Some fundamental considerations regarding voice identification

Thomas Shipp and E. Thomas Doherty
Speech Research Laboratory, VA Medical Center, San Francisco, California 94121

Harry Hollien
Institute for Advanced Study of the Communication Processes, ASB 63, University of Florida, Gainesville, Florida 32611

(Received 12 January 1987; accepted for publication 28 April 1987)

Several important areas need substantial clarification or expansion before the reported findings of Koenig, "Spectrographic voice identification: A forensic survey" [J. Acoust. Soc. Am. 79, 2088–2090 (1986)], can be readily accepted. They are: (1) the method of "voiceprint" analysis used, (2) "voiceprint" examiners' qualifications, and (3) the means for determining the FBI's correct identification.

PACS numbers: 43.70.Jt, 43.72.Pf

The conclusions reached by Koenig (1986) in his Letter to the Editor were provocative and, at the same time, disturbing. That is, his terse letter concluded that Federal Bureau of Investigation (FBI) voiceprint examiners exhibited error rates of less than 1% for those speakers they could either positively identify or eliminate (35% of 2000 evaluations). Thus it would appear that, for 15 years, the FBI has employed a method of spectrographic voice identification that is virtually error free, a conclusion we feel is unwarranted based on the strength of this report. Our concerns are, principally, ones of methodology.

Critical to accepting this report would be a description of the methods used by the FBI in voiceprint analysis and a detailed description of the perceptual procedures added (at some point) to the spectrographic evaluations. Instead, Koenig concentrates on the qualifications of the voiceprint examiners and their methods of tape handling, while presenting only incomplete descriptions of the actual methods used by the FBI in identifying talkers using spectrographic and aural/perceptual comparison.

Another matter needing clarification relates to Koenig's statement that exactly the same analysis techniques were used by FBI examiners over the designated 15 years. Such a constant seems highly unlikely (and, certainly, unwise) in light of the advances in technique and technology that have occurred during this period in the areas of signal recording and analysis and in talker identification.

Another area needing further elaboration concerns qualification procedures for the examiners. Pertinent to the FBI's certification process is the type used and other criteria employed for assessment of an examiner's proficiency or aptitude. To date, no evidence has been published to suggest that any examiner has passed objective and independent evaluations before being certified.

The assumption in this survey that a correct decision was made when the identification was consistent with the outcome of the case in question (when a jury's conviction or exoneration of a talker matched the FBI's conclusion) is inappropriate. We have little quarrel with statements that simply report the percent agreement or disagreement between that outcome and the examiners' decisions. On the other hand, we feel strongly that criteria such as the determination of guilt or innocence are not sufficient for establishing the correctness of the judgments. Identification or elimination "consistent with interviews and other evidence in the investigation" is not an adequate standard to confirm the accuracy of such decisions. Any technique such as this must be proven to be valid and reliable by some independent means before it can be used. Here, the technique is used to support the guilt or innocence of an individual, while the determination of such guilt or innocence is then used to show that the technique is proper. The data provided in this report cannot do both. First, the possibility exists that the judgment and/or the findings of the investigators may be incorrect. Also, the outcome of a case may be influenced by the report of the examiner. For example, an investigator, especially one who believes that voice identification is an accepted scientific procedure, may choose to pursue or ignore a suspect on the basis of the voice identification report contrary to his actions without such information.

The determination of correct identifications is further compromised since the methodology places the responsibility on the user agency for informing the investigator if additional evidence challenged earlier findings. A procedure that leaves to the contracting agency the burden for reporting altered decisions is suspect for use in research since the agency may fail to provide such data. For this survey, it would have been prudent to have contacted those agencies involved to confirm the ultimate disposition of the case.

Along similar lines, we are also uncomfortable with Koenig's conclusion that his results represent expected minimum error rates "since investigators are not always correct
in their evaluations of a suspect's involvement. There being no similar discussion of the fallibility of the examiners, one is led to conclude that the two false eliminations and the one false identification were the result of error by the field investigators and not the FBI examiners.

