Law Enforcement Executive

FORUM

Table of Contents

Legal, Scientific, and Forensic Controversies Over Spectrographic Voice Analysis for Identification or Elimination[1]

Herbert Joe, MA, JD, LLM, BCFE, Partner, Yonovitz & Joe, LLP
Al Yonovitz, PhD, CCC-A, Partner, Yonovitz & Joe, LLP

Since 1923[2], the standard for admissibility of scientific evidence in federal courts was governed by the "Frye Test," in which evidence was admitted if it was based on principles "generally accepted" within that scientific community. This basically meant that the relevant scientific community would in effect decide whether a particular scientific methodology would be admitted or not. One significant shortcoming of the Frye Test was its failure to distinguish new or novel scientific or technological procedure. That was the law of the land (at least for federal courts) for 70 years until the Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals*[3] ruled otherwise in 1993.

The U.S. Supreme Court overruled the Frye test in the *Daubert* case, which is the foundational case for the admissibility of scientific evidence and arguably affects tort reform more than any other single case. The *Daubert* case made the trial judge the "gatekeeper" in admitting the good sciences and barring the junk sciences. The *Daubert* test, now the rule of law for all federal courts, is a nonexclusive list of factors used by the federal courts to determine the reliability of the expert testimony or scientific principles utilized thereof. If an expert relies on unreliable data or methodology, then his or her entire expert opinion is likewise unreliable and should be excluded from the jury.

Some laws have come down since *Daubert* to clarify its rulings: In *General Electric v. Joiner*,[4] the U.S. Supreme Court held that a trial court's decision to admit or deny expert testimony is reviewed under an abuse of discretion standard at the appellate level. In *Kuhmo Tire v. Carmichael*,[5] the U.S. Supreme Court held that *Daubert* applied to all forms of expert testimony, not just scientific testimony, and in 2000, Federal Rule 702[6] was amended to include the logic of *Daubert* and *Kuhmo Tire*.

Currently, admissibility of expert testimony in federal courts is necessarily governed by *Daubert*; however, state courts are not bound by *Daubert*. A significant minority of states (11) have state statutes that specifically adopt *Daubert*[7] in determining the admissibility of expert testimony in state courts. Thirteen states—including Illinois—are still applying the *Frye* test,[8] and 11 states, along with the military courts, apply a variation of the *Frye* test.[9] The remaining states have a history of rejecting both *Frye* and *Daubert*.

Spectrographic voice analysis was developed at Bell Labs in 1941 and introduced forensically in 1961; however, forensic phonetics, which actually predates spectrographic voice analysis, was the topic of a widely published study on

comparing voices.[10] At any rate, the basic premise of spectrographic voice analysis is that each of us has unique vocal characteristics; sinus cavities, vocal chords, and articulators (i.e., lips, teeth, tongue, etc.) confer vocal characteristics, which are individually unique when spectrograms are made. Unfortunately, there is a (mis)perception by the general public that voice ID (VID) is straightforward. This misperception was reinforced by the introduction of the term *voiceprint* or the phrase *voiceprint identification*,[11] Hollywood dramas like *CSI*, and the like.

VID or voice spectrographic analysis has been controversial since its forensic outset. This is partially manifested by roughly as many jurisdictions that allowed expert testimony on VID as those that did not before *Daubert*.[12] Since *Daubert*, no federal appellate court has approved the admission of voice spectrographic expert testimony into evidence. The Fifth Circuit federal Court of Appeals has stated that the state of the law concerning expert voice identification is "ambiguous" in the wake of *Daubert*.[13] One state Supreme Court, however, has allowed expert testimony on VID since *Daubert*, namely, *State v. Coon*, 974 P.2d 386 (AK, 1999) (admission of testimony based on voice spectrographic analysis was not an abuse of discretion).

The Alaska Supreme Court did note that the scientific literature "permit[ted] a conclusion that there is significant disagreement among experts in the field of voice spectrographic analysis regarding the reliability of the technique."

The same expert who testified in the *Coon* case is the same expert in the *Angleton* case.[14] In *Angleton*, the federal District Court held that the proposed expert testimony on aural spectrographic voice identification method failed to meet the *Daubert* standard for reliability, and the protocol followed by the expert did not protect against several sources of error, further reducing the reliability of the expert's testimony. Stephen Cain, the expert in the *Angleton* case, testified that he followed the protocol of the American Board of Recorded Evidence (ABRE) in his analysis.

