

VOICE IDENTIFICATION: The Aural/Spectrographic Method

by: Michael C. McDermott (mike@mcldtd.com), Tom Owen (owlmax@aol.com),
Frank M. McDermott, Ltd.
Owl Investigations, Inc.

Table of Contents:

I.

INTRODUCTION

II.

The Sound Spectrograph

III.

The Method of Voice Identification

IV.

History

V.

Standards of Admissibility

VI.

Research Studies

VII.

Conclusion

VIII.

Table of Cases

IX.

Appendix 1

© 1996 Owl Investigations, Inc.

INTRODUCTION

The forensic science of voice identification has come a long way from when it was first introduced in the American courts back in the mid 1960's. In the early days of this identification technique there was little research to support the theory that human voices are unique and could be used as a means for identification. There was also no standardization of how an identification was reached, or even

training or qualifications necessary to perform the analysis. Voice comparisons were made solely on the pattern analysis of a few commonly used words. Due to the newness of the technique there were only a few people in the world who performed voice identification analysis and were capable of explaining it to a court. Gradually the process became known to other scientists who voiced concerns, not as to the validity of the analysis, but as to the lack of substantial research demonstrating the reliability of the technique. They felt that the technique should not be used in the courtroom without more documentation. Thus the battle lines were drawn over the admissibility of voice identification evidence with proponents claiming a valid, reliable identification process and opponents claiming more research must be completed before the process should be used in courtrooms.

Today voice identification analysis has matured into a sophisticated identification technique, using the latest technology science has to offer. The research, which is still continuing today, demonstrates the validity and reliability of the process when performed by a trained and certified examiner using established, standardized procedures. Voice identification experts are found all over the world. No longer limited to the visual comparison of a few words, the comparison of human voices now focuses on every aspect of the words spoken; the words themselves, the way the words flow together, and the pauses between them. Both aural and spectrographic analysis are combined to form the conclusion about the identity of the voices in question.

The road to admissibility of voice identification evidence in the courts of the United States has not been without its potholes. Many courts have had to rule on this issue without having access to all the facts. Trial strategies and budgets have resulted in incomplete pictures for the courts. To compound the problem, courts have utilized different standards of admission resulting in different opinions as to the admissibility of voice identification evidence. Even those courts which have claimed to use the same standard of admissibility have interpreted it in a variety of ways resulting in a lack of consistency. Although many courts have denied admission to voice identification evidence, none of the courts excluding the spectrographic evidence have found the technique unreliable. Exclusion has always been based on the fact that the evidence presented did not present a clear picture of the technique's acceptance in the scientific community and as such, the court was reluctant to rely on that evidence. The majority of courts hearing the issue have admitted spectrographic voice identification evidence.

THE SOUND SPECTROGRAPH

The sound spectrograph, an automatic sound wave analyzer, is a basic research instrument used in many laboratories for research studies of sound, music and speech. It has been widely used for the analysis and classification of human speech sounds and in the analysis and treatment of speech and hearing disorders.

The instrument produces a visual representation of a given set of sounds in the parameters of time, frequency and amplitude. The analog spectrograph is composed of four basic parts; (1) a magnetic tape recorder/playback unit, (2) a tape scanning device with a drum which carries the paper to be marked, (3) an electronic variable filter, and (4) an electronic stylus which transfers the analyzed information to the paper. The analog sound spectrograph samples energy levels in a small frequency range from a magnetic tape recording and marks those energy levels on electrically sensitive paper. This instrument then analyses the next small frequency range and samples and marks the energy levels at that point. This process is repeated until the entire desired frequency range is analyzed for that portion of the recording. The finished product is called a spectrogram and is a graphic depiction of the patterns, in the form of bars or formants, of the acoustical events during the time frame analyzed. The machine will produce a spectrogram in approximately eighty seconds. The spectrogram is in the form of an X,Y graph with the X axis the time dimension, approximately 2.4 seconds in length, and the Y axis the frequency range, usually 0 to 4000 or 8000 Hz. The degree of darkness of the markings indicates the approximate relative amplitude of the energy present for a given frequency and time.

Recent developments in sound spectrography have produced computerized digital sound spectrographs ranging from dedicated digital signal analysis workstations to PC-based systems for acquisition, analysis editing, and playback. These sophisticated computer-based systems provide high fidelity signal acquisition, high-speed digital processing circuitry for quick and flexible analysis, and CD-quality playback. The computerized-based systems accomplish all the same tasks of the analog systems, but with the computer-based systems the examiner gains a host of comparison and measurement tools not available with the analog equipment. The computer-based systems are capable of displaying multiple sound spectrograms, adjusting the time alignment and frequency ranges and taking detailed numeric measurements of the displayed sounds. With these advances in technology, the examiner widens the scope of the analysis to create a more detailed picture of the voice or sound being analyzed.

The accuracy and reliability of the sound spectrograph, either analog or digital, has never been in question in any of the courts and never considered an issue in the admissibility of voice identification evidence. This may be due in part to the wide use of the instrument in the field of speech and hearing for non-voice identification analysis of the human voice and, in part to the fact that given the same recording of speech sounds the sound spectrograph will consistently produce the same spectrogram of that speech.

The contest comes in the interpretation of the spectrograms. Proponents of the aural and spectrographic technique of voice identification base their decisions on the theory that all human voices are different due to the physical uniqueness of the vocal track, the distinctive environmental influences in the learning process of speech development, and the unique development of neurological faculties which

are responsible for the production of speech. Opponents claim that not enough research has been completed to validate the theory that intraspeaker variability is less than interspeaker variability.

THE METHOD OF VOICE IDENTIFICATION

The method by which a voice is identified is a multifaceted process requiring the use of both aural and visual senses. In the typical voice identification case the examiner is given several recordings; one or more recordings of the voice to be identified and one or more recorded voice samples of one or more suspects. It is from these recordings the examiner must make the determination about the identity of the unknown voice.

The first step is to evaluate the recording of the unknown voice, checking to make sure the recording has a sufficient amount of speech with which to work and that the quality of the recording is of sufficient clarity in the frequency range required for analysis.¹ The volume of the recorded voice signal must be significantly higher than that of the environmental noise. The greater the number of obscuring events, such as noise, music, and other speakers, the longer the sample of speech must be. Some examiners report that they reject as many as sixty percent of the cases submitted to them with one of the main reasons for rejection being the poor quality of the recording of the unknown voice.

Once the unknown voice sample has been determined to be suitable for analysis, the examiner then turns his attention to the voice samples of the suspects. Here also, the recordings must be of sufficient clarity to allow comparison, although at this stage, the recording process is usually so closely controlled that the quality of recording is not a problem.

The examiner can only work with speech samples which are the same as the text of the unknown recording. Under the best of circumstances the suspects will repeat, several times, the text of the recording of the unknown speaker and these words will be recorded in a similar manner to the recording of the unknown speaker. For example, if the recording of the unknown speaker was a bomb threat made to a recorded telephone line then each of the suspects would repeat the threat, word for word, to a recorded telephone line. This will provide the examiner with not only the same speech sounds for comparison but also with valuable information about the way each speech sound completes the transition to the next sound.

There are those times when a voice sample must be obtained without the knowledge of the suspect. It is possible to make an identification from a surreptitious recording but the amount of speech necessary to do the comparison is usually much greater. If the suspect is being engaged in conversation for the purpose of obtaining a voice sample, the conversation must be manipulated in

such a way so as to have the suspect repeat as many of the words and phrases found in the text of the unknown recording as possible.

The worst exemplar recordings with which an examiner must work are those of random speech. It is necessary to obtain a large sample of speech to improve the chances of obtaining a sufficient amount of comparable speech.

As in any other form of identification analysis, as the quality of the evidence with which the examiner has to work declines, the greater the amount of evidence and time necessary to complete the analysis, and the less likely the chance for a positive conclusion.

