COMMENTS

The following comments were written by students at Northwestern University School of Law. Contributors to the present issue are Philip E. Cutler, Carol R. Thigpen, Timothy R. Young, and Edward B. Mueller.

THE EVIDENTIARY VALUE OF SPECTROGRAPHIC VOICE IDENTIFICATION

Ever since it was heralded in 1962 as a foolproof method of personal identification, voice identification by spectrographic analysis, the “voiceprint” technique, has been in a legal limbo. Scientists could not agree whether the technique possessed any validity at all, or agree on who was competent to pass judgment on that threshold question. Because a judicial determination of the admissibility of scientific evidence and the expert testimony necessary to interpret it is traditionally keyed to scientific acceptance of the principle from which the testimony is derived, admissibility was withheld. Recent developments in both science and the law, however, indicate that despite initially adverse scientific and judicial reaction, spectrographic voice identification is perhaps coming of legal age. This comment will assess the impact of these developments on the technique’s status as admissible evidence.

THE TECHNIQUE

Speaker recognition by spectrographic voice analysis is a seemingly simple, but fundamentally complex, method of personal identification. In making an identification by this method, the

The constitutional ramifications of the technique are excluded from the scope of this comment. They are, however, quite real and as yet undefined. The Seventh Circuit, for example, has held that a grand jury may not subpoena a number of witnesses, solely for the purpose of obtaining voice exemplars (apparently for spectrographic comparison), without complying with fourth amendment “reasonableness” requirements. In re Dionisio, 442 F. 2d 276 (7th Cir. 1971), cert. granted, 92 S.Ct. 2056 (1972). Serious questions of “reasonableness” also arise in the method used to obtain the voice exemplar. See, e.g., People v. King, 266 Cal. App. 2d 437, 72 Cal. Rptr. 478 (1968); State ex rel. Trimble v. Hedman, ___ Minn. ___, 192 N.W.2d 432 (1971).

Currently, there are several methods of personal identification which purport to have forensic value. Fingerprints, sole prints, and palm prints are, of course, the most widely recognized. See Inbau, Scientific Evidence in Criminal Cases, 25 J. CRIM. L. & C. 500 (1934). Two rather novel methods of personal identification have been suggested recently: Hirachi, Identification of Earprints, 24 Kriminalistik 75 (1970); Suzuki & Tsuchihashi, Personal Identification by Means of Lip Prints, 17 J. FOR. MED. 52 (1970). It was suggested at one time that electrocardiograms could provide a reliable means of identification, though the author recognized its limited forensic use. Castellanos, Personal Identification by Electrocardiography, 23 J. CRIM. L. & C. 356 (1932).

There are three known methods of speaker recognition. The oldest method, of course, is identification by ear. Another method is computer analysis of speech spectrograms. The latter technique is identical to the subjective comparative analysis of speech spectrograms under consideration here, except that a computer is substituted for the human examiner. At present, this method, although potentially more reliable because
first step is to tape-record an exemplar of an individual’s voice. The sound spectrum of this speech sample is then scanned electronically by a high-speed sound spectrograph9 which produces a spectrogram, a visible amplitude-frequency-time display of the speech sounds recorded. This visible portrayal of the frequency variations in an individual’s voice can then be subjectively compared with spectrograms of phonetically identical sounds produced by “unknown” individuals. The sound patterns represented on the spectrogram are the product of the energy expelled during speech and are shaped and determined by the dynamic interplay between the individual’s vocal mechanism10 and the coupling and placement of his articulators.11 The validity of the technique as a means of personal identification rests on the premise that the sound patterns produced in speech are unique to the individual and that the spectrogram accurately and sufficiently displays this uniqueness.

The spectrogram is popularly referred to as a voiceprint. See Figure 1. Frequency is measured along the vertical axis, generally using a logarithmic scale to portray in greater detail the frequency ranges where speaker-dependent identity traits are thought to be manifested. HECKER, SPEAKER RECOGNITION at 40, 50–56, 62. Time is depicted along the horizontal axis. The “bar” type of spectrogram, which depicts the frequencies produced by energy vibrations in the vocal tract as vertical striations and those of phonetic formants as curved horizontal bars, displays amplitude in

---

9 The sound spectrograph electronically sorts out the frequency components in the sound spectrum of speech and mechanically produces a visual representation of those spectral characteristics as they change over time. For a technical discussion of the operating characteristics of the sound spectrograph see HECKER, SPEAKER RECOGNITION at 50–56; Presti, HIGH-SPEED SOUND SPECTROGRAPH, 40 J. ACOUSTICAL SOC’Y AM. 628 (1966). See generally Koenig, Dunn & Lacy, THE SOUND SPECTROGRAPH, 18 J. ACOUSTICAL SOC’Y AM. 19 (1946); Steinberg & French, THE PORTRAYAL OF VISIBLE SPEECH, 18 J. ACOUSTICAL SOC’Y AM. 4 (1946).

10 The vocal mechanism consists of the pharyngeal cavity, vocal folds, nasal cavity, tongue mass, soft palate, and oral (mouth) cavity. HECKER, SPEAKER RECOGNITION at 7. Organic differences in the structure of the vocal mechanism may be due to heredity, sex, and age. Id. at 4.

