

# Voiceprint Identification

**Money Laundering and Narcotics Update, Department of Justice, 1988, and  
The Legal Investigator 1990 by Steve Cain, Lonnie Smrkovski and Mindy  
Wilson**

Voiceprint identification can be defined as a combination of both aural (listening) and spectrographic (instrumental) comparison of one or more known voices with an unknown voice for the purpose of identification or elimination. Developed by Bell Laboratories in the late 1940s for military intelligence purposes, the modern-day forensic utilization of the technique did not start until the late 1960s following its adoption by the Michigan State Police. From 1967 until the present, more than 5,000 law enforcement related voice identification cases have been processed by certified voiceprint examiners.

Voice identification has been used in a variety of criminal cases, including murder, rape, extortion, drug smuggling, wagering-gambling investigations, political corruption, money-laundering, tax evasion, burglary, bomb threats, terrorist activities and organized crime activities. It is part of a larger forensic role known as acoustic analyses, which involves tape filtering and enhancement, tape authentication, gunshot acoustics, reconstruction of conversations and the analysis of any other questioned acoustic event.

## Theory

The fundamental theory for voice identification rests on the premise that every voice is individually characteristic enough to distinguish it from others through voiceprint analysis. There are two general factors involved in the process of human speech. The first factor in determining voice uniqueness lies in the sizes of the vocal cavities, such as the throat, nasal and oral cavities, and the shape, length and tension of the individual's vocal cords located in the larynx. The vocal cavities are resonators, much like organ pipes, which reinforce some of the overtones produced by the vocal cords, which produce formants or voiceprint bars. The likelihood that two people would have all their vocal cavities the same size and configuration and coupled identically appears very remote.

The second factor in determining voice uniqueness lies in the manner in which the articulators or muscles of speech are manipulated during speech. The articulators include the lips, teeth, tongue, soft palate and jaw muscles whose controlled interplay produces intelligible speech. Intelligible speech is developed by the random learning process of imitating others who are communicating. The likelihood that two people could develop identical use patterns of their articulators also appears very remote.

Therefore, the chance that two speakers would have identical vocal cavity dimensions and configurations coupled with identical articulator use patterns appears extremely remote. While there have been claims that several voices have been found to be indistinguishable, no evidence to support such allegations has been published, offered for examination or demonstrated to the authors.

Several studies have been published evidencing the ability to reliably identify voices under certain conditions, and a Federal Bureau of Investigation survey of its own performance in the examination of 2,000 forensic cases revealed an error rate of 0.31 percent for false identifications, and 0.53 per cent for false eliminations. (See Koenig, B.E., 1986, Spectrographic Voice Identification: a forensic survey, *Journal of the Acoustical Society of America*, 79:2088-2090.)

While there is disagreement in the so-called "scientific community" on the degree of accuracy with which examiners can identify speakers under all conditions, there is agreement that voices can, in fact, be identified.

To facilitate the visual comparisons of voices, a sound spectrograph is used to analyze the complex speech wave form into a pictorial display on what is referred to as a spectrogram. The spectrogram displays the speech signal with the time along the horizontal axis, frequency on the vertical axis, and relative amplitude indicated by the degree of gray shading on the display. The resonance of the speaker's voice is displayed in the form of vertical signal impressions or markings for consonant sounds, and horizontal bars or formants for vowel sounds. The visible configurations displayed are characteristic of the articulation involved for the speaker producing the words and phrases. The spectrograms serve as a permanent record of the words spoken and facilitate the visual comparison of similar words spoken between an unknown and a known speaker's voice.

#### Procedural Guidelines

The acoustic environment in many cases can be controlled at the receiving end of speech signal. Shutting off the radio, television or other signal- noise generating devices will reduce or eliminate unwanted background speech or noise. While not always possible, the investigator should attempt to select a reasonably quiet environment for controlled activities such as drug buys or other illegal operations being investigated. Many times these types of activities are carried out in bars, restaurants, car washes, billiard rooms and the like, and the investigator cannot always dictate the location.

It may require the recording of telephone conversations or face-to-face encounters under a variety of acoustic conditions in which someone is wearing a body recorder or transmitting the conversation via radio frequency to a remote location. Unfortunately, in many cases the investigators cannot control the acoustic environment. In situations involving an adverse environment,

investigators should use high technology stereo equipment to optimize recording capability.

The attempt to produce samples as parallel to the unknown as possible actually assists the examiner in his task because speaker variables are reduced to a minimum. Numerous studies have been conducted that indicate very reliable decisions can be made by trained professional examiners when samples are obtained in the manner described.

The notion proposed by some opponents that duplicating the unknown as closely as possible may cause error is not supported by any available evidence. Research studies have produced strong evidence that even very good mimics cannot duplicate another's speech patterns.