Once the evidence has been determined to be sufficient to perform the analysis, the examiner then begins the two step process of voice sample comparison; one aural (listening) and the other spectrographic (visual). These are two different but interwoven and equally important analytical methods which the examiner combines to reach the final conclusion. The first step is an aural comparison of the voice samples.² Here the examiner compares both single speech sounds and series of speech sounds of the known and unknown samples. At this stage the examiner is conducting a number of tasks; comparing for similarities and differences, screening out less useful portions of the samples, and indexing the samples for further analysis. An example of the initial aural comparison is the screening of the samples for pronunciation similarities or discrepancies such as the word "the" may be said with a short "a" sound or a long "e" sound. If the word is not pronounced in the same manner it loses comparison value.

Once the examiner has located those portions to be used for the analysis, a more detailed aural comparison is undertaken. This comparison can be accomplished in many different ways. One of the most commonly used methods of aural comparison is re-recording a speech sound sample of the unknown followed immediately by a re-recording of the same speech sounds of the suspect. This is repeated several times so that the final product is a recording of specific speech sounds, in alternating order, by the unknown speaker followed by the suspect. Such comparisons have been greatly facilitated by the use of audio digital recording equipment which allows for the digital recording, storage, and repeated playback of only the desired speech sounds to be examined.

During the aural comparison the examiner studies the psycholinguistic features of the speakers voice. There are a large number of qualities and traits which are examined from such general traits as accent and dialect to inflection, syllable grouping and breath patterns. The examiner also scrutinizes the samples for signs of speech pathologies and peculiar speech habits.

The second step in the voice identification process is the spectrographic analysis of the recorded samples. The sound spectrograph is an automatic sound wave analyzer with a high quality, fully functional tape recorder. The speech samples

to be analyzed are recorded on the sound spectrograph. The recording is then analyzed in two and one half second segments. The product is a spectrogram, a graphic display of the recorded signal on the basis of time and frequency with a general indication of amplitude.

The spectrograms of the unknown speaker are then visually compared to the spectrograms of the suspects. Only those speech sounds which are the same are compared.³ The comparisons of the spectrograms are based on the displayed patterns representing the psychoacoustical features of the captured speech. The examiner studies the bandwidths, mean frequencies, and trajectory of vowel formants; vertical striations, distribution of formant energy and nasal resonances; stops, plosives and fricatives; interformant features, the relation of all features present as affected during articulatory changes and any peculiar acoustic patterning.⁴ The examiner looks not only for similarities but also for differences. The differences are closely examined to determine if they are due to pronunciation differences or if they are indicative of different speakers.

When the analysis is complete the examiner integrates his findings from both the aural and spectrographic analyses into one of five standard conclusions; a positive identification, a probable identification, a positive elimination, a probable elimination, or no decision. In order to arrive at a positive identification the examiner must find a minimum of twenty speech sounds which possess sufficient aural and spectrographic similarities. There can be no differences either aural or spectrographic for which there can be no accounting.

The probable identification conclusion is reached when there are less than twenty similarities and no unexplained differences. This conclusion is usually reached when working with small samples, random speech samples or recordings of lower quality. The result of positive elimination is rendered when twenty differences between the samples are found that can not be based on any fact other than different voices having produced the samples. A probable elimination decision is usually reached when working with limited text or a recording of lower quality. The no decision conclusion is used when the quality of the recording is so poor that there is insufficient information with which to work or when there are too few common speech sounds suitable for comparison.

HISTORY

A good place to start examining the history of speech sound analysis goes back a little more than one hundred years to Alexander Melville Bell who developed a visual representation of the spoken word. This visual display of the spoken word conveyed much more information about the pronunciation of that word than the dictionary spelling could ever suggest. His depiction of speech sounds demonstrated the subtle differences with which different people pronounced the same words. This system of speech sound analysis developed by Bell is the phonetic alphabet which he called "visible speech".⁵ His method of encoding the

great variety of speech sounds was by handwritten symbols and was language independent. This code produced a visual representation of speech which could convey to the eye the subtle differences in which words were spoken. This system was used by both Bell and his son, Alexander Graham Bell, in helping deaf people learn to speak.⁶

It was in the early 1940's that a new method of speech sound analysis was developed. Potter, Kopp & Green, working for Bell Laboratories in Murray Hill, New Jersey, began work on a project to develop a visual representation of speech using a sound spectrograph. This machine, an automatic sound wave analyzer, produced a visual record of speech portraying three parameters; frequency, intensity and time. This research was intensified during World War II when acoustic scientists suggested that enemy radio voices could be identified by the spectrograms produced by the sound spectrograph. The war ended before the technique could be perfected.

In 1947, Potter, Kopp and Green published their work in a book, the title of which was borrowed from Alexander Melville Bell, *Visible Speech*. Their work is a comprehensive study of speech spectrograms designed to linguistically interpret visible speech sound patterns. This work was similar to that of Bell's in that speech sounds were encoded into a visual form. The difference is, instead of a pen, Potter, Kopp and Green used a sound spectrograph to produce the visual patterns.

Research in the area of speaker identification slowed dramatically with the end of World War II. It was not until the late 1950's and early 1960's that the research began again. It was at this time the New York City Police Department was receiving a large number of telephone bomb threats to the airlines.⁷ At that time Bell Laboratories was asked by law enforcement officers to provide assistance in the apprehension of the individuals making the telephone calls. The task of developing a reliable method of identification of a speaker's voice was given to Lawrence G. Kersta, a physicist at Bell Laboratories who had worked on the early experiments using the sound spectrograph. In two years Kersta had developed a method of identification in which he reported results yielding a correct identification 99.65% of all attempts.⁸

It was in 1966 that the Michigan State Police began the practical application of the voice identification method in attempting to solve criminal cases. A Voice Identification unit was established and the unit personnel received training from Kersta and other speech scientists. During the first few years the voice identification method was used only as an investigative aid.

The first court of published opinion to rule on the admissibility of voice identification analysis was in the case of *United States v. Wright*, 17 USCMA 183, 37 CMR 447 (1967). This was a court martial proceeding in which the

appellate court affirmed the admission of spectrographic voice identification evidence by the board of review. The lengthy dissent by Judge Ferguson based on the requirements for acceptance of scientific evidence spelled out in *Frye v. United States*, 293 Fed. 1013 (CA DC Cir) (1923), was the beginning of a controversy which continues today.

The first non-military case to review the admissibility of voice identification evidence was the New Jersey Supreme Court in *State v. Cary*.⁹ In this case the court stated that "the physical properties of a person's voice are identifying characteristics".¹⁰ The court also noted that trial courts in the states of New York and California have admitted voice identification evidence but that these admissions have not been subject of appellate review.¹¹ The court declined to rule on the admissibility issue and remanded the case to determine if the equipment and technique were sufficiently accurate to provide results admissible as evidence. The Superior Court of New Jersey, on appeal from a denial of admission after remand, held that the majority of evidence "indicates, not that the technique is not accurate and reliable, but rather that it is just too early to tell and at this time lacks the required scientific acceptance".¹² The New Jersey Supreme Court reviewed this decision and once again remanded for additional fact finding "in light of the far-reaching implications of admission of voiceprint evidence".¹³ The State of New Jersey was unable "to furnish any new and significant evidence" by the third time the New Jersey Supreme Court reviewed this issue and as such affirmed the trial court's opinion excluding voice identification evidence.¹⁴

California came to a similar holding when the issue first reached the appellate level in *People v. King*.¹⁵ The State brought in Lawrence Kersta as the voice identification expert to testify as to the reliability of the technique. The defense brought in seven speech scientists and engineers to rebut Kersta's claims. The court held that "Kersta's claims for the accuracy of the 'voiceprint' process are founded on theories and conclusions which are not yet substantiated by accepted methods of scientific verification".¹⁶ The court cited the *Frye* test as the proper standard for admissibility.¹⁷ The court also left the door open for future admission by saying when voice identification evidence has achieved the necessary degree of acceptance they will welcome its use.¹⁸

In *State ex rel. Trimble v. Heldman* ¹⁹, the Supreme Court of Minnesota held that "spectrograms ought to be admissible at least for the purpose of corroborating opinions as to identification by means of ear alone".²⁰ The court was impressed by the testimony of Dr. Oscar Tosi who had previously testified against the use of spectrographic voice identification evidence in courtrooms, but after extensive research and experimentation now described the technique as "extremely reliable".²¹ The court made reference to the *Frye* test and to the scientific community's acceptance of Dr. Tosi's study, but did not specifically apply the *Frye* test as the standard for the admissibility of the voice identification

evidence.²² In discussing the issue of admissibility the court held that it was the job of the factfinder to weight the credibility of the evidence.