11 The placement of the articulators, the tongue, lips, teeth, jaw muscles and soft palate, is an ability learned through imitation. Speech production is thus related to geographic, social, and cultural factors. Id.; Bolt, RELIABILITY at 598–99.
KERSTA: THEORY, EXPERIMENTS AND REPLICATION

Lawrence Kersta, an electrical engineer and physicist, has been the foremost advocate of the validity of personal identification by spectrographic voice analysis. His claim that the technique is accurate and reliable is founded on two propositions.

The theory of invariant speech is the cornerstone of his hypothesis that individuals can be identified by the spectral characteristics of their voices. The theory posits that the characteristic spectral patterns of phonetically identical utterances vary more between two individuals (interspeaker variability) than between two such utterances spoken by the same individual (intraspeaker variability). Although so far the theory has not been proven directly, Kersta has buttressed the theory by applying his own hybrid form of statistical probability to the acoustic theory of speech production. He argues that since both the dimensions of the vocal cavities and the coupling of the articulators, which define the spectrum for a given sound, are affected by heredity, sex, age, and socio-environmental factors, it is extremely unlikely that two individuals would develop spectrographically identical speech patterns. While a superficially attractive rationale for voice uniqueness, a proper application of probability theory demands substantially more precision than this.

Kersta also asserts that a trained examiner, by visually comparing a series of spectrograms for objective points of similarity, can, because of the speech pattern uniqueness reflected on the spectrogram, identify or eliminate one unknown speaker from a group of many known speakers.

Kersta’s initial experiments with spectrographic voice analysis demonstrated that the technique showed potential as a means of personal identification, but did not present an identification task. His experiments merely tested the ability of a trained examining team to match spectrograms with speakers in a closed set of exemplars. Using speaker groups of up to twelve members, error

---

2 See note 1 supra. Although Kersta attracted considerable attention with his claims to “voiceprint” infallibility, there is evidence that he was not the first to advance this theory. See Tosi, et al., The Current Status (1971) of Voice Identification, October 18, 1971 (unpublished paper presented at the 82d meeting of the Acoustical Society of America). Kersta is now president of Voiceprint Laboratories, Somerville, N.J. His company manufactures a sound spectrograph and offers consulting services to law enforcement agencies.

13 Although the individual spectral characteristics in the speech signal do not remain invariant, the theory presumes that speech habits, once formed, vary more between two individuals pronouncing the same sound than between two pronunciations of the same sound by one individual. An excellent technical discussion of the acoustic speech theory can be found in G. Fant, Acoustic Theory of Speech Production (1960) and P. Ladefoged, Elements of Acoustic Phonetics (1962).

14 But cf. text accompanying notes 72–73 infra.


16 See note 13 supra.

rates in these trials ranged from 0.8% for words spoken in isolation to 1.0% for words excerpts from context. In subsequent experiments performed under Kersta’s direction, the number of speakers represented by the set of spectrograms presented to the examiners was increased, but the matching format of the experiment was adhered to. These experiments confirmed his initial findings.

From these extremely low error rates, Kersta concluded that spectrograms of an individual’s speech patterns for particular words are as unique in their identifying characteristics as fingerprints, thus rendering the technique a reliable method of personal identification when performed by a trained examiner. In addition, Kersta claimed that neither passage of time nor conscious efforts at mimicry could frustrate a system of identification based on spectrographic voice analysis. He further maintained that the relatively higher pitch of the female voice would not affect the accuracy of such an identification technique.

Other experimenters, all reputable scientists in speech, phonetics, or associated fields, were unable to duplicate Kersta’s high accuracy rates, even for matching-to-sample tests. It appears that most of the difficulty other experimenters experienced can be traced to Kersta’s failure to publish detailed technical reports of his experiments and the objective identification factors used by his examiners. Kersta’s secrecy, coupled with his defensive attitude toward the technique and the extrapolations he drew from his limited experiments, created an atmosphere of distrust for the entire technique which may have obscured the real issues involved in gaining scientific acceptance for the technique. However, a study completed in 1970 under the direction of Oscar Tosi at Michigan State University not only replicated Kersta’s original experiments and confirmed his high accuracy rates for matching-to-sample tests but also tested the accuracy and reliability of spectrographic voice identification in experiments having a true identification format.

THE TOSI STUDY: FORMAT AND RESULTS

Since in the matching-to-sample tests conducted by Kersta and others a match for the “unknown” spectrogram always existed, the trial became merely a process of elimination. This type of trial has no relation to a forensic application of the technique and the results obtained through such trials cannot be extrapolated to validate the technique as a means of identification. The format of the Tosi study, however, was designed to test varying conditions which could be expected to fail.

& Woods, Speaker Authentication and Identification, 44 J. Acoustical Soc’y Am. 1396 (1968). Hennessy & Romig, supra, however, conclude that while the results of these attempted replications are in sharp contrast with Kersta’s near-perfect identification scores, and thus do raise questions as to the validity of the technique, the lack of training given the examiners, the lack of uniformity in the objective points of similarity which the examiners were instructed to look for, and the use of a different spectrograph all may have seriously impaired the validity of the test results.

The Tosi study was conducted under a grant from the United States Department of Justice, Law Enforcement Assistance Administration, to the Michigan State Police. The results are summarized and explained in Tosi, et al., An Experiment on Voice Identification: Excerpts from the Report SHSLR 171 (Michigan State University, July 1971) [hereinafter cited as Tosi, Experiment]. See also Michigan Department of State Police, Voice Identification Research: A Summary of the Report to the U.S. Dept. of Justice Law Enforcement Assistance Administration (Grant No. N1 70-004, Feb. 1971).

