J Forensic Sci, July 2004, Vol. 49, No. 4
Paper ID JFS2004083
Published 26 May 2004
Available online at: www.astm.org

BOOK REVIEW

Thomas L. Bohan,[1] *Ph.D., J.D.*

Review of: *Forensic Voice Identification*

REFERENCE: Hollien H. *Forensic voice identification*. Academic Press, SanDiego, CA, 2002, 240 pp.

Dr. Hollien states that his most important reason for writing this book[2] was to reduce the gap separating the two components of voice-identification professionals: engineers and phoneticians. As one of the latter group, he attributes the gap as arising, on the one hand, from the engineers' belief that the phoneticians are insufficiently quantitative and, on the other, the phoneticians' belief that the engineers are ignorant of the "soft" information that assists in voice identification.[3] Since, by profession and temperament, I value quantitative evaluations of forensic techniques and am impatient with hand-waving arguments, I approached this review biased toward the engineering camp, which meant that I would be able to gauge how well the author succeeds in his gap-narrowing effort. Stated differently, I felt that I could gauge how half of his target audience would receive the book.

As a further reviewer-disclosure, I note that I opened this book with the hope and expectation that it would educate me on two specific points: (1) whether a voice-identification specialist is better than a lay listener at voice recognition based simply upon listening to a voice or its recording; (2) the scientific and legal status of voice spectrograms (voiceprints). Because of the second interest, I turned first to Chapter 6, entitled Voiceprints, a topic the author says he would have liked to ignore completely.[4] Once in Chapter 6, my attention was immediately drawn to the author's mention of a talk on voiceprints that I had attended,[5] a talk that had remained fresh in my mind because of the audience's exceedingly hostile reaction, especially to the speaker's advocacy for the broader use of voiceprints, all as related by Dr. Hollien.[6] I was, however, disappointed by the rest of the voiceprint discussion. Although the author leaves no doubt of his own conviction that voiceprints are bogus and harmful (presumably in terms of misidentifications in criminal cases), he does not discuss the views of the scientific community—however defined—in such a way that one can draw his or her own conclusions about the technique's scientific acceptance. Neither does he provide any indication of acceptance by trial judges and appeals courts. One whose knowledge was based entirely or even primarily on Chapter 6 would be quite surprised to learn that many of state high courts addressing voiceprints have upheld their admission,[7] the most recent high-court acceptance being in 1999.[8]

Dr. Hollien does provide enough details regarding voiceprint technology to inform the reader that, in contrast to fingerprints, voiceprints depend strongly on the parameters used in their recording, parameters including the choice of bandwidth over which to average the signal.[9] Nevertheless, when they were first introduced to the public in the 1960s, they were touted as "voice fingerprints." Their gross appearance helped promote this identification with fingerprints. As with fingerprints, voiceprints as presented seem to feature ridges and valleys and even whorls. Also, like the fingerprint technician comparing specific features of a partial fingerprint from the crime scene with corresponding features of known full prints, the voiceprint technician compares features of the unknown voiceprint with those of voiceprints from known persons. And the similarity does not end there, since both types of practitioners claim the right to do subjective, non-quantifiable "matches." However, the voiceprint technician goes much further into the subjective realm than the fingerprint technician, who can and does point to match points. It seems that the process of declaring a voiceprint match is

[1] MTC Forensics, Portland, ME 04101.

[2] He depicts the book as an expansion of three chapters in his earlier *Acoustics of Crime*, Plenum Press, 1990.

[3] This is my paraphrasing of the author's words.

[4] He says he was torn between not wanting to publicize a technique best forgotten and a need to educate those entering the profession so that they will be ready for those diehards who continue to support the evidentiary use of voiceprints.

[5] It was in the Engineering Sciences Section of the 1996 Annual Meeting of the American Academy of Forensic Sciences. In a minor historical error, the book says that it was in 1998.

[6] Handwritten margin notes next to this talk's abstract in my 1996 Academy *Proceedings* include comments from the audience such as "Fraud" and my own comment "[speaker says] accepted by all the courts except Maryland and California."

[7] Interestingly, most of these cases came down in the late 1970s, just before and after the the National Academy of Sciences' report "On the Theory and Practice of Voice Identification" mentioned in Chapter 6. My count of the state high court score on voiceprints is eight, in favor nine against.

[8] *State of Alaska v Coon*, 974 P. 2d (Alaska, 1999), the case marking the shift of Alaska from the *Frye* standard for scientific evidence admission to *Daubert*. The Maine Supreme Court also chose a voiceprint case in which to take Maine out of the *Frye* column, though without providing trial judges any guidelines such as *Daubert* did– *State of Maine v Williams*, 388 A.2d 500 (Me. 1978).

[9] The voiceprint (referred to by its advocates as a "sonogram," a word to which the author takes exception, preferring the more general "spectrogram," devoid of any root referring to sound) is a two-dimensional plot of voice frequency and intensity as a function of time. Time is plotted on the x-axis and frequency on the y-axis, with the intensity (loudness) represented by the darkness of the trace. It appears as a multivalued function of time, because of the combination of frequencies that constitute any given phoneme or fragment of "a tiger's roared greeting" (see page 118).