"The opinion of an expert is admissible, if at all, for the purpose of aiding the jury or the factfinder in a field where he has no particular knowledge or training. The weight and credibility to be given to the opinion of an expert lies with the factfinder. It is no different in this field than in any other".²³

In 1972 the third and fourth District Courts of Florida, in separate opinions, held admissible the use of spectrographic voice identification evidence.²⁴ The court in *Worley* held that the voice identification evidence was admissible to corroborate the defendant's identification by other means. The court stated that the technique had attained the necessary level of scientific reliability required for admission, but since it was only offered as corroborative evidence, the court refused to comment as to whether such evidence alone would be sufficient to sustain the identification and conviction.²⁵

The third District Court of Appeals of Florida did not limit the admission of spectrograph evidence to corroborative status. In the *Alea* opinion the court does not mention the *Frye* test as the standard to be used for admission, but rather states that "such testimony is admissible to establish the identity of a suspect as direct and positive proof, although its probative value is a question for the jury".²⁶

In the case of *State v. Andretta* ²⁷, the New Jersey Supreme Court stated that there was much more support for the admission of spectrographic voice identification evidence than at the time they decided *Cary*, but refused to address the issue further since the only issue before them was whether the defendant should be compelled to speak for a spectrographic voice analysis.²⁸

In California the Court of Appeal affirmed the trial court's admission of voice identification evidence in the case of *Hodo v. Superior Court*.²⁹ Here the court found the requirements of *Frye* had been met in that there was now general acceptance of spectrographic voice identification by recognized experts in the field. The court cited Dr. Tosi's testimony that "those who really are familiar with spectrography, they are accepting the technique".³⁰ Tosi also pointed out that the general population of speech scientists are not familiar with this technique and thus can not form an opinion on it.³¹

The court in *United States v. Samples* ³² held that the *Frye* test of general acceptance precludes too much relevant evidence for purposes of the fact determining process at a revocation of probation hearing and the court allowed the use of spectrographic voice identification evidence to corroborate other identification evidence.³³

In 1974 the case of *United States v. Addison* ³⁴ rejected the admission of voice identification evidence saying that such evidence "is not now sufficiently

accepted" and as such the requirements of the Frye test were not met.³⁵ At the trial the court heard from two experts endorsing the technique, Dr. Tosi and a recent convert to the reliability of the technique, Dr. Ladefoged. Only one expert, Dr. Stuart, testified that he was still skeptical of the technique and thought that most of the scientific community was also.³⁶ Although the admission of spectrographic voice identification evidence was held to be error by the trial court, the appellate court refused to overturn the conviction due to overwhelming amount of other evidence supporting the conviction.³⁷

Attempted disguise or mimic were the grounds the California Court of Appeal used to reverse a conviction based in part on spectrographic voice identification in the case of *People v. Law*.³⁸ The court found that "with respect to disguised and mimicked voices in particular, the prosecution did not carry out its burden of proof to demonstrate that the scientific principles pertaining to spectrographic identification were beyond the experimental and into the demonstrable stage or that the procedure was sufficiently established to have gained general acceptance in the particular field in which it belongs".³⁹ The main concern of the court was that no experimentation had been completed studying the effects of attempts to disguise or mimic on the accuracy of the identification process. Without mentioning the Frye test this court used the standards set in Frye as the test of admissibility although the court seemed to be limiting the scope of the opinion to cases involving disguise or mimic.

In *United States v. Franks* ⁴⁰, the Sixth Circuit Court of Appeals held spectrographic voice identification evidence to be admissible. The court said it was "mindful of a considerable area of discretion on the part of the trial judge in admitting or refusing to admit evidence based on scientific processes".⁴¹ Quoting from *United States v. Stifel* ⁴², the court pointed out that "neither newness nor lack of absolute certainty in a test suffices to render it inadmissible in court. Every useful new development must have its first day in court. And court records are full of the conflicting opinions of doctors, engineers and accountants...".⁴³ The court in *Franks* found that extensive review was given to the qualifications of the experts and opportunity to cross-examine the experts to determine the proper weight to be given such evidence.

The Massachusetts Supreme Court, in *Commonwealth v. Lykus* ⁴⁴, allowed the admission of spectrographic voice identification evidence saying that the opinions of a qualified expert should be received and the considerations similar to those expressed in Frye should be for the fact finder as to the weight and value of the opinions. The court gave greater weight to those experts who had had direct and empirical experience in the field as opposed to those who had only performed a theoretical review of that work.⁴⁵ The court also stated that "neither infallibility nor unanimous acceptance of the principle need be proved to justify its admission into evidence".⁴⁶ The Massachusetts Supreme Court again, that same year, found no error in the use of spectrographic voice identification evidence in the case of *Commonwealth v. Vitello*.⁴⁷

The Fourth Circuit Court of Appeals, in the case of *United States v. Baller* 48, allowed the admission of spectrographic voice identification evidence saying unless it is prejudicial or misleading to the jury, it is better to admit relevant scientific evidence in the same manner as other expert testimony and allow its weight to be attacked by cross-examination and refutation.⁴⁹ The court listed six reasons supporting admission; the expert was a qualified practitioner, evidence in voir dire demonstrated probative value, competent witnesses were available to expose limitations, the defense demonstrated competent cross-examination, the tape recordings were played for the jury, and the jury was told they could disregard the opinion of the voice identification expert.⁵⁰

Voice identification evidence was admitted by the Sixth Circuit Court of Appeals in *United States v. Jenkins* 51 using the same logic as in *Baller*. Here the court said that the issue of admissibility was within the discretion of the trial judge and that once a proper foundation had been laid the trier of fact was able to assign proper weight to the evidence.⁵²

In 1976 the New York Supreme Court pointed out, in the case of *People v. Rogers* 53, that fifty different trial courts had admitted spectrographic voice identification evidence, as had fourteen out of fifteen U. S. District Court judges, and only two out of thirty- seven states considering the issue had rejected admission.⁵⁴ The *Rogers* court stated that this technique, when accompanied by aural examination and conducted by a qualified examiner, had now reached the level of general scientific acceptance by those who would be expected to be familiar with its use, and as such, has reached the level of scientific acceptance and reliability necessary for admission.⁵⁵ The court also pointed out that other scientific evidence processes are regularly admitted which as, or less, reliable than spectrographic voice identification; hair and fiber analysis, ballistics, forensic chemistry and serology, and blood alcohol tests.⁵⁶

The Supreme Court of California finally put an end to the see-saw ride of admissibility in that state in *People v. Kelly* 57 by rejecting admission because of insufficient showing of support. "Although voiceprint analysis may indeed constitute a reliable and valuable tool in either identifying or eliminating suspects in criminal cases, that fact was not satisfactorily demonstrated in this case".⁵⁸ In this case the court seemed to have the most trouble with the fact the only expert provided to lay the foundation for admission was the technician who performed the analysis, saying that a single witness can not attest to the views of the scientific community on this new technique and that this witness, who may not be capable of a fair and impartial evaluation of the technique since he has built a career on it, lacked the academic credentials to express an opinion as to the acceptance of the technique by the scientific community.⁵⁹

In *United States v. McDaniel* 60, it appears that District of Columbia Circuit Court of Appeals would have liked to admit the spectrographic voice identification evidence but had to reject it because the shadow of the *Addison* decision of two

years past "looms over our consideration of this issue".⁶¹ The court held the admission of the voice identification evidence to be harmless error in that the rest of the evidence was overwhelming. The court did recognize the trend toward admissibility and contemplated that it may be time to reexamine the holding of Addison "in light of the apparently increased reliability and general acceptance in the scientific community".⁶²

The Supreme Court of Pennsylvania rejected admission in *Commonwealth v. Topa* ⁶³ holding that the technician's opinion alone will not suffice to permit the introduction of scientific evidence into a court of law.⁶⁴ This was the same situation, in fact the same single expert, which confronted the Kelly court.