Since a matching-to-sample test utilizes only closed sets, supra note 21, only one type of error existed: false identification. Error rates averaged 1%, substantially the same as those experienced by Kersta. Tosi, Experiment at 19.

Hecker sets forth the major variables inherent in speaker recognition by subjective analysis of speech
have a major impact on the reliability of the technique in a forensic setting.

First, both open and closed trials\(^{38}\) were conducted and examiner-awareness of the nature of the trial was recorded. Because the examiner would never know whether the author of the "unknown" spectrogram was also included in the "known" speaker population in an objective forensic application, the importance of this variable to the technique is obvious.

Second, the effect of a reduction in cue material on identification accuracy rates was tested by presenting the examiners with nine cue-word spectrograms\(^{34}\) for each speaker in the first phase of the study, but permitting only six cue-word spectrograms for each speaker in the second phase.\(^{36}\)

A third very important feature of the Tosi study was its use of both contemporary and non-contemporary matching spectrograms in testing speaker identification.\(^{36}\) Since speech spectrography has shown that the same person rarely utters the same sound in exactly the same way,\(^{37}\) this feature of the study attempted to determine the effect of time-lapse\(^{38}\) on the ability of the examiners to make positive identification or elimination.

spectrograms. Heckier, Speaker Recognition at 56–55. See also Bolt, Reliability.

\(^{38}\) In an open trial, the speaker population against which the unknown spectrogram must be compared may or may not contain the author of the known spectrogram. In a closed trial, on the other hand, a matching spectrogram is always contained in the speaker population.

\(^{34}\) The cue-words used were: it, is, on, you, and, the, I, to, and me. A sufficient number of different sounds must be given to the examiner to enable him to judge the spectral patterns of the individual to be identified.

\(^{35}\) Overall results of the second phase did not differ significantly from those experienced in the first. Tosi, Experiment at 17–18. However, Tosi noted that increasing examiner proficiency very likely contributed to this result.

\(^{36}\) In certain of the trials, the match for the unknown speaker's spectrogram represented a spectrogram made from an exemplar recorded contemporaneously with that representing the "unknown". In other trials the match was a non-contemporary spectrogram, i.e., the exemplar from which the matching spectrogram was made was recorded by the same person but at a later time.

\(^{37}\) This aspect of speech is known as intraspeaker variability. The sources of intraspeaker variability are not well known, but experimental evidence suggests that aging, disease, and psychological stress may affect a person's speech production and consequently the spectral characteristics of his speech patterns. Heckier, Speaker Recognition at 16–18, 69–70. See also notes 91–92 infra. An obvious contributor to intraspeaker variability is the environmental factor which shapes the articulation of speech. See note 11 supra.

\(^{38}\) Only non-contemporary matching spectrograms are used in a forensic application of the technique.

A fourth variable tested was the effect of the context of the cue material and its mode of recording on the reliability of the technique. The cue-words from which both the "known" and "unknown" spectrograms were made were spoken first in isolation, then in a fixed context, and finally in a random context.\(^{39}\) In addition, the cue-words were recorded in three different ways: in a quiet environment directly into a tape recorder, over a telephone line in a quiet environment, and over a telephone line in a noisy environment.\(^{40}\)

In selecting a speaker population of two hundred fifty males, drawn from a population of twenty-five thousand at the university, Tosi attempted to meet one of the requirements for validation of the technique: homogeneity of the speaker group.\(^{41}\) The speakers selected had no speech defects.

The time-lapse between obtaining the contemporary spectrogram used as the "unknown" and the non-contemporary matching spectrogram used as the "known" (the match) was one month. Tosi, Experiment at 7, 8.

\(^{39}\) The "fixed context" refers to identical sentences from which the desired sound is excerpted. The "random context" refers to different sentences from which an identical sound is excerpted. The fixed context would thus test identification accuracy where an accused reads from a transcript of an unknown speaker's conversation; the random context trial would test accuracy in those situations where a voice exemplar of the exact words of the transcribed sample could not be obtained. Compare State v. Cary, 49 N.J. 343, 230 A.2d 384 (1967); State v. McKenna, 94 N.J. Super. 71, 226 A.2d 755 (Essex County Ct. 1967); with People v. King, 226 Cal. App. 2d 437, 72 Cal. Rptr. 478 (1963); State ex rel. Trimble v. Hedman, __Minn__, 192 N.W.2d 432 (1971).

\(^{40}\) Frequency and noise distortion can impair the reliability of spectrographic speaker recognition. Since forensic applications would involve varied environments, it was important to measure the effect of frequency and noise distortion on the technique's reliability. Heckier, Speaker Recognition at 61–62. In the Tosi study, fifty decibels of white noise, measured at the head of the speaker, was used for distortion purposes. Tosi, Experiment at 6.