Mr. Cain and Dr. Nakasone testified that the ABRE was formed after a dispute arose among the members of the voice identification board of the International Association of Identification (IAI)[15] over the standards for aural spectrographic analysis. In *Angleton*, Dr. Nakasone testified that the group that left the IAI to form the ABRE felt the IAI's standards for voice identification were too stringent because they required examiners to obtain a second opinion and to include a statement of accuracy in their reports. Furthermore, . . .

> The record before this court shows that the remaining proponents of the use of aural spectrographic voice identification for courtroom testimony are a handful of consultants who apply the techniques for the purpose of litigation. The proponents, including Cain, are not performing scientific research in aural spectrographic voice identification and testifying as experts as an aspect of their research work.[16]

In addition, Mr. Cain acknowledged that it is not uncommon for VID analysts reviewing the same recordings to have differing opinions as to the identity of the speakers on the recordings.[17]

The state of *forensic* flux of expert testimony on voice *spectrographic* analysis is also significantly manifested by the changing views of Dr. Hirotaka Nakasone.

Dr. Nakasone has been working in the field of speech recognition since 1977. While working on his doctorate in speech sciences from Michigan State in 1984, Dr. Nakasone conducted voice spectrographic research for the Los Angeles County Sheriff's Department and on several occasions testified as a proponent of the VID method in courts and administrative tribunals. Since 1992, Nakasone has worked for the Federal Bureau of Investigation (FBI), conducting research in audio forensic identification. His current research is in developing computer-assisted voice identification systems. As late as 1989, Dr. Nakasone testified in *U.S. v. Smith*, 869 F.2d 348, 353-54 (7th Cir. 1989) that the voice spectrographic technique was reliable and had a low error rate.

Recently, Dr. Nakasone has testified that his initial belief that the VID technique is sufficiently reliable for courtroom purposes has eroded over time, as a lack of research efforts have failed to support the underlying premises of the voice identification techniques or to produce reliable testing for error rates. Dr. Nakasone has testified credibly that this failure is the basis of the FBI's approach to voice spectrographic analysis. (The FBI does not permit the use of voice spectrographic analysis for courtroom identification but only for investigation.)

The bottom line, however, is that when the courts ask the wrong questions, they will only get the wrong answers. Likewise, when "experts" primarily or exclusively use a flawed technique, then false positive or false negative answers will occur. As emphasized above, there is a justifiable state of *forensic* flux of expert testimony on voice spectrographic analysis.[18]

VID analysis relies extensively on merely matching patterns between spectrograms. Thus, VID analysis (alone) is unreliable because several significant vocal characteristics are not available on spectrograms, and almost every putative expert on VID is not formally (academically or clinically) trained in the speech and hearing sciences[19]—a prerequisite to recognizing and distinguishing the vocal characteristics that are not available on spectrograms but are available through other instrumental means including expert aural perceptual analysis.

Specifically, the Aural-Acoustical methodology for Voice/Speaker Identification or Elimination incorporates a single dimensional scaling of the conclusion along a continuum. (Although the Aural-Acoustical method does not rely on spectrographic analysis as its principal basis, it can be supplemented by spectrographic analysis.) This continuum holds as its basis that at one end of the scale, a very high probability (or positive) "identification (match)" exists, and at the other end of the scale, a very high probability (or positive) "elimination (non-match)" exists.

Assuming samples of the recorded evidence and the exemplar recording contains sufficient and intelligible speech materials, the Aural-Acoustical method of speaker identification or elimination has as its basis both acoustical (objective) and aural (subjective) procedures. It is therefore possible that error exists; however, the probability of error decreases with the competency of the forensic scientist. The methodologies have been long established in the speech, hearing, and language sciences and represent ordered analyses of obtained data.

The speaker identification or elimination procedure employed is one in which an unknown voice is taken from an evidence tape and compared to exemplars of a

known voice. In this manner, samples of a number of comparisons between the unknown and known combinations are placed in pairs or composites for direct and repeated comparisons. The Aural-Acoustic method of analysis follows the protocol and standards described in publications as well as a number of presentations to professional organizations, including the Acoustical Society of America and the American Speech, Hearing and Language Association. The principles of this protocol are to provide a basis for voice/speaker identification or elimination that is consistent with the known principles and evidence-based practice of the hearing, speech, and language sciences.