In *People v. Tobey* ⁶⁵ the Michigan Supreme Court found, by applying the Frye test, that the trial court erred in admitting spectrographic voice identification evidence. The court found that neither of the two experts testifying in favor of the technique could be called disinterested and impartial experts in that both had built their reputations and careers on this type of work.⁶⁶ The court pointed out that not all courts require independent and impartial proof of general scientific acceptability and was quick to add that this decision was not intended in anyway to foreclose the introduction of such evidence in future cases where there is demonstrated solid scientific approval and support of this new method of identification.⁶⁷

In admitting voice identification evidence, the United States District Court for the Southern District of New York, in *United States v. Willaims* ⁶⁸, found that the requirements of the Frye test were met when the technique was performed "by aural comparison and spectrographic analysis".⁶⁹ The court stated that the concerns of the defendant that this technique had a mystique of scientific precision which may mask the ultimate subjectivity of spectrographic analysis, although they were valid concerns, could be alleviated by action other than suppression of the evidence, such as opposing expert opinion and jury instructions allowing the jury to determine the weight, if any, of the evidence.⁷⁰

In *People v. Collins* ⁷¹, the Supreme Court of New York rejected admission of spectrographic voice identification evidence saying that the Frye test alone was insufficient to determine admissibility and must be used in conjunction with a test of reliability.⁷² The court found that the proponents of the technique were in the minority and that the remainder of the relevant scientific community either expressed opposition or expressed no opinion.⁷³

In *Brown v. United States* ⁷⁴, the District of Columbia Court of Appeals rejected the use of voice identification evidence, but held the error to be harmless and affirmed the conviction in light of overwhelming non-spectrographic identification of the defendant as perpetrator of the crime. One of the main problems in this case was the fact that the exemplar of the defendant's voice was recorded in a defective manner but used anyway after the tape speed malfunction had been

corrected in a laboratory. Dr. Tosi, testifying as a proponent of the technique, stated that the technician should not have used the defective recording as a basis of comparison.⁷⁵ The court held the technique was not shown to be sufficiently reliable and accepted within the scientific community to permit its use in this criminal case, but that this decision did not foreclose a future decision as to admissibility of the technique.⁷⁶

In the civil case of *D'Arc v. D'Arc* ⁷⁷, the court found that the requirements of the Frye test had not been met and thus the evidence could not be admitted. The court believed that even with proper instructions to the contrary, this type of evidence "has the potentiality to be assumed by many jurors as being conclusive and dispositive" and thus should be subject to strict standards of admission.⁷⁸

The court in *State v. Williams* ⁷⁹ refused to apply the Frye standard citing instead the Maine Rules of Evidence, Rule 401, which states "all relevant evidence is admissible", with relevant being described as evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.⁸⁰

In *Reed v. State* ⁸¹ the court applied the Frye standard to determine admissibility with a rather wide definition of the scientific community which included "those whose scientific background and training are sufficient to allow them to comprehend and understand the process and form a judgment about it".⁸² The court said the trial court erred in using the more restricted definition of scientific community, "those who are knowledgeable, directly knowledgeable through work, utilization of the techniques, experimentation and so forth" and did not mean the broad general scientific community of speech and hearing science.⁸³

In a fifty-one page dissent to the *Reed* decision ⁸⁴, Judge Smith points out that the Frye standard is much criticized and has never been adopted in the state of Maryland, that this decision is out of step with other courts on related issues of fingerprints, ballistics, x-rays and the like, that this decision is out of step with prior Maryland holdings on expert testimony, that the majority of reported opinions have accepted such evidence, and that even if Frye were applicable it is satisfied.

In *United States v. Williams* ⁸⁵ the court did not apply the Frye standard but did note that acceptance of the technique appeared strong among scientists who had worked with spectrograms and weak among those who had not.⁸⁶ The court then focused on the reliability of the technique and the tendency to mislead. As to the reliability of the technique, the court noted the small error rate, 2.4% false identification, the existence and maintenance of standards of analysis, and the conservative manner in which the technique was applied.⁸⁷ As to the tendency to mislead, the court felt that adequate precautions were taken in that the jury could view the spectrograms and listen to the recording and the expert's qualifications, the reliability of the equipment and the technique were subject to

scrutiny by the defense, and the jury was instructed that they were free to disregard the testimony of the experts.⁸⁸

In the case of *People v. Bein* ⁸⁹ the court based admissibility on a two pronged test; general acceptance by the relevant scientific community, and competent expert testimony establishing reliability of the process. The court found that both tests had been met and allow the admission of the evidence.⁹⁰ The court described the relevant scientific community "to be that group of scientists who are concerned with the problems of voice identification for forensic and other purposes".⁹¹ The court also suggested that "it is no different in this field of expertise than in other fields, that where experts disagree, it is for the finder of fact to determine which testimony is the more credible and therefore more acceptable".⁹²

The Ohio Supreme Court, in *State v. Williams* ⁹³, relied on their own state rules of evidence, as did the Maine court in *Williams*, and rejected the use of the Frye standard. The court refused "to engage in scientific nose counting for the purpose of whether evidence based on newly ascertained or applied scientific principles is admissible".⁹⁴ The court noted, with approval, the playing of the recordings to the jury and, that the jury was free to reject the testimony of the expert.⁹⁵

In that same year, right across the border in Indiana, the court in *Cornett v. State*⁹⁶ rejected admission of voice identification evidence saying the conditions set out in Frye had not been met. Here the court used a wide definition of the scientific community which included linguists, psychologists and engineers who use voice spectrography for identification purposes.⁹⁷ Although the court held that the trial court erred in admitting the evidence, the error was found to be harmless and the conviction affirmed.⁹⁸

Likewise the court in *State v. Gortarez* ⁹⁹ rejected the admission of voice identification evidence but affirmed the conviction holding such admission to be harmless error. The court also used a wide definition of the scientific community in applying the Frye standard including experts in the fields of acoustical engineering, acoustics, communication electronics, linguists, phonetics, physics and speech communications and found that there was not general acceptance among these scientists.¹⁰⁰

In the case of *United States v. Love*¹⁰¹, the admissibility of spectrographic voice identification was not at issue. The fourth circuit Court of Appeals was reviewing whether the trial judge's comments about a voice identification expert were considered error. The trial judge told the jury that they, the jury, were to assign whatever weight they wanted to the testimony of the expert and even disregard his testimony if they "should conclude that his opinion was not based on adequate education, training or experience, or that his professed science of voice

print identification was not sufficiently reliable, accurate, and dependable."¹⁰² The Court of Appeals found no error in the judge's instruction to the jury.

In admitting spectrographic voice identification evidence, the Supreme Court of Rhode Island, in *State v. Wheeler* 103, declined to apply the Frye standard holding instead "the law and practice of this state on the use of expert testimony has historically been based on the principle that helpfulness to the trier of fact is the most critical consideration".¹⁰⁴ The court reviewed the cases around the country, both state and federal, and noted that the majority of circuit courts that have considered admission of spectrographic evidence have decided in favor of its admission.¹⁰⁵ The court pointed out that the defendant had all the proper safeguards such as cross-examination, rebuttal experts, and the jury had the right to reject the evidence for any one of a number of reasons.¹⁰⁶

In *State v. Free*¹⁰⁷ the Court of Appeals of the State of Louisiana did not rely on the Frye test for guidance in determining the admissibility of spectrographic voice identification evidence but instead applied a balancing test set forth in *State v. Catanese*¹⁰⁸). One individual, accepted as an expert in voice identification, testified as to the theoretical and technical aspects of the spectrographic voice analysis method. No other witnesses were called to either support or show fault with the admission of the voice identification testimony. The Court of Appeals found that voice identification evidence, when offered by a competent expert and obtained through proper procedures, "is as reliable as other kinds of scientific evidence accepted routinely by courts" and "can be highly probative"¹⁰⁹. Using the Catanese balancing test the Court of Appeals found that trier of fact was likely to give almost conclusive weight to the voice identification expert's opinion, consequently, misleading the jurors. The Court of Appeals was also concerned that there were not enough experts available who could critically examine the validity of a voice identification determination in a particular case. Nine rules were suggested as a basis for which voice identification evidence could be accepted¹¹⁰). The Court of Appeals held that Catanese prohibits admission of the voice identification evidence at this time¹¹¹ and found the admission of that evidence to be harmless error.