The greatest applicability of spectrographic voice identification in the criminal area is to crimes in which the telephone is used. See, e.g., In re Dionisi, 442 F.2d 276 (7th Cir.), cert. granted, 92 S.Ct. 2056 (1972) (interstate transmission of gambling information); United States v. Phoneix, No. 70-CR-428 (S.D. Ind. April 15, 1971) (bomb threat); United States v. Wright, 17 U.S.C.M.A. 183, 37 C.M.R. 447 (1967) (obscene telephone call); State v. Alea, No. 70-9397 (Dade County, Fla., Crim. Ct. Sept. 18, 1971), appeal docketed, No. 71-1419, Fla. App., Dec. 21, 1971 (extortion); State ex rel. Trimble v. Hedman, __Minn__, 192 N.W.2d 432 (1971) (enforcement of police officer into ambulance). Hecker, Speaker Recognition at 57. See also Bolt, Reliability at 601–02, 610–12. Homogeneity of the speaker population is an important factor in determining the validity of any method of speaker identification because interspeaker variability in speech can be expected to be lowest among members of ethnic, racial,
fects and utilized a standard American English dialect.42

Each of the twenty-nine examiners used in the experiment was given one month of training in basic acoustic speech principles and in the interpretation of speech spectrograms. Moreover, the examiners were given several objective points of similarity to look for when making comparisons between spectrograms.43 This training was far more extensive than that given examiners in previous speaker recognition experiments utilizing spectrograms.44

In each of the nearly thirty-five thousand random trials,45 aural comparison of the speech samples was prohibited. The examiner was forced to come to a positive conclusion, either rejecting or accepting one of the “known” spectrograms as identical with the “unknown”, and an average time of only fifteen minutes was devoted to study of the spectrograms before a conclusion was demanded. These inhibiting factors, while necessary as a control in the experiment, would not be present in a forensic application of the technique.46

Regional, dialectic or similar groups, while intraspeaker variability does not vary appreciably between homogeneous groups. HECKER, SPEAKER RECOGNITION at 16. The speakers in the Tosi study were selected because they exhibited no special regional accent or other identifying characteristic. However, “homogeneity” of this speaker group was achieved on the basis of aurally perceptual similarity in speech patterns, whereas the crucial variable in spectrographic speaker recognition is the similarity in appearance of spectrograms. Unfortunately, very little is known of the perceptual and physical correlates of this type of speaker homogeneity. Id. at 57. The impact of this variable on the Tosi study as a general validation of the technique is discussed infra, notes 74–87 and accompanying text.

Tosi, EXPERIMENT at 8.

Although Kersta has never published a report detailing exactly which characteristics he uses to determine speaker identity, Tosi became familiar with Kersta’s methods while attending a course at Kersta’s Voiceprint Laboratories. Tosi instructed his examiners to consider the following objective points of similarity: similar mean frequencies of vowel formants, formants’ bandwidths, gaps and types of vertical striations, slopes of formants, durations, and characteristic patterns of fricatives and interformants’ energeries. These factors were selected because they tend to be speaker-dependent (interspeaker variability), rather than a function of an individual’s own variation in speech patterns (intraspeaker variability). Tosi, EXPERIMENT at 9. See generally G. FANT, supra note 13, at 20; Kamine, supra note 3, at 216.

44 See Hemmery & Romig, supra note 20, at 187–97.

45 "Power of 'known' speakers in each trial varied from ten to forty.

46 In a forensic application of spectrographic voice identification the examiner would make an initial comparison by ear to select the speech sounds for spectrographic comparison. In addition, he would be permitted to take as much time as necessary to reach a conclusion. Furthermore, he would have three, not two, general decision options: identical spectrograms indicating identity of speakers, non-identical, and insufficient information to reach a conclusion. Tosi, EXPERIMENT at 21–22.

47 Id. at 19–20.

48 However, the existence of either type of error demonstrates the degree to which intraspeaker and interspeaker variability interact to confuse the examiner. This, of course, highlights the examiner-dependent nature of the accuracy of spectrographic voice identification.

Tosi, EXPERIMENT at 20. Although the decision of elimination or identification in each trial was arrived at jointly by an examining panel, Tosi had each member rate his degree of confidence in the decision on a scale of one to four. In examining the relationship of this confidence-rating to errors, Tosi found that the less confident the examiner, the more likely that his decision was in error. Tosi calculated that had his examiners been permitted to express no opinion in those cases where they were uncertain as to the correctness of their decision, the gross error rate would fall to 7%. Of this, 2% would be attributable to false identification; 5% to false elimination. Id.

49 The first forensic use of spectrographic voice analysis apparently took place in a New York perjury prosecution. People v. Strahle, Crim. No. 9525/64 (Westchester Co., N.Y., Sup. Ct., 1966). See NOTE, VOICEPRINT METHOD OF IDENTIFICATION—RESTATEMENT OF THE COURTS TOWARD ACCEPTANCE OF SCIENTIFIC EVIDENCE, 12 N.Y.L.F. 50I (1966). The admissibility question, however, never reached the appellate level because the jury, unable to determine the weight to be accorded such evidence, failed to reach a verdict.
King the principal objection expressed to the admission of such evidence was that experimentation had not yet progressed to the point where the technique's reliability could be sufficiently demonstrated. Shortly before the King decision, however, a divided United States Court of Military Appeals in United States v. Wright ruled expert testimony regarding a spectrographic voice identification admissible, relying solely on Kersta's testimony as to the technique's reliability.

**THE FREY RULE**

As with all other types of evidence, relevance is the principal determinant in the admissibility of scientific evidence. The relevance of proffered testimonial or tangible evidence is determined by a judicial consideration of both the materiality and probativeness of that which is offered. In applying these evidentiary rules to scientific evidence, including the expert testimony necessary to interpret it, courts have usually relied on scientific experts to initially approve the worth and validity of a technique, thereby assuring the court that the proffered evidence possesses at least a minimum probative value. The general rule was first set forth by the District of Columbia Circuit in Frye v. United States, where the court stated:

> [While courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.]