In an attempt to obtain proper speech samples, investigators should not hesitate to ask suspects for the samples they need. Surprisingly, many suspects will voluntarily give a sample of their voice for comparison purposes.

In the event you are dealing with some type of vocal disguise, attempt to obtain a similarly produced known exemplar in addition to the suspect's normal voice. It should be noted that vocal disguises can be very difficult for the examiner to deal with and the probability of determination is less than with normal voice samples.

If a suspect refuses to cooperate with the investigator, a court order may be acquired compelling the suspect to produce voice recordings for the purpose of comparison. Courts have repeatedly held that requiring the accused to submit voice exemplars for the purpose of comparison for identification or elimination does not violate the suspect's Fifth Amendment rights. In *Wade*, 388 U.S. 218 (1967), the Court held that the privilege against self-incrimination offers no protection from compulsion to submit to speaking for purpose of voice identification, or to writing, photographing, finger-printing and measurements.

Several problems have been encountered in obtaining known voice exemplars even with the use of a court order. If the court order is vague, the suspect may utter a few words of the text involved, speak too softly, too fast, or too slowly, or otherwise disguise the sample and claim compliance with the order.

To prevent such problems, the investigator is wise to request that the court order specify in detail, that the suspect give a sample of his or her voice, repeating the phrases of the questioned call in a natural conversational voice (or in a similar disguise, if that is the case) and that such sample shall be given at least three times and to the reasonable satisfaction of the investigator. Voice exemplars obtained with such specific instructions are usually very satisfactory for comparison purposes.

Before terminating the recording session, check the recording to determine whether or not a satisfactory exemplar was obtained.' Remember that once a suspect is released, a second known sample may be very difficult to obtain.

Whatever the recording circumstance, background noise and the distance between the talker and the receiving device should be minimized for optimal recording. Good quality tape recording equipment should be used, as well as magnetic recording tape. As a rule of thumb, recording tape with standard 120 equalization, normal bias and no more than a 5 dB drop at 6 KHz should be used.

After the development of a suspect, the next task is to properly obtain known voice samples for comparison purposes. Do not hesitate to ask a suspect for a speech sample. If the suspect refuses, a court order may be obtained requiring compliance with the request. See *Schmerber v. California*, 384 US. 757(1966). and *Gilbert v. California*, 388 US. 263 (1967). Both are landmark cases. There are also many additional decisions at both state and federal court levels that may be cited to support such a request. Court orders should clearly spell out the minimum number of samples to be obtained, the manner of speech, and the method to be employed.

The next task for the investigator is to obtain proper speech samples for comparison purposes. Probably the best guide here is attempting to duplicate the recording of the questioned call. Known samples should be obtained via the telephone and recorded in the same manner as the questioned call. If possible, the same recorder and telephone pickup should be used. In some cases, even the same telephone has been employed. If there is room on the questioned tape, the known sample may be placed on it. If there is not, another tape of the same type and brand should be used if at all possible.

Speech samples obtained should contain exactly the same words and phrases as those in the questioned sample because only like speech sounds are used for comparison. Because the voice, like handwriting, is dynamic and variant, several samples of each spoken phrase are desired for analysis. Unless the questioned call sounds like a read statement, the suspect should not be allowed to read the phrases from a transcript but should repeat each phrase after it is spoken by someone else. To avoid an unnatural verbal response, the suspect should repeat the first phrase and proceed in the same manner with each successive phrase.

When all phrases have been recorded, the same procedure should be repeated at least two more times beginning with the first word or phrase. The suspect may be asked to read the phrases if a very poor job of repeating is done. Some people do a better job of reading than repeating the phrases.

It is important that the known sample be spoken in the same manner as the questioned sample; therefore, the investigator should be familiar with the voice,

manner of speech and the text. If the caller's voice was disguised, the suspect should give a normal sample and a disguised one as in the questioned call.

Recorded evidence should be wrapped in tinfoil to protect it from possible contact with a magnetic field if it is submitted by mail. The evidence should be shipped in a secure container that will prevent the evidence from tearing through the packaging material. Do not submit a copy of your investigative report with the evidence. The examiner does not want to know the details of the case. It is important, however, to provide the examiner with information regarding the recording method, the number of calls and suspects involved, and any other information that may assist the examiner in the examination of the evidence.

Upon receipt of the evidence by the laboratory, it is properly marked and a case number is assigned. The analysis and comparison of known and questioned voice samples may take several hours or days to complete, depending on the number of samples involved and the complexity of the examination. Both an aural (listening) and visual (spectrographic) examination and comparison is conducted. Aural and spectrographic cues examined should compliment one another in the event the voices are in fact the same.