In 1987 the Supreme Court of New Jersey again addressed the issue of admissibility of spectrographic evidence in the civil case of *Windmere v. International Insurance Company*.¹¹² In affirming the judgment of the Appellate Division, the Supreme Court of New Jersey ruled that the Appellate court's affirmation of the admission of the spectrographic evidence by the trial court was improper. The court stated the admissibility of the spectrographic voice analysis is based on the scientific technique having sufficient scientific basis to produce uniform and reasonably reliable results and contribute materially to the ascertainment of the truth ¹¹³, a standard the court admits bears "a close resemblance to the familiar Frye test".¹¹⁴ The court relies upon the "general acceptance within the professional community" to establish the scientific reliability of the voice identification process. In reaching a determination of

general acceptance, the court on a three prong test which includes; (1) the testimony of knowledgeable experts, (2) authoritative scientific literature, and (3) persuasive judicial decisions which acknowledge such general acceptance of expert testimony.¹¹⁵ The court found that none of the three prongs indicated that there was a general acceptance of spectrographic voice identification in the professional community. The court criticized the proponent experts as being too closely tied to the development of this identification analysis to represent the opinions of the community.¹¹⁶ The court found that the trial court did not undertake to resolve the issue of conflicting scientific literature and they would make no effort to resolve the conflict.¹¹⁷ The court also reviewed the judicial decisions regarding admissibility and found a split among the jurisdictions as to the reliability of the identification process.¹¹⁸

The New Jersey Supreme Court specifically limited its decision in *Windmere* excluding spectrographic voice identification evidence to the present case. The court stated that the future use of voice identification evidence "as a reasonably reliable scientific method may not be precluded forever if more thorough proofs as to reliability are introduced" ¹¹⁹ and they will "continue to await the more conclusive evidence of scientific reliability".¹²⁰

The Court of Appeals of Texas in the case of *Pope v. Texas*¹²¹ refused to address the issue of admissibility of voice identification evidence stating that "the overwhelming evidence against appellant renders this error, if any, harmless"¹²²). Justice McClung in his dissenting opinion states that the trial court did err in admitting the voice identification evidence and that the error was not harmless¹²³. He suggests that the Frye test is the proper standard for assessing the admissibility issue and that the "relevant scientific community" should be defined broadly¹²⁴. When this aspect of the test is so defined the "general acceptability" criterion is not met.

In February of 1989, the United States Court of Appeals for the Seventh Circuit affirmed the decision of the United States District Court for the Northern District of Illinois admitting spectrographic voice identification evidence in the criminal case of *United States of America v. Tamara Jo Smith*.¹²⁵ The Seventh circuit now joins the Second, Fourth and Sixth Circuits in affirming the use of spectrographic voice identification evidence.¹²⁶ The Appellate court used the Frye standard to hold expert testimony concerning spectrographic voice analysis admissible in cases where the proponent of the testimony has established a proper foundation.¹²⁷ The court noted that this technique was not one-hundred percent infallible and that the entire scientific community does not support it, however, neither infallibility nor unanimity is a precondition for general acceptance of scientific evidence.¹²⁸ The Seventh circuit found that a proper foundation had been established in that the expert testified to the theory and the technique, the accuracy of the analysis and the limitations of the process.¹²⁹ The court noted that variations from the norm result in an increase of false eliminations.¹³⁰ The jury was not likely to be misled in that they had the

opportunity to hear the recordings, see the spectrograms, hear the limitations of the process, witnessed a rigorous cross-examination of the expert and could reject the testimony of the expert.¹³¹

In *United States v. Maivia*,¹³² the United States District Court admitted spectrographic evidence after a four day hearing on the issue. The court examined the various sub- tests of the Frye test and found that spectrographic voice identification evidence met these tests. The court also noted that "inasmuch as the admissibility of spectrographic evidence to identify voices has received judicial recognition, it is no longer considered novel within the Frye test and consequently the test is inapplicable" ¹³³. The court also looked to the Federal Rules of Evidence, specifically rule 403, in deciding the admissibility of spectrographic voice identification evidence.

In affirming the order of the Appellate Division, the New York Supreme Court, in the case of *People v. Jeter*¹³⁴, concluded that the trial court was not able to properly determine that voice identification evidence is generally accepted as reliable based on case law and existing literature. The Court stated that the trial court should have held a preliminary inquiry into the reliability of voice spectrographic evidence. In the light of the other evidence, the admission of the voice identification evidence was held to be harmless error in this case.

STANDARDS OF ADMISSIBILITY

Prior to 1993 there were two main standards of admissibility which had been applied to voice identification evidence; the Frye test and the Federal Rules of Evidence (and the rules of evidence of the various states). The Frye test originated from Court of Appeals of the District of Columbia¹³⁵ in a decision rejecting admissibility of a systolic blood pressure deception test (a forerunner of the polygraph test). The court stated that admission of this novel technique was dependent on its acceptance by the scientific community.

"Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and 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".¹³⁶

Out of forty published opinions prior to 1993 deciding the admissibility of voice identification evidence, twenty-three courts applied the Frye standard or a standard very similar to Frye. Sixteen of the twenty-three courts rejected the admission of such evidence. Six of these courts held the admission of voice identification evidence by the trial court was harmless error and affirmed the conviction or judgment. Eight of the sixteen stated that although voice

identification evidence had not yet met the required standard of scientific acceptability, their decision was not intended to foreclose future admission when such standards were met. Two of these courts denied admission because they felt a single witness could not speak for the entire scientific community regarding the acceptance issue.

Seven courts applied the test and found the requirements of Frye had been met. Of the thirteen courts applying a standard of admissibility different from Frye, only one, the Free court¹³⁷, rejected voice identification evidence.

There are three problems with the Frye standard; at what point is the principle of "sufficiently established" determined, at what point is "general acceptance" reached, and what is the proper definition of "the particular field in which it belongs".

These three areas have been major stumbling blocks for the courts in deciding the issue of the admissibility of voice identification evidence due to the small number of voice scientists who have performed research in this field. The trial court in *People v. Siervonti* ¹³⁸ noted the lack of research in this area saying "one only wishes that the last twelve years had been spent in research and not in attempting to get the method into the courts".¹³⁹

The Frye test has been criticized as not being the appropriate test to use for the admission of voice identification evidence. This standard was established and applied to the admission of a type of evidence which is very different from voice identification. In Frye the court was concerned with the admission of a test designed to determine if a person was telling the truth or not. This type of evidence invades the province of the finder of fact. Voice identification evidence belongs in the general classification of identification evidence which does not impinge on the role of the finder of fact. As such it shares common traits with the other identification sciences of fingerprinting, ballistics, handwriting, and fiber, serum and substance identification.

Another criticism of the application of the Frye test as the standard for admission of voice identification evidence is that general acceptance by the scientific community is the proper condition for taking of judicial notice of scientific facts. McCormick states that general scientific acceptance is a proper condition for taking judicial notice of scientific facts, but not a criterion for the admissibility of scientific evidence.¹⁴⁰

The court in *Reed v. State* ¹⁴¹ seemed to note this difference between the standard for the taking of judicial notice and that for admission of evidence such as voice identification. The court said that validity and reliability may be so broadly accepted in the scientific community that the court may take judicial notice of it. If it can not be judicially noticed then the reliability must be demonstrated before it can be admitted.¹⁴² The court then applied the Frye test,

general acceptance by the scientific community, to determine reliability and thus, admissibility.

