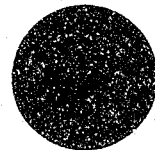


146791

146992

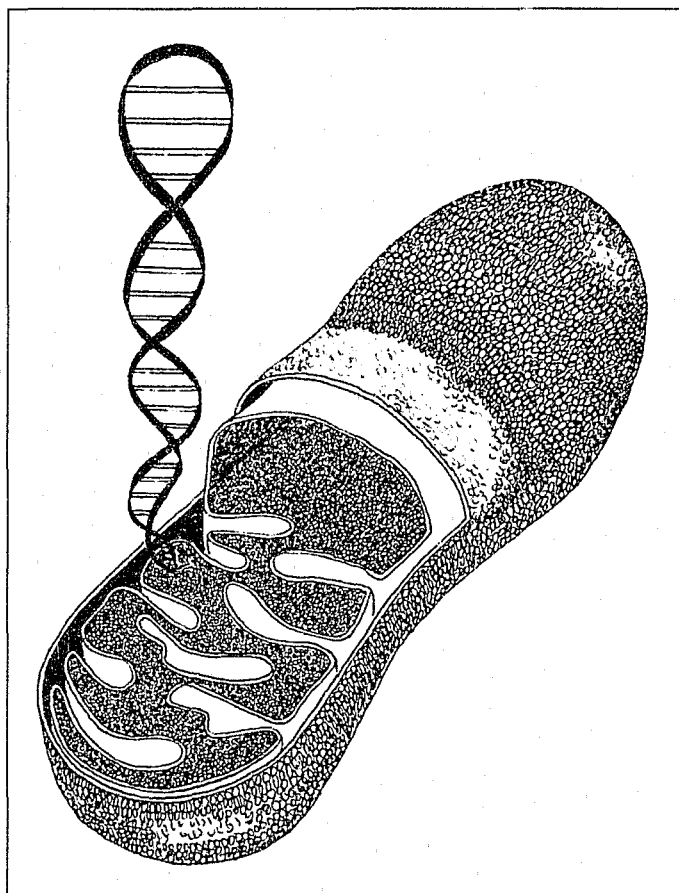
U.S. Department of Justice
Federal Bureau of Investigation



CRIME LABORATORY DIGEST

Volume 20, Number 4

October 1993



IN THIS ISSUE

Guidelines for the Use of Mitochondrial
DNA Sequencing in Forensic Science

Selected Topics in Forensic Voice Identification

CRIME LABORATORY DIGEST

Volume 20, Number 4

October 1993

ARTICLES

68

Guidelines for the Use of Mitochondrial DNA Sequencing in Forensic Science

by Mark R. Wilson, Mark Stoneking, Mitchell M. Holland,
Joseph A. DiZinno, and Bruce Budowle 146991

78

Selected Topics in Forensic Voice Identification

by Bruce E. Koenig 146992

DEPARTMENTS

67

Editor's Note

67

Book Review

67

Employment Opportunity

82

Meeting Announcements

Special thanks to Glenda H. Wilson for providing her original artwork, which is depicted on the cover of this issue.

The *Crime Laboratory Digest* (ISSN 0743-1872) is a quarterly forensic science journal published by the FBI Laboratory Division in cooperation with the American Society of Crime Laboratory Directors (ASCLD). It is intended to serve as a means of communication between crime laboratories, permitting information of interest and value to be disseminated among crime laboratory scientists.

Postmaster: Send address changes to Editor, *Crime Laboratory Digest*, FSRTC, FBI Academy, Quantico, VA 22135.

146991-

**U.S. Department of Justice
National Institute of Justice**

146992

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Justice.

Permission to reproduce this ~~copyrighted~~ material has been granted by

Public Domain/FBI

U.S. Department of Justice

to the National Criminal Justice Reference Service (NCJRS).

Further reproduction outside of the NCJRS system requires permission of the ~~copyright~~ owner.