The “general acceptance” test imposed by Frye has troubled many commentators in the years since its inception. The major argument against Frye is that a too literal application of the rule would unjustifiably tend to restrict the acceptance of new scientific techniques as admissible evidence. However, while Frye remains the foundation upon which courts base admissibility determinations in the field of scientific evidence, several modern courts have interpreted Frye's mandate in a decidedly progressive vein, attending to the spirit and policy of Frye rather than to the letter of its rule.

In attempting to refine somewhat the “general acceptance” rule of Frye, a California appellate court in People v. Williams, stated:

> 293 F. at 1013–24 (emphasis added). With regard to expert testimony based on scientific evidence, Wigmore stated his own version of the Frye rule:

> When the testimony thus appearing to the ordinary layman to lack a rational basis is founded on observations made with esoteric methods or apparatus . . . the method should be explained by the witness, and if it be vouched for in his branch of learning, it suffices to admit his testimony.

2 Wigmore § 659.

60 See, e.g., McCormick § 170; J. Richardson, Modern Scientific Evidence § 6.18 (1961); Note, supra note 50, McCormick feels that a “general acceptance” test is appropriate only for a court’s taking judicial notice of scientific facts. He would not impose such a test for determining admissibility. McCormick § 170.


The results of tests of the type here under attack, as well as opinions based thereon, are admissible only if the tests have gained acceptance in the field of learning in which they are in use.

Although noting that the expert witnesses had all acknowledged that it could not "truthfully be said that the medical profession as a whole" had generally accepted the test in question, the court concluded that:

All of the medical testimony points to the reliability of the test. It has been generally accepted by those who would be expected to be familiar with its use. In this age of specialization more should not be required.

A further refinement of Frye was undertaken by the court in Coppolino v. State. The admissibility of expert testimony relating to a novel and specially designed medical test for the presence of a certain highly toxic poison in the deceased was at issue. The court noted that prior to the discovery and use of the test by the state's expert, medical scientists believed it impossible to detect either the presence of the toxin or its component parts in the body. However, in upholding the trial judge's ruling admitting the testimony of experts who testified to the abnormally high presence of succinic acid, a component of the toxin, the court noted the Frye principle and then stated:

The problem presented to the trial judge was, were the scientific tests performed by Umbarger so unreliable and scientifically unacceptable that their admission into evidence was error.

Coppolino thus appears to hold that if the offered evidence or testimony is based on a test or technique which is analytically and scientifically valid, then notwithstanding the relative newness of the test or technique and lack of exposure to the profession, evidence derived from it may be admitted. However, the implicit premise of Frye, Williams, and Coppolino is that judicial and scientific acceptance of a novel technique is also dependent on the capacity of the test or experiment to produce the offered results or extrapolations.

THE TOSI STUDY AS A VALIDATION OF THE TECHNIQUE

"Scientific Acceptance"

Completion of the Tosi study and publication of the results removed a major scientific objection to the forensic use of spectrographic voice identification: the total lack of large-scale experiments investigating the technique's reliability in circumstances similar to those found in court cases. In addition, the methodology employed by Tosi gave a certain statistically replicable character to his results. But the admissibility requirements of Frye, Williams, and Coppolino go beyond mere statistical replicability. For the Tosi study to have any positive effect on the legal status of spectrographic voice identification, the results of the study must be of a kind which permit scientific extrapolation to validate the technique as a general method of personal identification. It is Tosi's contention that the results achieved by using his relatively homogeneous speaker population can be extrapolated to so validate the technique.

Recognizing that the present state of scientific knowledge was inadequate to prove directly the theory of invariant speech upon which the validity of spectrographic voice identification depends, Tosi felt nevertheless that such proof could be obtained inferentially from a proper evaluation of empirical data derived from a well constructed speaker identification experiment. However, the validity


Ladefoged has expressed the opinion that the Tosi study was a model piece of scientific methodology, "well designed and carried out with true scientific objectivity." Letter, supra note 2. See authorities cited in note 32 supra.


Tosi, Experiment at 2. Becker noted in his survey.
of such an evaluation, and consequently the admissibility of evidence derived from use of the technique, is itself dependent upon the sampling and testing format of the experiment.\textsuperscript{78} Furthermore, Tosi's conclusion that the validity and reliability of spectrographic voice identification has been proven by empirical data obtained in his experiment is grounded solely on an assumption that the inter- and intraspeaker variations within a group of criminal suspects or within any other speaker group would not differ substantially from those which existed within his test group.\textsuperscript{74}

In discussing the error rates reported by Tosi, Peter Ladefoged, formerly a vehement critic of spectrographic voice identification,\textsuperscript{76} noted that errors were often the result of examiner confusion between certain pairs of voices. Conceding that if Tosi's speaker population represented a statistically valid sample of criminal suspects the relatively low error rate of 6\% false identification could be taken as a probable maximum, Ladefoged felt nonetheless that the "confusability" factor should be further examined before fully accepting all of Tosi's claims.\textsuperscript{74}

Ladefoged's hesitancy in accepting the 6\% rate for false identification as a maximum\textsuperscript{77} stems from his assessment of the homogeneity of the speaker population used in the experiment. Although Tosi was careful to achieve homogeneity in the general sound of speech patterns,\textsuperscript{78} the college males constituting the universal population, and consequently those in the sample itself, almost certainly came from different backgrounds. Because interspeaker variations in spectrographic speech patterns can be expected to be greater the more heterogeneous the speaker population,\textsuperscript{79} thus resulting in fewer confusable voices, it is debatable whether a substantially similar number of voices exist in a smaller, more homogeneous community where interspeaker variations are fewer. Although it may be reasonable, as Tosi claims,\textsuperscript{80} to assume that the variation in probabilities will not be statistically significant, that claim remains for the moment a mere hypothesis, unsupported by empirical data. It is at least questionable whether the criteria for admissibility are met by the results and extrapolations of this one study.\textsuperscript{81}

This view, that the Tosi study, although methodologically sound, simply did not go far enough was advanced by Ladefoged in testimony during the trial court's hearing on the defendant's habeas corpus petition in \textit{State ex rel. Trimble v. Hedman},\textsuperscript{82} and by two former associates of Kersta at Bell Telephone Laboratories in \textit{State v. Andretta}.\textsuperscript{83} Although Ladefoged has now apparently changed his mind as to the degree to which the Tosi study validated spectrographic voice identification,\textsuperscript{84} Peter Denes, one of the witnesses testifying in the \textit{Andretta} case and the co-author of a 1970 report\textsuperscript{85} setting forth the requirements for scientific validation of the technique, felt that the Tosi study "only goes a small way towards clearing up a very large area of uncertainty before something more definite could be said about the reliability of this method."\textsuperscript{86} Denes did not feel that the results of the Tosi study would change

\begin{quote}
\textsuperscript{78} Hecker, Speaker Recognition at 16.
\textsuperscript{79} Tosi, Experiment at 21.
\textsuperscript{80} It should be pointed out that there is a fundamental difference between the test at issue in Coppolino and Tosi's experimentation with spectrographic voice identification. In Coppolino the test involved a chemical reaction, the interrelationship of discernible physical properties. Tosi's study, however, involved not a constant, but rather a host of variables. While the results of similar Coppolino-type tests can be predicted with certainty once the proper formulas are deduced, the results of Tosi's study are contingent upon variables which are serially dependent on both other variables and assumptions for their validity.
\textsuperscript{81} Minn. 192 N.W.2d 832 (1971), discussed in the text accompanying notes 98–107.
\textsuperscript{84} See Bolt, Reliability.
\end{quote}
in any way the negative conclusion as to the technique's scientific validity reached in the 1970 report.\textsuperscript{87}

The "confusability" factor to which Ladefoged refers in his discussion of the Tosi study points up the wholly examiner-dependent nature of the reliability of spectrographic voice identification.\textsuperscript{88} The ultimate identity inference to which the expert testifies in court is derived not from scientific principles, but rather from a purely subjective comparative analysis of similarities in the sound pattern parameters exhibited on a pair or series of spectrograms. At the moment not even these similarities can be quantitatively described.\textsuperscript{89} As both Tosi and Ladefoged note,\textsuperscript{90} the ethics and training of the examiner are highly important. The examiner must be aware of and able to evaluate the relationships between spectrum variations on the spectrogram and speaker-dependent identity traits.

A further difficulty in accepting Tosi's extrapolations, apart from the fact that his experiment did not adequately test the confusability of voices or the degree to which inter- and intraspeaker variability overlap among homogeneous population groups, is that so little is understood about speaker variabilities. Speech is a studied process of imitation and one's speech habits are not only subject to significant change prior to mature development, but the effect of subsequent physiological and environmental changes on the speech pattern has not yet been determined.\textsuperscript{91} Moreover, due to inherent intraspeaker variability, the causes of which have yet to be determined, no two spectrograms of phonetically identical utterances made by the same speaker are ever exactly alike. Nor is it fully understood yet what effect psychological stress, quite possibly a frequent factor in criminal acts, has on speech production. Two studies have indicated, for example, that spectrograms made of speech generated under stress conditions exhibit significant change from those generated under "normal" conditions.\textsuperscript{92}

Notwithstanding these reservations about the validity of spectrographic voice identification, the scientific reaction to Tosi's study and extrapolations is in marked contrast to the feelings expressed several years ago when Kersta published his theories.\textsuperscript{93} Ladefoged, for example, has testified in court\textsuperscript{94} that the dominant view of the scientific community is presently in accord with an opinion he expressed in a letter to Dr. Edward David, Science Adviser to the President of the United

87 Id. at 14.

88 This characteristic highlights the technique's similarity to polygraph testing. Although error rates of less than 10% have been reported for the polygraph, see F. INbau & J. Reed, Truth & Deception 226-34 (1966); Horvath & Reid, The Reliability of Polygraph Examiner Diagnosis of Truth & Deception, 62 J. Crim. L.C. & P.S. 276 (1971), neither courts nor scientists have responded favorably to this technique. Also, the nearly 50% rate of inconclusiveness experienced in many spectrographic voice analyses, see Michigan State Police, supra note 30 at 38-39, parallels the usefulness of the now debunked d'ermal nitrate test for gunpowder. See Turkel & Lipman, Unreliability of Dermal Nitrate Test for Gunpowder, 46 J. Crim. L.C. & P.S. 281 (1955).


90 Tosi, Experiment at 21-22. See also letter, supra note 2.

91 Mysak, Pitch and Duration Characteristics of Older Males, 2 J. Speech Hearing Res 46 (1959), found that aging is often accompanied by a rise in median fundamental frequency, greater variability in fundamental frequency, and a slight reduction in speaking rate. Sander, Maloney & Jackson, Phonatory and Related Changes with Advanced Age, 9 J. Speech Hearing Res. 333 (1966), speculated that the reduced range of fundamental frequency which they found associated with aging is caused by the aging of the laryngeal cartilages and muscles.

92 Various diseases of the chest, larynx and central nervous system are known to affect particular facets of speech production, but their effect on the spectral characteristics portrayed on the spectrogram is not known. See Hecker, Speaker Recognition at 18.


94 See Hecker, Stevens, von Bismark & Williams, Manifestations of Task-Induced Stress in the Acoustic Speech Signal, 44 J. Acoustical Soc'y Am. 993 (1968); Williams & Stevens, On Determining the Emotional State of Pilots During Flight: An Exploratory Study, 40 Aerospace Med. 1369 (1969). Stevens has suggested that because forensic applications of spectrographic voice identification may often be made in situations where a speaker is undergoing stress, further research into the physiological and emotional correlates of the acoustic speech signal may be essential to determine the validity of the technique. Interview with Dr. Kenneth N. Stevens, Professor of Electrical Engineering, Massachusetts Institute of Technology, in Cambridge, Mass., Nov. 23, 1971.

95 The Technical Committee on Speech Communication of the Acoustical Society of America expressed alarm at its 1966 meeting at the practical infallibility attributed to spectrographic voice identification by Kersta and the popular press which reported his conclusions. See Kersta, Nature; N.Y. Times, April 12, 1966, § L, at 1, col. 2. Several members of the committee undertook to review the reliability of the technique and to dispel certain allusions to fingerprint identification. Their findings and recommendations for validation of the technique are presented in Bolt, Reliability 606-08. A general summary may be found in Bolt, et al., Identification of a Speaker by Speech Spectrograms, 166 Science 338 (1969).

SPECTROGRAPHIC VOICE IDENTIFICATION

States. Ladefoged there stated that he would accept Tosi's 6% false identification figure as a "rough estimate" of the technique's accuracy.

Judicial Reaction

Those reported cases which have discussed the admissibility of spectrographic voice identification testimony since publication of the Tosi study have justified admitting such testimony by referring to the general rule laid down in Frye, arguing that scientific opinion of the technique is now favorable. These courts have felt that the Tosi study should be taken to validate the technique generally as a reliable method of voice identification, leaving the further unresolved questions as to voice similarity among smaller, more homogeneous groups as affecting only the weight and credibility of the identification testimony. It can be argued with some force, though, that Tosi's extrapolations, like those of Kersta, are premature. While Tosi's results are impressive, they do not support the further inference that the technique is reliable as a general principle in other than the laboratory test conditions: the quantum of reliability in other applications yet knows no limit. Notwithstanding the readiness now of the scientific community to accept the technique, the requirement of analytical validity expressed in Coppolino is not met by an experiment which produces a range of error which is presently undefined.

The text of Ladefoged's opinion reads, in pertinent part:

"If I were asked to testify on the validity of the system, I would have to emphasize that we do not at the moment know the probable error rate. But I would accept a minimum of 6% as a rough estimate of the possibility of making a misidentification (assuming, of course, that there was no question of the samples being involved, and that the identification had been made by an experienced, responsible, investigator)."

Ladefoged, testifying for the defense in the original habeas corpus hearing, stated that although he and other members of the scientific community accepted Tosi's experiments as far as they went, he would exercise more caution in extrapolating those results to prove the general validity and reliability of the technique. In his testimony for the state, Tosi disagreed, claiming that provided the examiner is properly trained, listens to the speech samples to be compared spectrographically, and is permitted to spend as much time as he requires, the technique is highly reliable. Tosi also disagreed with Ladefoged's contention that the sex of the speaker impaired the technique's reliability in any way.

With regard to the degree of reliability required of a scientific technique, the court noted that mere disagreement over the fallibility of a scientific technique would not render an expert's opinion, based on the technique, inadmissible.

---

54 The text of Ladefoged's opinion reads, in pertinent part:

55 All evidence, must, of course, be probative. See note 56 supra. When there is no support for the proposition that a technique is accurate within certain tolerance limits, the probativeness of evidence based on the technique is substantially reduced. A strong argument can be made that such evidence should be "excluded because altogether lacking in probative value." See Coolidge v. New Hampshire, 403 U.S. 443, 448 n.2 (1971), citing Tribe, supra note 18, at 1342 n.40. See also note 81 supra.

though the Trimble court was correct in stating that the infallibility of a scientific technique or device is not a prerequisite to admissibility, the expected rate of error is a matter to be considered in deciding whether to admit identification testimony based on a novel technique. If Ladefoged's reservation regarding the confusability of voices in other samplings is valid, then the expected rate of error has only a likely minimum, with no real evidence to sustain a probable maximum.

While the Trimble court recognized that the precise question before it was narrower than the broader admissibility issue faced by Cary, King, and Wright, it nevertheless noted that:

[In the trial of the case spectrograms [sic] ought to be admissible for the purpose of corroborating voice identification by aural means if a sufficient foundation is laid to satisfy the trial judge that the expert whose opinion is sought is qualified to assist the fact finder in coming to the right conclusion.]

This statement, although dicta, has already had substantial legal effect. The court seemed convinced that the technique had advanced to the point where judicial notice could be taken of its reliability, the only bar to its admissibility being its use by an unqualified voice identification expert.