Scientific evidence has long been admitted before it was judicially noticed, as with the case of fingerprints. The admission of fingerprint identification evidence was first challenged in the case of *People v. Jennings*¹⁴³ in 1911. The court in *Jennings* allowed the admission of fingerprint evidence saying "whatever tends to prove any material fact is relevant and competent".¹⁴⁴ It was not until thirty-three years later that fingerprint evidence was first judicially noticed.¹⁴⁵

The majority of courts which have decided the issue of admissibility in favor of allowing voice identification into the courtroom have used similar standards which permit the finder of fact to hear the evidence and determine the proper weight to be assigned to it. Their logic runs parallel to the Federal Rules of Evidence which state that all relevant evidence is admissible with the word "relevant" being defined as evidence tending to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.¹⁴⁶ A qualified expert may testify to his opinion if such opinion will assist the trier of fact in better understanding the evidence.¹⁴⁷

Many of the courts which have upheld the admission of voice identification evidence have done so because the trial court had set up a number of precautions to insure the evidence was viewed in its proper light. These precautions include allowing the jury to see the spectrograms of the voices in question, allowing the jury to hear the recordings from which the spectrograms were produced, the expert's qualifications and opinions as well as the reliability of the equipment and technique are subject to scrutiny by the other side, the availability of competent witnesses to expose limitations in the process, and instructions to the jury that they were free to assign whatever weight, if any, to the evidence they felt it deserved.

The United States Supreme Court in 1993 changed the long-standing law of admissibility of scientific expert evidence by rejecting the *Frye* test as inconsistent with the Federal Rules of Evidence in the case of *Daubert v. Merrell Dow Pharmaceuticals*¹⁴⁸. The Court held that the Federal Rules of Evidence and not *Frye* were the standard for determining admissibility of expert scientific testimony. *Frye*'s "general acceptance" test was superseded by the Federal Rules' adoption. Rule 702 is the appropriate standard to assess the admissibility of scientific evidence. The Court derived a reliability test from Rule 702.

In order to qualify a scientific knowledge, an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation - i.e., good grounds, based on what is known. In short, the requirement that an expert's testimony pertain to scientific knowledge establishes a standard of evidentiary reliability¹⁴⁹

The Daubert decision concerns statutory law and not constitutional law. The Court held that the Federal Rules, not Frye, govern admissibility.. The only Federal Circuit to reject spectrographic voice analysis has been the District of Columbia. Daubert may cause the District of Columbia to change its stance the next time such evidence is introduced.

Since Daubert is not binding on the states, it will be difficult to determine just how much impact Daubert will have on the admissibility standards of the states. Many states have adopted evidence rules based on the Federal Rules of Evidence and may not be effected by this holding. Other states which have adopted the Frye test will have to decide to either continue following Frye or change their standard to Daubert. The Arizona Supreme Court declined to follow Daubert saying that it was "not bound by the United States Supreme Court's non-constitutional construction of the Federal Rules of Evidence when we construe the Arizona Rules of Evidence."¹⁵⁰

RESEARCH STUDIES

The studies that have been produced over the years have run the gambit in type, parameter, and result. A quick review of the available published data would leave one with the impression that the spectrographic method of voice identification was only somewhat more accurate than flipping a coin. The diversity of the relatively low number of studies and the range of results has only added to the confusion as to the reliability and validity of this method of identification. When one takes the time and expends the effort to analyze the studies in this field, a very different conclusion becomes evident. When the individual parameters of the studies are taken into account, who was being evaluated, what information was given to the examiner to assess, and what limitations were placed on the examiner's conclusions, a much clearer picture of the accuracy of the spectrographic voice identification method develops. The picture is not one of a marginally accurate technique but rather a picture that clearly shows that a properly trained and experienced examiner, adhering to internationally accepted standards will produce a highly accurate result. The studies also show that as the level of training diminishes and/or the conclusions an examiner may reach are artificially limited, the error rate goes up dramatically.

The training for accurately performing the spectrographic voice identification method has been established as requiring completion of (1) a formal course of study, usually 2 to 4 weeks duration, in the basics of spectrographic analysis, (2) two years of study completing 100 voice comparison cases, usually in a one-to-one relationship with a recognized expert, (3) examination by a board of experts in the field of spectrographic voice identification analysis.

For the most accurate results from the spectrographic voice identification method, a professional examiner (1) will require the original recordings or the best quality re-recordings if the original is not available; (2) will perform a critical

aural review of the suspect and known recordings; (3) will produce sound spectrograms of the comparable words and phrases; (4) will produce a comparison recording juxtaposing the known and unknown speech samples; (5) will evaluate the evidence and classify the results into one of five standard categories [1 - positive identification, 2 - probable identification, 3. - positive elimination, 4 - probable elimination, and 5 - no decision]. The final decision is reached through a combined process of aural and visual examination.

It is important to remember that the spectrographic method of voice identification is a process that interweaves the visual analysis of the sound spectrograms with the critical aural examination of the sounds being viewed. Taking the results from all of the studies produced shows that if the examiner's ability to analyze both the graphic representations of the voice and the aural cues found in the recordings is limited or restricted, accuracy suffers. Likewise, the amount of training has a direct bearing on the level of accuracy of the results.

In a survey of 18 studies¹⁵¹ of the accuracy of the spectrographic voice identification method, the results fall into two categories; those with proper training, using standard procedures produce very accurate results, whereas those with inadequate training, using limited analysis methods, produce inaccurate results.

In a study¹⁵² in 1975 authored by Lt. L. Smrkovski of the Voice Identification Unit of the Michigan State police, error rates in voice identification analysis comparisons, based on three levels of training and experience, were evaluated. The following table summarizes the results of that study.

Error type Novice Trainee Professional

False Ident. 5.0% 0.0% 0.0%

False Elim. 25.0% 0.0% 0.0%

No Decision 2.5% 2.5% 7.5%

Lt. Smrkovski's results show that proper training is essential. The fact that his results show a higher no decision rate among the professional examiners than the trainee examiners may indicate that the professional is a bit more cautious in his analysis than the trainee.

Mark Greenwald, in his 1979 thesis¹⁵³ for his M.A. degree at Michigan State University, studied the performance of three professional examiners (each with eight years experience) and five trainees (each with less than two years experience) using standard spectrographic voice identification methods (visual and aural) and result classifications. Greenwald found that the professional examiners produced no errors when using full frequency bandwidth recordings.

When the frequency band width was restricted, the professional examiners still produced no errors, but did increase their percentage of no decision classifications. Greenwald also found that the training level was an important factor and that the trainees in this study had an error rate of 6.1% for false identifications in the restricted frequency bandwidth trials.

In 1986, the Federal Bureau of Investigation published a survey of two thousand voice identification comparisons made by FBI examiners¹⁵⁴. This survey was based on 2000 forensic comparisons completed over a period of fifteen years, under actual law enforcement conditions, by FBI examiners.¹⁵⁵

The examiners had a minimum of two years experience, completed over 100 actual cases, completed a basic two week training course and received formal approval by other trained examiners.¹⁵⁶

The results of the survey are depicted in the chart 157 below.

DECISIONS NUMBER PERCENT(%)

No or low confidence 1304 65.2

Eliminations 378 18.9

Identifications 318 15.9

ERRORS

False eliminations 2 0.53

False identification 1 0.31

The FBI results are consistent with the Smrkovski study in that properly trained examiners, utilizing the full range of procedures, produce quite accurate results.

By way of contrast, the 1976 study¹⁵⁸ by Alan Reich used four speech science graduate students with previous experience with speech spectrograms (but untrained in spectrographic voice identification analysis) to examine, using visual comparison only, nine excerpted words. This study produced an accuracy rate in the undisguised trials of 56.67%. When disguise was introduced into this study paradigm the accuracy rate decreased significantly.