Selected Topics in Forensic Voice Identification



Bruce E. Koenig
Federal Bureau of Investigation
Quantico, VA

Ongoing law enforcement operations throughout the world are continually capturing the voices of suspects with miniature transmitter/receiver systems, analog and digital on-the-body recorders, telephone intercept devices, and concealed room microphones. Since these recordings are normally utilized for investigative leads and/or legal proceedings, specific speakers must be accurately identified. Voice identifications that occur through self-recognition of one's voice, eye-witness information, surveillance logs, and the use of a person's name in the conversation are usually readily accepted. However, voice identifications that involve listening only and/or laboratory tests are often more difficult to evaluate accurately. To provide a better understanding of these voice comparison topics, two types of aural-only comparisons will be discussed, and an update on the spectrographic technique is included.

Aural Identification of Familiar Voices

Recognition of familiar voices is a daily occurrence for most people, as they identify spouses, children, coworkers, friends, and business associates after only a few words spoken over the telephone or by hearing them from an adjacent room. This process involves long-term memory, where recognition occurs through a prior knowledge of speech characteristics, including such attributes as accent, speech rate, pronunciation, pitching, vocabulary, and vocal variance (intraspeaker variability).

Some of the relevant scientific research and opinions that address the accuracy of identifying familiar voices include the following:

1. Researchers used 7 listeners who were familiar with the 16 chosen speakers through daily contact. The speakers had no pronounced speech defects or accents. Groups of two to eight speech samples of varying lengths were played back to the listeners, which resulted in an identification accuracy of better than 95% for samples lasting from about 1 to 2 seconds. Voice samples were also frequency restricted, but the results reflected only a limited loss of accuracy under conditions normally encountered in law enforcement investigations. In tests involving whispered speech, the duration had to be somewhat greater than three times longer than normal speech samples to obtain equivalent levels of identification (Pollack *et al.* 1954).
2. Sixteen listeners with no hearing losses, who had known the recorded 10 male coworkers for at least 2 years, were chosen. None of the 10 recorded individuals had either pronounced regional accents or speech abnormalities.

When the listeners heard sentences of less than 3 seconds duration from the 10 coworkers, their median accuracy rate of identification was 98% (range of 92% to 100%). When only a disyllable (e.g., mama) was spoken, the median accuracy rate dropped to 88% (range of 73% to 98%) (Bricker and Pruzansky 1966).

3. In a study of coworkers, recordings were made on different telephone lines of four women and seven men, each talking for 30 seconds to 1 minute on a neutral topic such as the weather. An additional recording was prepared of another male, who was relatively unfamiliar to most of the listeners. The recordings were arranged in a random order and played to 10 of the other coworkers, who were asked to identify the speakers. "All the listeners except one correctly identified all the 11 [coworkers]... The one listener who made an error... confused two speakers who were not well-known to him. Three of the 10 listeners knew [the eighth male, who was not a coworker], and correctly identified him. Of the remaining seven listeners, only two said that they could not recognize this speaker. Five listeners wrongly identified this speaker as..." another one of their coworkers. "It is worth noting that four of the five listeners who made the wrong identification were highly skilled, experienced phoneticians..." with doctoral degrees in the field (Ladefoged 1978). This experiment reflects a 100% identification rate for the coworkers' voices that were well-known to them and an overall average accuracy rate of 96% when the relatively unfamiliar voice was added.
4. Twenty-four individuals were asked to listen to speech samples of 24 coworkers (15 males and 9 females) whom they had known for several years and 4 speakers unknown to the listeners. The speech samples averaged about 30 seconds in length and contained at least 12 utterances of 2 to 4 words each. Listeners rated each coworker on a scale of very familiar to totally unfamiliar prior to the testing. They listened to the samples for as long as they wished and then rated their decisions as follows: (1) guessing, (2) fairly sure, or (3) very sure. Deleting the results of any voice rated totally unfamiliar to the listener, the results showed a 90.4% correct identification rate and 4.3% incorrect identification rate, with 5.3% who said they did not know the speaker. If the 5.3% are deleted, the correct identification rate is 95.4%. "This rate is probably fairly representative of situations where a limited vocabulary is required and can be expected to be even higher in informal conversations where more of the individual speaker's speech habits are present as cues for identification" (Schmidt-Nielson and Stern 1985).

5. In an introduction to another research paper, the author states that the "Identification of speakers by their voices is a common experience. Most listeners have little difficulty in identifying the voices of familiar speakers over the telephone or on the radio. Recognition of voices of familiar speakers in the darkness or when the speaker is out of sight of the listener is also a common occurrence" (Compton 1963).

This research reflects that the identification accuracy rate for familiar voice samples lasting 1 second or longer ranged from 92% to 100% and averaged 95% to 100%. Samples recorded through the telephone or other limited bandwidth systems had little effect on accuracy. The effects of noise and loss of high frequency information were studied in another experiment (Clarke *et al.* 1966) which found that aural speaker identification was only slightly degraded when progressing from high-quality voice samples to typical investigative recordings. It is obvious from everyday experience and the cited research that identifying familiar voices can be an accurate method for identifying voices recorded in forensic applications, even with the limiting factors of noise and attenuated high frequencies.

Aural Identification of Unfamiliar Voices

Aural comparisons of unfamiliar voice samples rely on short-term memory. For example, a woman receives a number of different telephone inquiries regarding a classified advertisement. She then receives an obscene telephone call, and she tries to remember if any of the voices match. In a judicial proceeding, a judge and/or a jury may have to decide if a particular crucial comment on an investigative recording was spoken by the defendant, who readily admits to saying the other statements attributed to him on the transcript, or to someone else involved in the conversation. Examiners using the spectrographic technique, described later, play back the separate voice samples concurrently on separate devices or computer files with an electronic patching arrangement to allow rapid aural switching between them or by recording short phrases or sentences from each sample on the same recording (Voice Comparison Standards 1991).

The *de facto* study of unfamiliar voice comparisons (Clarke *et al.* 1966) determined the following:

1. Sentence length over the range of 5 to 11 syllables is not an important variable in identification accuracy.
2. Correct identifications decreased from approximately 90% to 80% when the signal-to-noise ratio (SNR) was reduced from 30 decibels (dB) to 0 dB.
3. Correct identifications decreased from approximately 88% to 78% when the frequency response was reduced from 4,500 hertz (Hz) to 1000 Hz.

Since most investigative recordings have a SNR of 10 dB to 40 dB and a frequency response of 2,500 Hz to 5,000 Hz, the range of expected correct identifications of unfamiliar voices would be 78% to 90%, with most identifications in the 78% to 83% range.

The use of expert testimony for aural identifications of unfamiliar voices provides no assistance to the court and/or to the jury. The notes of the advisory committee on Rule 901 of the Federal Rules of Evidence appropriately reflect this fact as follows: "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..." (Federal Criminal Code and Rules 1991). Additionally, the voice comparison standards of the International Association for Identification (IAI) specifically state that it "... does not support or approve the use of... aural only expert decisions..." for voice comparisons (1991).

Spectrographic Comparisons

The spectrographic laboratory technique is the most well known and possibly the most accurate of the laboratory testing procedures presently available for comparing verbatim voice samples under forensic conditions. However, some scientists believe that aural identifications of very familiar voices are more accurate (Hecker 1971). The spectrographic technique has been described in numerous forensic and scientific publications, including an overview article published in the *Crime Laboratory Digest* (Koenig 1986). Therefore, a detailed explanation will not be rendered here; the following paragraphs provide a brief summary of the examination, a review of the new comprehensive standards passed by the IAI, and its status in government and private laboratories.

When properly conducted, spectrographic voice identification is a relatively accurate but not conclusive examination for comparing a recorded unknown voice sample with a suspect repeating the identical contextual information over the same type of transmission system (e.g., a local telephone line). The examiner uses both the short-term memory process previously detailed and a spectral pattern comparison between identically spoken sounds on spectrograms. Figures 1A and 1B are sound spectrograms of different male speakers saying "salt and pepper." The horizontal axis represents time, divided into 0.1-second intervals by the short vertical bars near the top, and the vertical axis is frequency, ranging linearly from 80 Hz to 4000 Hz, with horizontal lines every 1000 Hz. The speech energy is reflected in the gray scale from black (highest level) to white (lowest level). The frequency range of the voice is analogous to the range of a musical instrument, where the lowest notes are at the lowest frequency and the highest notes at the highest frequency. The mostly horizontal bands of darkness reflect the vocal resonances and are called formants. The closely spaced vertical striations represent fundamental frequency (voice pitch) or the actual vibrations of the vocal cords. The spectrographic technique requires comparison of identical phrases between the voice samples, with a decision made at one of a number of confidence levels. The scientific support of this examination is limited, and the actual error rate under most investigative conditions is unknown. The research to date indicates that the technique has a certain error rate that is independent of examiner-induced errors, with errors of false elimination (the voice samples were

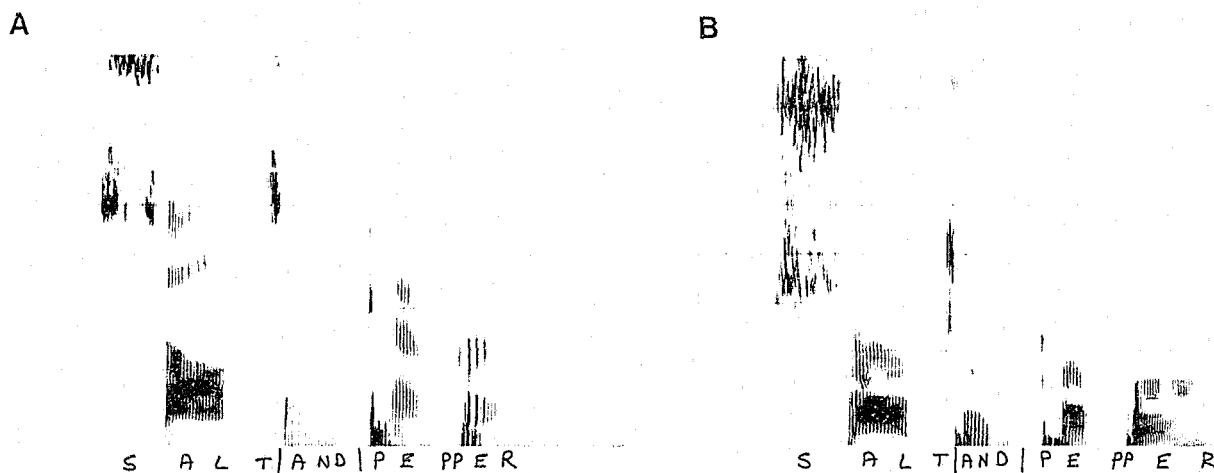


Figure 1 (A) and (B). Sound spectrograms of different male speakers saying "salt and pepper."

actually from the same person, but the examination found that they did not match) appreciably higher than false identification (the voice samples were actually from different persons, but the examination found that the samples matched).

In July 1991, the Voice Identification and Acoustic Analysis Subcommittee of the IAI passed and published its first set of comprehensive spectrographic voice identification standards. These requirements, which became effective January 1, 1992, for all certified IAI members, include examiner qualifications, evidence handling, preparation of exemplars, preparation of copies, preliminary examination, preparation of spectrograms, spectrographic/aural analysis, work notes, testimony, certification, and miscellaneous subjects. Table 1 lists the minimum qualifications for spectrographic examiners of the IAI and the FBI and updates a similar table published in an earlier issue of the *Crime Laboratory Digest* (Koenig 1986). Table 2 is another updated and expanded table from the same article concerning minimum criteria for spectrographic comparisons. Tables 1 and 2 and the previously published tables reflect that the upgraded IAI standards are now appreciably closer to the FBI's criteria. The FBI's standards require higher educational levels, more words for lower confidence decisions, enhancement proce-

dures when needed, and a higher frequency voice range. The most important legal difference is the FBI's policy not to provide testimony on spectrographic comparisons due to the inconclusive nature of the examination and the unknown error rate under specific investigative conditions.