At least two federal courts have followed the lead of the Minnesota Supreme Court. In United States v. Phoenix the trial judge admitted evidence of a spectrographic identification after hearing extensive testimony by both Tosi and Ernest Nash, a voice identification expert by the Michigan State Police. In a written opinion, a District of Columbia trial judge in United States v. Raymond value of a technique is in perfect equipoise and the high degree of probativeness required of scientific evidence is not present. Given the serious questions with regard to the relationship of the non-tested variables to the accuracy of the technique, disagreement among experts should in fact render evidence based on the technique inadmissible.

It is doubtful that infallibility for a scientific technique is or should be a prerequisite for its admissibility. See 3 Wigmore § 990; McCormick, Deception Tests and the Law of Evidence, 15 Calif. L. Rev. 484, 500 (1927).

See note 88 supra.

...Minn. at 192 N.W.2d at 444 (emphasis added).


No. 70-CR-428 (S.D. Ind. April 15, 1971).
the matter, has not moved to accept spectrographic voice identification in other than investigative
situations.\textsuperscript{113}

A recent New Jersey case, \textit{State v. Andretta},\textsuperscript{114} however, has adhered to the exclusionary result
reached by the court in \textit{State v. Cary}. Despite the precedent set by Trimble and Raymond, both cited
in the opinion, and the radical change in scientific thought on the technique's accuracy and reliabil-
ity as a result of the Tosi study, the court noted that:

The evidence offered for the purpose of showing scientific acceptance did not go so far as to include
acceptance of the proposition that the technique would result in a scientifically reliable identifica-
tion or elimination by the expert, but \textit{merely that the study of Dr. Tosi, so far as it went, was con-
ducted in accordance with valid scientific standards.}\textsuperscript{115}

The court in \textit{Andretta} was not presented squarely with an admissibility question. Rather, the state
had requested a court order to compel the defendants in the case to submit voice exemplars
for spectrographic comparison with spectrograms made from exemplars taken nearly four years
before. The court was unimpressed with the testi-
mony of Tosi, Ladehoff, and Nash that time-
lapses of greater than one month would have no appreciable effect on the accuracy of an examiner's
identification or elimination. Because the Tosi study had only tested the effect of a one month
time-lapse on the accuracy of the technique, the
court did not feel that the further inference that
no statistically significant differences would be
found in greater time-lapse situations could be
drawn.\textsuperscript{116}

\textsuperscript{113} The late FBI Director J. Edgar Hoover stated, in
response to a request for FBI policy on the question,
that in view of the many unresolved questions which
may strike at the heart of the technique's reliability:
[We feel that the comparison of voiceprints is use-
ful as an investigative guide but has not been proven
sufficiently well authenticated to serve as a reliable
basis for expert testimony, as to identity, at this
time.

Letter from J. Edgar Hoover, Director, Federal Bureau
of Investigation, to the author, December 21, 1971, on
file in the editorial offices of the \textit{Journal of Criminal
Law, Criminology & Police Science}. The position of the
Bureau remains unchanged. Letter from L. Patrick
Gray III, Acting Director, Federal Bureau of In-
vestigation, to the author, June 15, 1972, on file in
the editorial offices of the \textit{Journal}.

\textsuperscript{114} Crim. Indict. Nos. 445–66/446–66 (Middlesex

\textsuperscript{115} \textit{Id.}, slip op. at 7–8.

\textsuperscript{116} \textit{Id.}, slip op. at 16–20.

\begin{center}
\textbf{Conclusion}
\end{center}

The Tosi study and the scientific respect ac-
corded it clearly go a long way toward establishing
spectrographic voice analysis as a reliable method
of voice identification. Since the technique may
be used to exclude\textsuperscript{117} as well as to identify, its
forensic applicability will benefit prosecution and
defense alike.

Although recent case law supports the admiss-
ibility of such identification testimony, there are
significant unquantified variables which may have
an as yet undetermined effect on the accuracy
and reliability of spectrographic voice analysis.
The courts ought to look carefully at the evidence
marshalled to support the general admissibility of
such an identification until further experimen-
tation establishes the relationship which these
variables have on the examiner's proficiency.
While the technique certainly possesses sufficient
reliability to render an identification admissible
on stipulation, at least in those jurisdictions which
allow such use of polygraph test results,\textsuperscript{118} the
courts ought not be too eager to embrace a tech-
nique which has only recently been rescued from
an abyss of scientific scorn—a rescue accomplished
by the narrow experimental data produced in
one study.

Even in those jurisdictions where evidence of
a spectrographic voice identification is adjudged
admissible, the bench or jury can and should be
made fully aware of the inadequate knowledge
speech scientists have of the many factors which
influence speech production. While the fact finder
is entitled to have before it all available data which
possesses a reasonable level of probativeness, the
danger of prejudice in using a technique, the ac-
curacy of which is so heavily dependent on un-
defined variables and the ethics, training, and
general responsibility of the examiner, would seem
to tip the scales toward exclusion, notwithstanding
the technique's "scientific acceptance." Courts can-
not wave \textit{Frye} as a talisman directing admissibility;
they may not abdicate their responsibilities by
blindly following general scientific acceptance.

\textsuperscript{117} See note 112 supra.

\textsuperscript{118} See, e.g., \textit{State v. Valdez}, 91 Ariz. 274, 371 P.2d
894 (1962). The \textit{Valdez} opinion, however, conditioned
even this restricted use of the polygraph technique's
results on requirements that the court and jury be
made fully aware of the possibilities for error inherent
in the technique.