While Koenig's letter raises many other concerns and questions, it seems important to note that speaker identification based on time-frequency-amplitude spectrograms has been rejected by the courts primarily because of demonstrated high error rates by independent investigators. More importantly, the voiceprint approach to talker identification has not been accepted by the scientific community (Bolt et al., 1973). Reports such as Koenig's certainly provide interesting archival information; however, they add little to the scientific database for determining the validity and/or reliability of the technique. Worse yet, such reports tend to be counter-productive as they may well provide a type of surface validity for using this approach as a means for solving the talker recognition problem.

A topic as controversial as talker identification warrants a more detailed report than provided in a letter to the editor. The scientific community has a paramount need for all such studies to be reported with clarity so that data may be weighed fairly in determining the validity of techniques used in the identification of talkers from recordings.


Reply to "Some fundamental considerations regarding voice identification" [J. Acoust. Soc. Am. 82, 687-688 (1987)]

Bruce E. Koenig, Donald V. Ritenour, Jr., Barbara A. Kohus, and Artese Savoy Kelly
Engineering Section, Technical Services Division, Federal Bureau of Investigation, 8199 Backlick Road, Lorton, Virginia 22079

(Received 3 April 1987; accepted for publication 28 April 1987)

This letter is a reply to comments made by T. Shipp et al. on a previous article by B. E. Koenig [J. Acoust. Soc. Am. 79, 2088-2090 (1986)].

PACS numbers: 43.70.Jt, 43.72.Pf

We would like to thank the authors of "Some fundamental considerations regarding voice identification"1 for their thoughts on particular aspects of spectrographic voice identification in forensic applications. Comments are always appreciated, both formal and informal, to clarify and enhance scientific endeavors in this field.

Most of the opinions expressed by T. Shipp et al. are in agreement with the findings of "Spectrographic voice identification: A forensic survey,"2 though many areas not directly related to the survey, or previously covered in other papers, were not set forth. To allow further clarification, the following additional information is provided.

(1) We do not consider a technique with an error rate of greater than 0.5% to be "virtually error free" in a forensic laboratory examination, since, for example, fingerprint comparisons are 100% accurate. A wrong voice comparison decision can potentially put an innocent individual in prison or free a murderer or drug dealer. All Federal Bureau of Investigation (FBI) reports state that the comparison "is not considered a positive means of speaker identification."

(2) The actual method of conducting comparisons varies slightly between examiners, but the general methodology was set forth in the survey article. Further information on comparison practices is available in other literature.3-15 Analysis techniques have been relatively consistent over the last 15 years, with the exception of better enhancement devices, which have slightly decreased the percentage of no- or low-confidence decisions. Spectrum analyzers, special tape recorders with azimuth alignment and speed correction adjustments, etc., have been available in the FBI laboratory during the entire survey period.

(3) Assessment of proficiency and aptitude of individuals being trained is done continually by intensive on-the-job apprenticeship, direct observation, and testing. Factors considered in the evaluations include accuracy of decisions under known conditions, ability to handle pattern recognition tasks, acuity of aural discrimination, etc. Education and other criteria have previously been set forth.2,16

(4) Investigative judgments are not a fail-safe method of determining the exact accuracy of spectrographic comparisons, as stated in the first conclusion of the survey.2 However, experimental results alone are not always directly applicable to actual examinations, since testing of only a few variables cannot imitate the great number of factors encountered in forensic cases. Varying telephone noise and channel degradations, voice disguises,17-19 emotion induced changes,20 drug and alcohol physiological effects, etc., interact independently and cumulatively to vastly complicate the voice comparison task. The survey is thus a realistic attempt to report the approximate range of error expected when

B. E. Koenig
Federal Bureau of Investigation
8199 Backlick Rd.
Lorton, VA 22079

(Received 3 April 1987; accepted for publication 28 April 1987)

This letter is a reply to comments made by T. Shipp et al. on a previous article by B. E. Koenig [J. Acoust. Soc. Am. 79, 2088-2090 (1986)].

PACS numbers: 43.70.Jt, 43.72.Pf

We would like to thank the authors of "Some fundamental considerations regarding voice identification"1 for their thoughts on particular aspects of spectrographic voice identification in forensic applications. Comments are always appreciated, both formal and informal, to clarify and enhance scientific endeavors in this field.