The Aural-Acoustic method has evolved from earlier standards developed by the IAI[20] (and the ABRE whose standards practically mirror those of the IAI). For example, section VII.B.5 of the 1996 "Voice Comparison Standards" of the Voice Identification and Acoustic Analysis Subcommittee (VIAAS) of the IAI and section 7.2.5 of the "Voice Comparison Standards" of the ABRE are entitled "Speech Characteristics." Speech and hearing scientists and phoneticians are particularly skilled in forensically assessing speech characteristics.[21] Examiners trained in spectrogram pattern matching receive little or no training in the assessment of speech characteristics.

Section VII.B.5.i ("Vocal Quality") of the 1996 "Voice Comparison Standards" of the Voice Identification and Acoustic Analysis Subcommittee (VIAAS) of the IAI states . . .

> Vocal quality is the perception of the complex, dynamic interplay of laryngeal voicing (pitch, intonation, and stress), articulator movement, and oral cavity resonances. *Since each individual's voice is relatively unique in vocal quality, comparisons can provide important information regarding similarities and differences between the voice samples.* (emphasis added, quoted verbatim by the ABRE in its "Voice Comparison Standards," Section 7.2.5.i. "Vocal Quality"[22])

This Aural-Acoustical method uses a number of instrumental or digital signal processing procedures that delineate the microstructure of various vocal qualities or characteristics, such as those described in the acoustic process below. It utilizes, with due caution, the use of these measures, not to overextend the conclusions that may be offered. Two publications discussing the Aural-Perceptual methods at length are Hollien's *Acoustics of Crime* (Plenum, 1990) and Hollien and Hollien's *Forensic Voice Identification* (Academic Press, 2001). The very significant vocal qualities or characteristics present in speech but not in simple spectrograms include the following:

- Complex co-articulation patterns of vowels and consonants
- Voice quality[23] (e.g., measurements of resonance, vocal fry, and/or nasality)
- Linguistic and paralinguistic features (e.g., prosody, rate, and/or melodic patterns)
- Speech abnormalities (e.g., misarticulation and/or fluency)
- Dialect
- Fundamental frequency[24]—absolute and variable
- Jitter[25]
- Shimmer[26]

In summary, there has been and will always be a great need for various applications of speech processing, speech sciences, and speech technology. Voice or speaker identification or elimination is one of those applications. The forensic comparison of voices or speakers for identification or elimination purposes—by any means—is not infallible. VID, especially by those untrained academically and clinically in the speech and hearing sciences, phonetics, linguistics, etc. is not a reliable means to identify or eliminate voices or speakers.[27] There are several significant vocal characteristics that may assist in identifying or eliminating voices or speakers that are not available on spectrograms. The Aural-Acoustic method, with or without supplementing with spectrograms, provides a basis for voice/speaker identification or elimination that is consistent with known principles of the speech, hearing, and language sciences. Under the proper conditions via an academically, clinically, and forensically competent scientist, voice/speaker identification or elimination may be reliably made and academically and forensically tenable between two vocal samples.

Endnotes

[1]   Copyright is not claimed as to any part of the original work prepared by any government entity. The information you obtain in this article is not, nor is it intended to be, legal advice. You should consult an attorney for individual advice regarding your own situation. Except as stated below, none of the material or derivative works thereof may be reproduced, distributed, republished, downloaded, displayed, posted, transmitted, or copied in any form or by any means, without the prior written permission of the copyright owner. Permission is granted to display, copy, distribute, and download the materials in this article solely for personal, noncommercial use, provided that you make no modifications to the materials and that all copyright and other proprietary notices contained in the materials are fully retained. Any unauthorized use of any material contained in this article may violate domestic and/or international copyright laws, trademark laws, the laws of privacy and publicity, and communications regulations and statutes.

[2]   Frye v. U.S., 293 F. 1013 (1923).

[3]   Daubert v. Merrell Dow Pharmaceuticals, Inc., 113 S.Ct. 2786 (1993).

[4]   General Electric Company v. Joiner, 522 U.S. 136 (1997).

[5]   Kuhmo Tire v. Carmichael, 119 S.Ct. 1167 (1999).

[6]   Federal Rules of Evidence Rule 702 (2000) now states that if scientific, technical, or other specialized knowledge will assist the trier of fact in understanding the evidence or help determine a fact issue, then the witness, by virtue of his or her knowledge, skill, training, education, or experience can testify, in the form of an opinion or otherwise, if (i) the testimony is based on sufficient facts or data, (ii) the witness uses scientific methods that are reliable, and (iii) the witness properly applied those reliable scientific methods to the facts of the case.

[7]   Connecticut, Indiana, Kentucky, Louisiana, Massachusetts, Missouri, New Mexico, Oklahoma, South Dakota, Texas, and West Virginia

[8] Alaska, Arizona, California, Colorado, Florida, Illinois, Kansas, Maryland, Michigan, Nebraska, New York, Pennsylvania, and Washington

[9] Arkansas, Delaware, Georgia, Iowa, Minnesota, Montana, North Carolina, Oregon, Utah, Vermont, and Wyoming

[10] McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology, 17*, 249-271.

[11] Kersta wrote a 1962 article published in *Nature* using those words. See Kersta, L. G. (1962). Voiceprint identification. *Nature, 196*, 1253-1257.

[12] See generally United States v. Leon, 966 F.2d 1455 (6th Cir. 1992) (admissible); United States v. Smith, 869 F.2d 348 (7th Cir. 1989) (admissible); United States v. Williams, 583 F.2d 1194 (2d Cir. 1978) (admissible); United States v. McDaniel, 538 F.2d 408 (D.C. Cir. 1976) (inadmissible); United States v. Jenkins, 525 F.2d 819 (6th Cir. 1975) (admissible); United States v. Baller, 519 F.2d 463 (4th Cir. 1975) (admissible); United States v. Franks, 511 F.2d 25 (6th Cir. 1975) (admissible); United States v. Addison, 498 F.2d 741 (D.C. Cir. 1974) (inadmissible); United States v. Maivia, 728 F. Supp. 1471 (D. HI. 1990) (admissible); United States v. Williams, 443 F. Supp. 269 (S.D.N.Y. 1977) (admissible); United States v. Sample, 378 F. Supp. 44 (E.D.P.A. 1974) (admissible); United States v. Raymond, 337 F. Supp. 641 (D.D.C. 1972) (admissible); United States v. Wright, 37 C.M.R. 447 (1967) (admissible); State v. Gortarez, 686 P.2d 1224 (AZ 1984) (inadmissible); People v. Kelly, 549 P.2d 1240 (CA 1976) (inadmissible); People v. Law, 114 Cal. Rptr. 708 (Cal. 1974) (inadmissible); Hodo v. Superior Court, 106 Cal. Rptr. 547 (Cal. 1973) (admissible); People v. King, 72 Cal. Rptr. 478 (Cal. 1968) (inadmissible); People v. Drake, 748 P.2d 1237 (CO 1988) (inadmissible); Brown v. United States, 384 A.2d 647 (D.C. 1978) (neither); Alea v. State, 265 So. 2d 96 (Fla. App. 1972) (admissible); Worley v. State, 263 So. 2d 613 (Fla. App. 1971) (admissible); Cornett v. State, 450 N.E.2d 498 (IN 1983) (inadmissible); State v. Free, 493 So. 2d 781 (LA 1986) (inadmissible); State v. Williams, 388 A.2d 500 (ME 1978) (admissible); Reed v. State, 391 A.2d 364 (MD 1978) (inadmissible); Commonwealth v. Lykus, 327 N.E.2d 671 (MA 1975) (admissible); People v. Tobey, 257 N.W.2d 537 (MI 1977) (inadmissible); State ex rel. Trimble v. Hedman, 192 N.W.2d 432 (MN 1971) (inadmissible); Windmere, Inc. v. International Insurance Company, 522 A.2d 405 (NJ 1987) (inadmissible); D'Arc v. D'Arc, 385 A.2d 278 (N.J. Super. Ct. 1978) (inadmissible); State v. Cary, 239 A.2d 680 (N.J. Super. Ct. 1970) (inadmissible); People v. Bein, 453 N.Y.S.2d 343 (N.Y. Sup. Ct. 1982) (admissible); People v. Collins, 405 N.Y.S.2d 365 (N.Y. Sup. Ct. 1978) (inadmissible); People v. Rogers, 385 N.Y.S.2d 228 (N.Y. Sup. Ct. 1976) (admissible); State v. Williams, 446 N.E.2d 444 (OH 1983) (admissible); State v. Olderman, 336 N.E.2d 442 (OH 1975) (admissible); Commonwealth v. Topa, 369 A.2d 1277 (PA 1977) (inadmissible); State v. Wheeler, 496 A.2d 1382 (RI 1985) (admissible)

[13] See U.S. v. Drones, 218 F.3d 496, 503 (5th Cir.2000). Like the *Coon* court, the *Drones* court noted the "uncertainty of the current state of the law regarding the reliability and admissibility of expert voice identification evidence" (218 F.3d at 504).

[14] U.S. vs. Angleton, 269 F.Supp.2d 892 (S.D.TX 2003).

[15] One of the authors, Dr. Yonovitz, is a former member of the certification and standards committee of the International Association of Identification (IAI).

16 *Id.* at 902.

17 *Id.* at 904, citing Docket Entry No. 160, p. 158, 1.10-1.14.

18 One of the authors, Al Yonovitz, is an associate professor of the speech and hearing sciences with a doctorate in acoustics and having been involved in the academic, scientific, clinical, research, and forensic aspects of vocal production for over 30 years, opines that voice/speech/speaker identification via the Aural-Acoustic method, with or without secondary analyses via spectrography, is an accepted academic method for VID.

19 The authors recognize Dr. H. Nakasone, Dr. Harry Hollein, and Dr. Tito Poza as peer experts in the forensic area of the speech sciences. There may be other U.S. expert or forensic examiners qualified to perform proper VID via the Aural-Acoustic method.

20 The IAI ceased certifying voice identification examiners in 1999 and ceased all voice identification activity in December 2002.

21 One of the authors, Dr. Al Yonovitz, has written numerous publications, given international presentations, produced abstracts, and taught undergraduate and graduate courses related to speech and hearing.

22 Approved by the ABRE Voice ID Board in April 1999.

23 Voice quality encompasses the perception of the listener of the overall sound of the talker's voice. Just as different musical instruments produce different wave compositions, the human voice is similar. It is this overtone structure or timbre that can differentiate one voice from another.

24 The perceived pitch is the psychophysical correlate of fundamental frequency.

25 Jitter is a frequency perturbation of the glottal source signal.

26 Shimmer is amplitude perturbation of the glottal source signal.

27 Spectrograms may be very useful to engineers and as a voice analysis (e.g., speech pathology) tool.

Herbert Joe is an attorney who has four degrees, including two science degrees and two law degrees. The following are a result of his expertise in the area of forensic audio/video: Board Certified Forensic Audio/Video Examiner; Diplomat, American Board of Forensic Examiners; Diplomat, American Board of Law Enforcement Experts; Licensed Instructor, Texas Board of Private Investigators; Board of Legal Advisors, American Guild of Court Videographers; Fellow, American Guild of Court Videographers; and Fellow, American College of Forensic Examiners. His 21-year involvement in the area of forensic audio/video includes regional, national, and international presentations. He has also authored publications and been interviewed by ABC, BBC, and FOX television. Joe is a partner in Yonovitz & Joe, LLP, a team of forensic audio/video analysts, experts, and consultants.

Al Yonovitz, PhD, has a doctorate in physiological and psychological acoustics and 30+ years of teaching and research, including appointments at the Speech and Hearing Institute; Graduate School of Biomedical Sciences; the School of Public Health at the University of Texas Health Science Center, Houston; the Baylor College of Medicine; the Department of Biomedical Engineering at the University of Houston; and the Conley Speech and Hearing Center, University of Maine. He has been a consultant to the VA Hospital, Houston, on voice research in psychiatric patients. Undergraduate and graduate classes taught by Professor Yonovitz include Auditory Systems and Disorders, Audiology, Seminar in Fluency Disorders, Biomedical Instrumentation, Industrial Audiometry and Hearing Conversation, Computer Applications in Speech Pathology and Audiology, Hearing and Speech Science, Research Methods in Speech Pathology and Audiology, Physiological and Psychological Acoustics, Special Topics: Middle Ear Mechanics, Special Topics: Audiometry with the Difficult to Test, Special Topics: Measurement of Voice, Hearing Impairment, Anatomy of the Speech and Hearing Mechanism, Aural Rehabilitation, Speech Science, Introduction to Audiology, and Audition. He is a partner in Yonovitz & Joe, LLP.