Taken as a whole the 18 studies support the conclusion that accurate results will be obtained only through the combined use of the aural and visual components of the spectrographic voice identification method as performed by a properly trained examiner adhering to the established standards. Those studies with poor accuracy results are important in that they demonstrate the weaknesses of

improperly performed examinations that do not adhere to the internationally accepted professional standards.

A large part of the debate over the admissibility of spectrographic voice identification analysis in the courts appears due to the fact that the parameters of these studies have not adequately been demonstrated to the courts in the necessary detail which would allow the courts to examine the overall meaning of these studies. Many of these studies look at only one or two aspects of the spectrographic voice identification method. Frequently the results of these restricted scope studies have been misapplied to the entire spectrographic voice identification method resulting in inaccurate information being used as the basis for deciding the admissibility of spectrographic voice identification analysis. It is important to provide an accurate picture of all the studies so the courts will have the foundational information necessary to make an informed decision regarding the admissibility of spectrographic voice identification analysis.

CONCLUSION

The technique of voice identification by means of aural and spectrographic comparison is still an unsettled topic in law. Although the spectrographic voice identification method has progressed greatly since it was first introduced to a court of law back in the mid 1960's, it still faces stiff resistance on the issue of admissibility in the courts today. One of the reasons for such opposition regarding admissibility is that the method has evolved greatly since its initial application. Court decisions based on early methods of voice identification analysis are not applicable to the methods used today. No longer are voices compared on the basis of a limited group of key words. Today's aural/spectrographic voice identification method takes advantage of the latest in technological advancements and interweaves several analyses into one procedure to produce an accurate opinion as to the identity of a voice. This modern technique combines the experience of a trained examiner performing the visual analysis of the spectrograms and aural analysis of the recordings with the use of the latest instruments modern technology has to offer, all in a standardized methodology to assure reliability. Court decisions reviewing the early voice identification cases may not be relevant to present day cases because the older decisions were based on less sophisticated procedures. Most of the courts which have rejected admission have been aware of continuing work in this field and have specifically left the door open as to future admissibility.

Proper presentation and explanation of the research pertaining to spectrographic voice identification analysis will allow the courts to better understand the accuracy and reliability of the spectrographic voice identification method. When the research is properly presented, the studies show that properly trained individuals, using standard methodology, produce accurate results.

The current trends in the admissibility issue of voice identification evidence indicate that courts are more willing to allow the evidence into the courtroom when a proper foundation has been established which then allows the trier of fact to determine the weight to be assigned to the evidence.

TABLE OF CASES

1. FRYE v US 293 F 1013 (D.C. Ct. App. 1923)
2. US v WRIGHT 37 CMR 447 (1967)
3. STATE v CARY 230 A.2d 384 (N.J. 1967)
4. STATE v CARY 239 A.2d 680 (N.J.Super. 1968)
5. PEOPLE v KING 266 C.A.2d 437 (1968)
6. STATE v CARY 250 A.2d 15 (N.J. 1969)
7. STATE v CARY 264 A.2d 209 (N.J. 1970)
8. STATE EX REL. TRIMBLE v HEDMAN 192 N.W.2d 432 (Minn. 1971)
9. US v RAYMOND 337 F.Supp. 641 (DCDC 1972)
10. WORLEY v STATE 263 So.2d 613 (Fla. 1972)
11. ALEA v STATE 265 So.2d 96 (Fla. 1972)
12. US v ASKINS 351 F.Supp. 408 (1972)
13. STATE v ANDRETTA 296 A2d 644 (N.J. 1972)
14. HODO v SUPERIOR COURT 30 C.A.3d 778 (Calif. 1973)
15. PEOPLE v CHAPTER 13 CrL 2479 (Calif. 1973)
16. US v SAMPLE 378 F.Supp. 44 (Penn. 1974)
17. US v ADDISON 498 F.2d 741 (DCDC 1974)
18. PEOPLE v LAW 40 C.A.3d 69 (Calif. 1974)
19. US v FRANKS 511 F.2d 25 (6th Cir. 1975)
20. COMMONWEALTH v LYKUS 327 N.E.2d 671 (Mass. 1975)

21. COMMONWEALTH v VITELLO 327 N.E.2d 819 (Mass. 1975)
22. STATE v OLDERMAN 336 N.E.2d 442 (Oh. 1975)
23. US v BALLER 519 F.2d 463 (4th Cir. 1975)
24. US v JENKINS 525 F.2d 819 (6th Cir. 1975)
25. PEOPLE v ROGERS 385 N.Y.S.2d 228 (N.Y. 1976)
26. PEOPLE v KELLY 549 P.2d 1240 (Calif. 1976)
27. US v MCDANIEL 538 F.2d 408 (D.C. Cir 1976)
28. COMMONWEALTH v TOPA 369 A.2d 1277 (Penn. 1977)
29. PEOPLE v EVANS 393 N.Y.S.2d 674 (1977)
30. PEOPLE v TOBEY 257 N.W.2d 537 (Mich. 1977)
31. US v WILLIAMS 443 F.Supp. 269 (S.D.N.Y. 1977)
32. PEOPLE v COLLINS 405 N.Y.S.2d 365 (1978)
33. BROWN v US 384 A.2d 647 (D.C.C.A. 1978)
34. D'ARC v D'ARC 157 N.J.Super. 553 (1978)
35. STATE v WILLIAMS 388 A.2d 500 (Me. 1978)
36. REED v STATE 391 A.2d 364 (Md. 1978)
37. US v WILLIAMS 583 F.2d 1194 (2nd Cir. 1978)
38. PEOPLE v BEIN 453 N.Y.S.2d 343 (N.Y. 1982)
39. STATE v WILLIAMS 4 OHIO ST.3d 53 (1983)
40. CORNETT v STATE 450 N.E.2d 498 (Ind. 1983)
41. STATE v GORTAREZ 686 P.2d 1224 (Ar. 1984)
42. PEOPLE v SIERVONTI, unpublished, Municipal Court of the Chico Judicial District, State of California (1985)

43. STATE v WHEELER 496 A.2d 1382 (R.I. 1985)
44. STATE v. FREE 493 So.2d 781 (La., 1986)
45. POPE v. STATE of TEXAS 756 S.W.2d 401 (Texas 1988)
46. UNITED STATES v. MAIVIA 728 F. Supp 1471 (D. Hawaii, 1990)
47. PEOPLE v. JETER 80 N.Y. 818 (NY 1992)
48. DAUBERT v. MERRELL DOW PHARMACEUTICALS 113 S. Ct. 2786 (1993)

APPENDIX 1

The following are summaries of studies of spectrographic voice identification and an FBI survey of forensic cases..

Greenwald, M., "The Effects of Decreased Frequency Bandwidth on Speaker Identification by Aural and Spectrographic Examination of Speech Samples", Master Thesis, Michigan State University, 1979

Hall, M. C., "Spectrographic Analysis of Interspeaker and Intraspeaker variables of Professional Mimicry", Master Thesis, Michigan State University, 1975

Hazen, B., "Effects of Different Phonetic Contexts on Spectrographic Speaker Identification", 54 J. Acoust. Soc. Am. 650, 1973

Hollien, H., & McGlone, R., "The Effect of Disguise on Voiceprint Identification", In the Proceedings of the Carnahan Crime Countermeasures Conference, University of Kentucky, University of Kentucky Press, Lexington, KY, 1976

Kersta, L. G., "Voiceprint Identification", 196 Nature Magazine 1253, Dec. 29, 1962

Reich, et al., "Effects of Selected Vocal Disguises upon Spectrographic Speaker Identification", 60 J. Acoust. Soc. Am. 919, 1976

Reich & Duke, "Effects of selected vocal disguises upon speaker identification by listening", 66 J. Acoust. Soc. Am. 1023, 1979

Smrkovski, L. L., "Collaborative Study of Speaker Identification by the Voiceprint Method", 58 J. AOAC 453, 1975