As with the identification of finger prints, there is presently no universal standard for the number of words required for identification. It does, however, vary from a minimum of 10 for some agencies and 20 for others. The Internal Revenue Service has chosen to use 20 or more like speech sounds between an unknown and known sample with the degree of certainty based on quality and excellence of the evidence examined. Obtaining a second, independent decision is standard practice in this field as in other forensic sciences.

Visual comparison of spectrograms involves, in general, the examination of spectrograph features of like sounds as portrayed in spectrograms in terms of time, frequency and amplitude. Specific features, the result of producing consonants, vowels and semi-vowels in isolation or in combination (coarticulation), include the following but certainly not all-inclusive clues: pitch, bandwidth, mean frequency, trajectory of vowel formants, distribution of formant energy, nasal resonance, stops, plosives, fricatives, pauses, inter formant features and other idiosyncratic and pathological features.

Special aural comparison tapes are prepared facilitating comparison of psycholinguistic features via short-term memory. Aural cues compared include resonance quality, pitch, temporal factors, inflection, dialect, articulation, syllable grouping, breath pattern, disguise, pathologies and other peculiar speech characteristics.

Some agencies offer court testimony, others do not. The IRS laboratory is the only federal agency that presently offers testimony. All other certified examiners, whether in state agencies or in private practice, also offer court testimony.

## Court Admissibility

Court testimony involving aural- spectrographic voice comparison essentially started having an impact on the courts after the Tosi Study in December 1970. Since then there have been between 150 and 200 trials in local, state or federal courts. Because of a difference based on evidentiary philosophical reasons, some courts have admitted aural-spectrographic voice evidence and others have not.

There are two general "rules" or "standards" by which scientific evidence is accepted in courts of law in the United States. The first, commonly referred to as the Frye "rule" or "test," is based on a 1923 District of Columbia case and basically requires "general acceptance in the particular field in which it belongs." See *Frye v. United States*, 54 App. D.C. 46, 293 F. 1013 (1923). The second is based on the argument of McCormick (See "McCormick on Evidence," 3rd Ed., 203 at 608.) McCormick states: "General scientific acceptance is a proper condition for taking judicial notice of scientific facts, but it is not a suitable criterion for the admissibility of scientific evidence. Any relevant conclusion supported by a qualified expert witness should be received unless there are distinct reasons for exclusion." See Rule 702 of the Federal Rules of Evidence.

Many state and federal courts have abandoned Frye and adopted the argument of McCormick. The supreme courts of Minnesota, Maine, Ohio and Rhode Island have admitted aural-spectrographic voice evidence following McCormick. Intermediate appellate courts in California, Maryland and Michigan admitted such evidence following Frye but were reversed by their respective supreme courts, which held that the Frye test had not been met. The Massachusetts Supreme Court held aural-spectrographic voice evidence admissible applying the Frye test, while those of Arizona, Indiana and Pennsylvania did not.

In the federal court system, we are aware of 30 trials in which the question of aural-spectrographic voice evidence was addressed. All but three admitted the evidence based on Frye or McCormick. On appeal, the Second, Fourth and Sixth Circuits held the evidence admissible, applying McCormick, while the District of Columbia did not, applying Frye. See *United States v. Williams*, 583 F.2d 1194 (2d Cir.), cert. denied 439 US.

1117 (1978); *United States v. Bailer*, 519 F.2d 463 (4th Cir.), cert. denied

423 US. 1019 (1975); *United States v. Franks*, 511 F.2d 25 (6th Cir.) cert. denied 422 US. 1042 (1975), and *United States v. McDaniel*, 538 F.2d 408 (D.C. Cir. 1976).

In *United States v. Williams*, supra at 1198, the court said: "The 'Frye' test is usually construed as necessitating a survey and categorization of the subjective views of a number of scientists, assuring thereby a reserve of experts available

to testify. Difficulty in applying the 'Frye' test has led a number of courts to its implicit modification." Also see United States v. Bailer, supra at n.6.

Since 1970, the forensic application of aural-spectrographic voice identification has been reliably applied in the investigation of several thousand cases. While there is disagreement on the reliability of the method under all conditions, there is agreement that voices can be identified and eliminated when the proper conditions exist and the analysis is carefully conducted by qualified examiners.

Several state appellate and supreme courts have admitted the evidence, as have three of four federal appellate courts. The United States Supreme Court has refused to review and decide the three cases brought before it. While the admission of aural-spectrographic voice evidence continues to be decided in various courts, the method continues to be a very important tool in the arsenal against crime.

Other areas of acoustic analysis include, in part, gun shot analysis, tape enhancement and tape authentication. While not discussed in this article, it should be noted that laboratory analysis related to these problems is available in some laboratories.

NDAABulletin December 1993