The use of the spectrographic technique since the mid-1980s continues to show a steady decline by both government laboratories and private examiners. As of mid-1993, the New York City Police Department and the FBI were the only government laboratories in this country regularly conducting these examinations. The private sector efforts were limited to less than a dozen part-time examiners. Professional meetings in the field have been sparsely attended, and no major spectrographic research is known to be under way. Problems still persist in the spectrographic voice identification field. Examples of these problems include the following: (1) separate sets of certified examiners making high-confidence decisions for both identification and elimination in the same case;¹ (2) individuals with no experience, training, or education in the voice identification discipline making conclusive decisions under oath in court; and (3) examiners testifying that an unknown voice is not the defendant's, although admitting their decisions are really inconclusive based upon accepted standards.

Table 1. Minimum Qualifications for Spectrographic Examiners of the IAI and FBI

Qualification	IAI	FBI
Education	High School Diploma	BS Degree
Periodic Hearing Test	Yes	Yes
Length of Apprenticeship	Usually 2 Years	2 Years
Number of Comparisons Conducted	100	100
Attendance at a Spectrographic School	Yes	Yes
Formal Certification	Yes	Yes

Table 2. Minimum Criteria for Spectrographic Comparison for the IAI and FBI

Criteria	IAI	FBI
Words Needed for Highest Confidence Level	20	20
Words Needed for Lowest Confidence Level	10	20
Affirming Independent Second Decision	Yes	Yes
Original Recording Required	Yes	Yes
Allows Testimony	Yes	No
Completely Verbatim Known Samples	Usually	Usually
Speech Frequency Rate	Above 2 KHz	Above 2.5 KHz
Accuracy Statement in Report	Yes	Yes
Enhancement Procedures, When Needed	Optional	Yes
Speed Correction of All Recordings	Yes	Yes
Track Determination of All Recordings	Yes	Yes
Azimuth Alignment Correction	Yes	Yes

Summary and Conclusion

Under investigative conditions, individuals can reliably identify voices that are well known to them, but the accuracy rate drops to approximately 78% to 83% when unfamiliar voices are compared to known voice samples. The use of expert witnesses does not improve the accuracy rate of aural only voice comparisons. The use of the spectrographic technique continues to decline, even with the establishment of new standards in 1992.

References

- Bricker, P. D. and Pruzansky, S. Effects of stimulus content and duration on talker identification, *Journal of the Acoustical Society of America* (1966) 40:6:1441-1449.
- Clarke, F. R., Becker, R. W., and Nixon, J. C. *Characteristics that Determine Speaker Recognition*. Technical Report ESD-TR-66-636, Electronic Systems Division, US Air Force, 1966.
- Compton, A. J. Effects of filtering and vocal duration upon the identification of speakers, aurally, *Journal of the Acoustical Society of America* (1963) 35:11:1748-1752.
- Federal Criminal Code and Rules*. West, St. Paul, MN, 1991, p. 289.
- Hecker, M. H. L. *Speaker Recognition: An Interpretive Survey of the Literature*. American Speech and Hearing Association, Washington, DC, 1971.
- Koenig, B. E. Spectrographic voice identification, *Crime Laboratory Digest* (1986) 13:4:105-118.
- Ladefoged, P. Expectation affects identification by listening, *Language and Speech* (1978) 21:4:373-374.
- Pollack, I., Pickett, J. M., and Sumby, W. H. On the identification of speakers by voice, *Journal of the Acoustical Society of America* (1954) 26:3:403-406.
- Schmidt-Nielson, A. and Stern, K. R. Identification of known voices as a function of familiarity and narrow-band coding, *Journal of the Acoustical Society of America* (1985) 77:2:658-663.
- Voice comparison standards, *Journal of Forensic Identification* (1991) 41:5:373-392.

Note

1. Los Angeles Board of Civil Service Commissioners. Threat case decided March 25, 1992, in which three IAI examiners made an identification at a high-confidence level, while two IAI examiners eliminated the suspect.