Smrkovski, L. L., "Study of Speaker Identification by Aural and Visual Examination of Non-Contemporary Speech Samples", 59 J. AOAC 927, 1976

Stevens, et al., "Speaker Authentication and Identification: A Comparison of Spectrographic and Auditory Presentations of Speech Material", 44 J. Acoust. Soc. Am. 1596, 1968

Tosi, et al., "Experiment on Voice Identification", 15 J. Acoust. Soc. Am. 2030, 1972

Tosi & Greenwald, "Voice Identification by Subjective Methods of Minority Group Voices", Paper presented at the 6th Meeting of the International Association of Voice Identification, New Orleans, La., 1978

Young, M. A., & Campbell, R. A., "Effects of Context on Talker Identification", 42 Acoust. Soc. Am. 1250, 1967

KERSTA

1962

Examiners: 8 high school girls Training duration: 1 week

Method: visual Speaker population: 123

Number of words: 10 words excerpted from sentences Context type: isolated random context

Temporal sequence: contemporary Type of trial: closed

Total number of trials: 2000

Type of decision: forced decisions limited sample limited time random context no aural examination examiners lacked sufficient experience Results: closed trials range of errors for false ID - 0.35 to 1.0% 10 words excerpted 0.00 to 2.0%

YOUNG & CAMPBELL

1967

Examiners: 7 PhD candidates in ASC 3 assistant professors in ASC Training duration: 1 week

Method: visual Speaker population: 5 adult males

Number of words: 2 words (you/it) in isolation & excerpted from 4 short sentences Context type: 1 word in isolation 2 words from random context

Temporal sequence: contemporary Type of trial: closed

Total number of trials: 1046

Type of decision: forced decisions limited sample random context no aural examination examiners not trained Results: closed trials range of errors for false ID - "you" in isolation: 10.4 to 18.0% 'it' in isolation: 22.7 to 33.0% "you/it" from random context in trial 1 of 15: mean error: 62.7%

STEVENS

1968

Examiners: college students 6 in the open trials 4 in the closed trials Training duration: 1 week

Method: aural vs. visual but not combined Speaker population: 24 males

Number of words: catalogue of 11 words in different random order - only 1 word used in most trials Context type: 1 to 4 words

Temporal sequence: non-contemporary (1 week) Type of trial: closed & open

Total number of trials: 216

Type of decision: forced decisions limited sample (1 to 4 words) random context no aural examination examiners not trained Results: open trials: range of errors for false ID for 4 examiners/1 word visual trials - 31.0 to 47.0% aural trials - 6.0 to 8.0% closed trials: range of errors for false ID - 1 - 4 discrete words visual trials 20.0 to 30.0% aural trials 5.0 to 18.0%

TOSI ET AL

1968 - 1970

Examiners: 29 of various backgrounds Training duration: 1 month

Method: visual Speaker population: 250 males randomly selected from a population of 25,000

Number of words: 6 & 9 words Context type: isolated, fixed and random context

Temporal sequence: contemporary & noncontemporary (1 month) Type of trial: closed & open

Total number of trials: 34,992

Type of decision: forced decisions, but allowed to rate confidence level limited sample limited time no aural examination examiners lacked sufficient experience Results: range of errors for all trials false ID - 0.51 to 6.43% when only 'fairly & almost' certain decisions are combined, the error of false ID reduces to 2.4%

HAZEN

1972

Examiners: college students (7 panels of 2) Training duration: 5 lectures and 3 practice sessions

Method: visual Speaker population: 60 males

Number of words: 5 words in the same context, 5 words physically excerpted from random conversation Context type: fixed and random context

Temporal sequence: contemporary Type of trial: closed & open

Total number of trials: 280

Type of decision: forced decisions limited sample (5 words) no aural examination random & fixed context examiners lacked sufficient experience used the most dissimilar spectrographic utterances compared sounds from totally different words studying changing phonetic context examiners could not evaluate effects of coarticulation due to questionable word boundaries Results: closed trials errors for false ID - fixed context range:10.0 to 30.0% mean: 20.0% random context range:50.0 to 90.0% mean: 74.29% open trials errors for false ID - fixed context range:16 to 66% mean: 42.86% random context range:66 to 100% mean: 83%

SMRKOVSKI

1974

Examiners: 7 police & private Training duration: more than 2 years experience/less than 2 years experience

Method: combined aural and visual Speaker population: 7 male & female

Number of words: 38 to 54 words Context type: fixed context

Temporal sequence: noncontemporary (1 week) Type of trial: open

Total number of trials: 84

Type of decision: no forced decisions allowed 1 to 5 conclusions no limited time aural & visual examination trained and experienced examiners Results: open trials trainees w/less than 2 yr experience: false ID - 0.0% false elim. 5.0% no decision 25.0% 0.35 to 1.0% examiners w/more than 2 yr experience: false ID - 0.0% false elim. 0.0% no decision 22.0%

SMRKOVSKI

1975

Examiners: 12 scientists, police and private Training duration: novice: no training trainee: < 2 yr Professional: > 2 yr

Method: combined visual and visual Speaker population: 20 male & female

Number of words: 9 words Context type: fixed context

Temporal sequence: noncontemporary Type of trial: open

Total number of trials: 120

Type of decision: no forced decisions allowed 1 to 5 conclusions no limited time aural & visual examination compared words in context - trainees, novices and experienced examiners Results: open trials: errors novices false ID 5.0% false elim 25.0% no decision 2.5% trainee false ID 0.0% false elim 0.0% no decision 2.5% Professional false ID 0.0% false elim 0.0% no decision 7.5% HALL

1975

Examiners: 4 professional and 20 college graduates Training duration: IAVI certified voice identification examiner

Method: combined visual and visual / visual only Speaker population: professional mimic and 6 celebrity voices

Number of words: mimic (mean of 25 sec.), celebrities (mean of 35 min.) Context type: quasi-fixed and random context

Temporal sequence: contemporary/ noncontemporary Type of trial: open

Total number of trials: aural (20/examiner) visual (200/examiner) Type of decision: same, different or undecided 5 IAVA classifications

Results: Interspeaker variability does not exist between a mimicked, disguised voice and the nature voice of the subject mimicked. Intraspeaker variabilities are minute and not significant when comparing mimics' voice and the nature voice of

the mimic. Aurally: The smaller signal-to-noise ratio within the recording and the more similar the context, the greater the percentage of accuracy in distinguishing between speakers. AURAL EXAMINATION: Grand means: RIGHT WRONG UNDEC. Grad. students 0.74 0.18 0.08 Professional 0.92 0.082 0.0

HOLLIEN/McCLONE

1975-76

Examiners: 5 faculty 1 graduate student Training duration: "the authors were familiar with the 'voiceprint' method of speaker identification"

Method: visual only (spectrograms were cut & mounted) Speaker population: 25 faculty and graduate students of the University of Florida

Number of words: 7 words Context type: "I do not set the same store"

Temporal sequence: contemporary Type of trial: open

Total number of trials: 25/examiner

Type of decision: record a match/ indicate none was possible Results: ". . . even skilled auditors such as these were unable to match correctly the disguised speech to the reference (normal) samples as much as 25% of the time . . . these groups were able to disguise their voices in such manners that their identification by the 'voiceprint' technique became little more than a matter of chance."

REICH ET AL

1976

Examiners: 2 PhD candidates in speech science 2 PhD candidates in speech pathology Training duration: 3 courses in speech science plus previous experience with speech spectrograms: 4 weeks at 10-15 hr/wk

Method: visual only (words excerpted and mounted) Speaker population: 40 adult males (mean: 27.3 yrs)

Number of words: 9 words Context type: fixed context

Temporal sequence: noncontemporary Type of trial: open

Total number of trials: 105 (7 matching tasks w/15 known & 15 unknown)

Type of decision: 1 to 5 certainty scale Results: The examiners were able to match speakers with a moderate degree of accuracy (55.67%) when there was

no attempt to vocally disguise. Disguised speech significantly interfered with speaker identification. