr. Osborne, who is, of course, a witness for the state. I want to caution you, you may use these transcripts as you listen to the tapes, so I want to caution you that this transcript doesn't necessarily mean what that [sic] tapes say. You are able to use that as it may help you and it may not. It is not evidence of the fact, it is merely Mr. Osborne's opinion as to what it says. I want to caution you that you must use that with extreme caution, as far as knowing what the truth or what this tape actually says. It is up to you, as jurors in this case, to decide what is said by the conversations on this tape, and to apply them in whatever manner you may seem [sic] fit in the trial of this case. I just wanted to give you those precautionary instructions.

While the comment about "extreme caution" may be a bit strong in the average case, in general we agree with the overall gist of the instruction.

376. Neil v. Biggers, 409 U.S. 188, 199–200 (1972).

Because these factors are oriented toward eyewitness identification, in earwitness cases courts must adapt them to apply to speakers. This adaptation is not always straightforward. For example, the third *Biggers* criterion, the accuracy of the victim's prior description of the criminal, will typically have little application in speaker identification cases. Most people, in our experience, have fairly limited abilities to describe voices, beyond perhaps commenting that a voice is high or low, nasal or tinny. Nonetheless, we would not discard previous identification entirely. It is certainly relevant that a witness described a voice and then later picked a voice matching that description. Moreover, a witness may be able to detect a regional accent, dialectal features, or other nonacoustic information that could help confirm a later identification. Overall, however, this factor should receive less weight than it does in eyewitness cases.

Other *Biggers* factors, such as the witness's degree of attention, the length of exposure, and the time between the crime and the confrontation are indeed relevant to reliability. For cases involving the identification of a voice on tape, the issue would be the time between the exposure to a person's voice and the identification. Research also reveals that delay has a very significant impact on reliability and interacts significantly with the amount of exposure. Courts should bear in mind that their intuitions about these matters are not likely to be reliable. Recall the research showing that accuracy drops significantly after just two or three weeks and can drop to as low as 13% after a mere five months.

The fourth *Biggers* factor, the level of certainty demonstrated by the witness at the confrontation, does not correlate all that well with reliability and should probably not be used as a factor for that reason. This conclusion is bolstered by the fact that witnesses tend to vastly overrate the importance of witness confidence. Nonetheless, we believe that *lack* of certainty should be considered. The fact that a witness is unsure about an identification is probably significant in most cases.

The *Biggers* factors are not sufficient in evaluating reliability of voice identification, however. The research reviewed in this Article has revealed several other significant factors that should be taken into account. One is that people are better at identifying the voice of someone who was previously familiar to them. A second is that the number of exposures to a voice is at least as important to the length of exposure in predicting reliable recall. Changes in tone of voice are also highly significant: A speaker who is under stress or who is very emotional may speak quite differently from how she normally speaks. This difference in tone of voice, as we have seen, makes identification more difficult (and hence, less reliable). The same is true when a speaker whispers or disguises her voice.

Other factors are relevant in assessing reliability. People are better at identifying the voice of someone speaking a language that they understand, and better yet if the speaker does not have an accent.[377] Very old and very young people are not as good at speaker identification. Nonetheless, the empirical results do not seem sufficiently strong to require that these factors be taken into account in deciding whether an identification has met due process standards. Although individual variability may well be an excellent predictor of reliability, current research does not permit inferences strong enough to exclude an identification on that basis at this time.

Research into voice identification continues, so this list is obviously subject to change. We have included only those factors that seem to us to be relatively well accepted by the academic community. Application of these factors will, we believe, lead to admission of more reliable evidence and will consequently reduce the chance of misidentifications and false convictions.

4. Provide the Jury with Helpful Instructions on How to Assess the Accuracy of a Voice Identification

Once an earwitness identification is admitted into evidence, the question goes to the jury. This would be more acceptable if jurors were given some guidance on how to determine whether a voice identification is likely to be reliable. Several jurisdictions presently provide such an instruction in the case of eyewitness identification.[378] In contrast, courts seldom, if ever, give such an instruction in a case of voice identification. In one federal case the trial judge refused to instruct the jury that the "'value [of the voice identification] depends upon the opportunity the witness had to hear the offender at the time of the offense and later make a reliable identification[.]'"[379] The Seventh Circuit affirmed the refusal, holding that "[a]s long as the basic requirement of familiarity with the voice is met, lay opinion testimony is an acceptable means for establishing a speaker's identity."[380] This misses the point. The defendant did not ask that the jury be instructed to ignore the identification. Rather, the request was to advise the jurors that they should "value" the identification only to the extent that they believe it to have been reliable. The jury's ability to make this evaluation is typically given as the

---

377. Obviously, having an accent is a relative term. Tiersma believes that he speaks ordinary American English, but that Solan speaks with a slight New York accent, while Solan believes that his English is standard and that Tiersma has a slight California accent.

378. *See, e.g.*, California Jury Instructions, Criminal (CALJIC) No. 2.91 (6th ed. 1996).

379. United States v. Magana, 118 F.3d 1173, 1208 (7th Cir. 1997) (quoting the defendant's submitted jury instruction).

380. *Id.* (citation omitted).

justification for having such a low threshold for admissibility. Likewise, the Supreme Court of Kansas affirmed a lower court's refusal to give an instruction on earwitness identification.[381] The trial court found that there was "no extensive scientific basis that 'earwitness' identification is as susceptible to the same misidentification as eyewitness identification."[382] As we have seen, voice identification is probably even more problematic than eyewitness testimony. We see no reason for refusing to give an instruction that could help jurors decide more analytically how much weight to give an identification.

We recommend that where voice identification evidence has been admitted, the jury be instructed on how to evaluate that identification. The court should give the jury concrete guidance not only on the factors they should consider, but in some cases on how the factor might influence their decision. Most existing jury instructions fail to do so. Thus, eyewitness instructions may inform jurors that race is relevant but will almost never tell them what the consequence is: that people make less reliable identifications of a person who differs racially from them. In some cases this may be so evident as not to need mention. Where it is not intuitively obvious (as with the impact of a foreign accent), the court should explain its relevance. Since there are no models for such an instruction, we have drafted one and include it in the Appendix to this Article.

5. Where Appropriate, Courts Should Allow Expert Witnesses to Testify on the Reliability of Earwitness Identification

More and more courts have been permitting expert testimony at trial on the unreliability of eyewitness testimony under appropriate circumstances.[383] Courts have also been allowing "educational experts" to point out the shortcomings with other methods of identification, such as identification by handwriting experts, to demonstrate the limits of those techniques.[384] We believe that courts should do the same when it comes to earwitness identifications.

In contrast to jury instructions, which tend to be quite generic, experts can explain to jurors exactly how research into voice

---

381. State v. Burnison, 795 P.2d 32, 39–40 (Kan. 1990).

382. *Id.* at 40. Another reason for refusing the instruction was that the court held the identification was sufficiently reliable.

383. *See* United States v. Downing, 753 F.2d 1224 (3d Cir. 1985). For a recent holding, see *People v. Lee*, 750 N.E.2d 63 (N.Y. 2001). For a discussion of the admissibility of such expert testimony under various circumstances, see FAIGMAN ET AL., *supra* note 306, § 11-1.0.

384. For further discussion, see D. Michael Risinger, *Navigating Expert Reliability: Are Criminal Standards of Certainty Being Left on the Dock?*, 64 ALBANY L. REV. 99, 142 n.171 (2000).

identification is relevant to the case at hand. Not only might experts be able to give testimony more attuned to the actual facts, but they can also bring to bear the most recent research.

To date, the issue has arisen only occasionally, because the scientific evidence surrounding the identification of speakers has been largely absent from the legal literature. A Massachusetts court rejected Professor Yarmey as an expert because the court deemed the identification in that particular case reliable enough to meet the concerns that Yarmey would have raised; nevertheless, the court found that the question of voice identification was in general a proper subject for expert testimony.[385] A federal district court permitted Professor Hollien to testify about characteristics of a defendant's voice that might have an impact on a witness's identification.[386] While courts should limit such testimony to ensure that it is tailored to the issues raised in a particular case, it should certainly not be excluded in principle.

6. In Cases Involving Recorded Evidence of Speech, Experts Should Be Allowed to Testify on Their Identification of the Speakers

We remain skeptical of expert testimony based solely on spectrographic analysis, or voiceprints, largely because it creates the illusion of being more reliable and accurate than it really is. Nonetheless, expert testimony can and should sometimes play a role in identifying speakers. In fact, Professor Hollien has testified in this capacity in many cases.[387]

As we mentioned previously, expert voice identification is not relevant in most earwitness cases, because there will usually have been no recording of the voice at the scene of the crime. Moreover, Rule 901 generally does not envision expert voice identification testimony being necessary to admit a tape recording into evidence. It does not follow, however, that experts must be excluded at trial when the defendant has the opportunity to offer a defense. Once a tape is admitted into evidence, the substantive question of guilt remains to be determined, and at this point the identification of speakers on the tape becomes a more serious issue.

In particular, if an officer is asked to identify a voice on a tape but has no particular personal knowledge about the recording itself and did not participate in the conversation, that officer's testimony is no more competent than an expert in phonetics who has studied the same voice samples. Moreover, as discussed above, experts tend to

---

385. Commonwealth v. Pagano, 710 N.E.2d 1034, 1040 (Mass. App. Ct. 1999).
386. Virgin Islands v. Sanes, 57 F.3d 338, 341 (3d Cir. 1995).
387. *See* United States v. Leon, No. 90-6571, 1992 U.S. App. LEXIS 14323, at **5, 10 n.3 (6th Cir. June 12, 1992).

perform better on voice identification tasks than do non-experts. When experts testify in such cases, they should be competent in fields such as acoustics, phonetics, and linguistics.

As Harry Hollien has pointed out, "while there may be no attribute within a person's speech of sufficient strength and uniqueness to permit that particular individual to be differentiated from all other talkers, the use of a group of features may permit successful recognition."[388] Research has shown that experts do indeed perform better than lay people at this task. In appropriate cases, they should be permitted to testify.

*****

So, was Bruno Hauptmann guilty of kidnapping the Lindbergh baby? Did Captain Hulet chop off the king's head? We really don't know, but we do know that Guy Paul Morin was falsely convicted of raping and murdering a young child. Questions of guilt or innocence are left to the judge or jury. When the evidence includes the identification of the defendant's voice, the legal system must take into account what we have learned over the past half century. Only then can we ensure that those who claim to have heard voices have actually heard them.

Appendix

Proposed Jury Instruction on Reliability of Earwitness Identification

Although this instruction is similar to some existing instructions on eyewitness testimony,[389] it differs in two important respects. First, and rather obviously, it is tailored to the issue of voice identification. Of course, not all of the questions will be relevant to any specific case. Those in square brackets are meant to be optional, for instance, and others may also sometimes be inappropriate.

*****

A witness has identified the voice of the defendant [as being Speaker A on tape recording Number 1][as being the person who committed the charged crime]. In deciding whether this identification was accurate, you should ask yourselves the following questions:

How well could the witness hear the speaker? You should consider factors like background noise, the distance between the speaker and the witness, and similar circumstances.

Was the witness paying close attention to the voice?

---

388. HOLLIEN, *supra* note 179, at 191.

389. This instruction tracks, to some extent, a proposed instruction on eyewitness testimony prepared by the California Judicial Council Task Force on Jury Instructions (Criminal) (on file with Peter Tiersma, who is a member of the Task Force).

Did the witness ever waver or express uncertainty about the identification?

Did the witness hear the live voice of the speaker, or was the voice on tape or on the telephone?

[If the voice was on tape or telephone, what was the quality of the recording or telephone connection?]

How familiar was the witness with the defendant's voice when asked to identify it?

[If the witness was not previously familiar with the voice, how much exposure did the witness have to the voice before being asked to identify it? On how many occasions was the witness exposed to the voice? How much time passed between the witness's original exposure to the voice and being asked to identify it?]

[If the witness gave a description of the speaker's voice, how similar was that description to the voice of the defendant?]

Was the witness ever asked to identify the speaker before trial?

[If the witness was asked to identify the speaker before trial, how many voices did the witness have to choose from? Did the witness ever fail to identify the defendant on these previous occasions? Did the witness change his or her mind?]

[Did the witness identify the voices of other participants in the crime? If so, were these identifications accurate?]

[Is the witness experienced at identifying voices, or is there other evidence that the witness is particularly good at identifying voices?]

Did the speaker have an unusual voice or speech mannerisms that would make his or her speech easier to identify?

Did the speaker try to disguise his or her voice, or was the speaker under stress, making identification more difficult?

[Did the speaker have a foreign accent or speak a foreign language? If so, you should consider that it is often more difficult to identify someone's voice under these circumstances.]

Was there other evidence, besides the defendant's voice, that suggested that the defendant was the speaker?

Were there any other circumstances that would have made identification more or less accurate?

You must consider all the evidence and be convinced beyond a reasonable doubt that it was the defendant who committed the crime. Otherwise, you must find the defendant not guilty.

***