Copyright © 2004 by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959.

essentially if not totally subjective, with little or no attempt made to articulate what the would-be matching is based on. One must just let the patterns wash over one. That, combined with the dependence of the voiceprint on its means of collection would seem to provide more than enough grounds for a repudiation of such evidence on quantitative grounds, namely that there are no quantitative means of testing the reliability of the technique, in spite of the 99% accuracy rates that the author reports being claimed by its advocates. Dr. Hollien does refer to a lack of standards and even uniformity in presenting voiceprint evidence.

I was also disappointed in the author's failure to satisfy my curiosity regarding the expert ear versus the "lay ear" when it came to aural matching. Although he *says* that the experts are clearly better than the non-experts, he fails to provide clear discussions of experiments quantitatively supporting this statement. Indeed, in at least one of the tests he reports, the results seem to contradict the statement.[10] This is important, since it goes to the question of whether voice identification specialists should be allowed to give expert testimony regarding aural voice identification.

The author distinguishes "speaker identification" from "speaker verification." The former (labeled with the unpleasant acronym SPID[11]) requires the solution of an "open" puzzle.[12] Given a voice (heard, but not recorded, in most cases), whose was it? When a recording of the unknown voice is available, and a comparison is to be made with a number of known voices, the analogy can be made to partial fingerprints recovered at a crime scene and the attempt to match them with prints in a database.

Although the author repudiates the naked use of voiceprints for identification, he accepts speech spectrograms and other machine-generated speech information in general as helpful aids to voice recognition, machine-assisted SPID being pointed to as the future of the field. An entire chapter is devoted to a discussion of the Semi-Automatic Speaker Identification (SAUSI) system developed by the author, who reports it to have been in use throughout the field over the past 35 years.

There is an extended, and helpful discussion of "earwitness lineups," with much analogizing to the eyewitness lineups, along with the author's observation that the analogy has some fundamental problems. For example, when it comes to "simultaneous" visual lineups of the traditional kind,[13] the witness can continually glance back and forth between the persons in the lineup, comparing and contrasting. When it comes to aural identification, however, the earwitness is presented with a sequence of recorded speech samples of different people. Depending on the protocol followed in the particular jurisdiction, there is no going back for second or third listenings. (One concern is that as the sample becomes more familiar, the witness will begin to think that it is familiar because it resembles the speech of the perpetrator.)

The author writes in a style that may be particularly off-putting to engineer-types, in spite of his warning at the outset that he will use a style that is unusual because it is jargon-free. In spite of his weakness for non-intuitive acronyms, he does avoid jargon. However, it was not the use of non-technical language that I found distinctive, but rather the rambling sentences which tended to madden me. Because the author has been writing books for a long time, I infer that the rambling is deliberate, perhaps related to a belief that it makes the text seem less intimidating to non-specialists and in particular to non-technical readers. However; since this book is expressly directed to engineers (among others), I think that the style was a mistake. Engineers and physical scientists tend to like bare statements of fact in preference to statements that through a series of circumlocutions creep up on the fact. Having said that, I also say that the author's style will suit many readers just fine; my mentioning it is not a fundamental criticism of the book so much as it is an alert to sensitive engineers.

Finally, a grammatical point that I do not believe to be just a quibble because I do not think that it is just a personal failing of mine that I lose my reading rhythm whenever I strike a mid-sentence subject change from second person to third person. There were hundreds of such rhythm losses in my reading of this book (typically two or three per page), sentences that began recounting something about a person—a witness, an investigator, a student, a criminal, etc.—only to have that person turn into a multitude by mid-sentence, and sometimes back again by sentence end. Typical is the sentence at the bottom of page 8: "Worse, *the suspect* may attempt to disguise *their voice*." [Emphasis added.] This type of sentence, with its second-to-third switch being immediately followed by third-to-second switchback, could give a fast reader whiplash. Where was the Academic Press editor? There is no excuse for this, especially in a book which also includes a number of sentences where subject number is conserved throughout. Also, it should be embarrassing for *some*body that, in a book about scientific evidence, the name of the key scientific-evidence case for most of the 20th century is misspelled throughout.

In conclusion and summary, I list the overall structure of *Forensic Voice Identification*. After a folksy introduction that presages the style used throughout, the author presents the following chapters: History (10 pp), Aural-Perceptual Approaches (36 pp), The Professionals (28 pp), Voiceprints (20 pp), Machine Approaches (20 pp), and SAUSI (36 pp). There are then 32 pages of references, 554 keyed to points within the individual chapters, an author index, and a skimpy subject index (less than the length of the author index).

[10] For example, the book's Table 3.1 reports that 100% of the five test subjects who were forensic phoneticians correctly matched voice samples recorded four weeks apart and that 93% of the *sixty-seven* lay-listener subjects did so. Given the size of the respective samples, it appears that there was essentially no difference between the two groups' success rate.

[11] A Google search shows that it is also an acronym with a different voice-related meaning.

[12] Speaker verification, on the other hand, is a closed puzzle, the task being to determine whether a particular voice matches one of a limited number of previously selected voices.

[13] This traditional type of lineup should be a vanishing exercise, one susceptible to a witness's tendency to settle for the *closest* match when unable to make a positive identification.