Most of the opinions expressed by T. Shipp et al. are in agreement with the findings of "Spectrographic voice identification: A forensic survey,"2 though many areas not directly related to the survey, or previously covered in other papers, were not set forth. To allow further clarification, the following additional information is provided.

(1) We do not consider a technique with an error rate of greater than 0.5% to be "virtually error free" in a forensic laboratory examination, since, for example, fingerprint comparisons are 100% accurate. A wrong voice comparison decision can potentially put an innocent individual in prison or free a murderer or drug dealer. All Federal Bureau of Investigation (FBI) reports state that the comparison "is not considered a positive means of speaker identification."

(2) The actual method of conducting comparisons varies slightly between examiners, but the general methodology was set forth in the survey article. Further information on comparison practices is available in other literature.3-15 Analysis techniques have been relatively consistent over the last 15 years, with the exception of better enhancement devices, which have slightly decreased the percentage of no- or low-confidence decisions. Spectrum analyzers, special tape recorders with azimuth alignment and speed correction adjustments, etc., have been available in the FBI laboratory during the entire survey period.

(3) Assessment of proficiency and aptitude of individuals being trained is done continually by intensive on-the-job apprenticeship, direct observation, and testing. Factors considered in the evaluations include accuracy of decisions under known conditions, ability to handle pattern recognition tasks, acuity of aural discrimination, etc. Education and other criteria have previously been set forth.2,16

(4) Investigative judgments are not a fail-safe method of determining the exact accuracy of spectrographic comparisons, as stated in the first conclusion of the survey.2 However, experimental results alone are not always directly applicable to actual examinations, since testing of only a few variables cannot imitate the great number of factors encountered in forensic cases. Varying telephone noise and channel degradations, voice disguises,17-19 emotion induced changes,20 drug and alcohol physiological effects, etc., interact independently and cumulatively to vastly complicate the voice comparison task. The survey is thus a realistic attempt to report the approximate range of error expected when
comparisons are conducted under actual investigative conditions with well-trained and equipped personnel.

(5) The FBI and other reputable laboratories are constantly striving to reduce examiner errors. Since the onset of its laboratory program in the early 1930's, the FBI has built a reputation based on objectivity and the elimination of analysis mistakes. In spectrographic comparisons all decisions are double-checked by the assigned examiner, independently reexamined by at least one additional qualified expert, and then the written report and notes are reviewed by a knowledgeable supervisor.

(6) The statement that "speaker identification based on time-frequency-amplitude spectrograms has been rejected by the courts primarily because of demonstrated high error rates by independent investigators" is not generally true, since no legal consensus has been reached concerning the admissibility of spectrographic comparisons. The examiners of the FBI have never testified to a voice comparison analysis using the spectrographic (voiceprint) method, and do not support its use as a positive, or conclusive, technique of comparing voices in legal applications. The FBI uses the comparisons for "investigative guidance" only, and requires contributors to state in writing that they understand this policy.

(7) The opinion of the appropriate "scientific community" is of considerable importance, which is why the FBI funded the 161-page, 1979 National Academy of Sciences report on spectrographic theory and practice. The FBI continuously monitors scientific endeavors and opinions in the spectrographic field, and supports research and surveys that add to the scientific and legal understanding of the technique's strengths and limitations.

Spectrographic comparisons are not a new and may never be the final solution to the forensic voice identification problem. However, it is an excellent investigative aid for law enforcement agencies that has resulted in not only identifying suspects in cases involving espionage, kidnapping, child molestation, consumer product tampering, etc., but over half the FBI decisions have contributed to the elimination of a possible suspect. The voiceprint technique, when properly applied under actual investigative conditions, allows objective decisions that are not presently available using any other voice identification system.

15See Ref. 15, p. 116.
31Contributors sign a form that states in part: "For your information, it is the established policy of the FBI that the results of a voiceprint examination be used only for your investigative assistance. No testimony by FBI experts will be provided. Should you find the above condition acceptable please indicate by signing this letter in the space provided. If you desire that no voiceprint examination be conducted in this matter, so designate by checking the space at the very bottom of this letter. The evidence will be promptly returned to you via registered mail." An enclosed self-addressed envelope has been provided for your convenience. No postage is necessary. The examination is being held in abeyance pending receipt of the requested information."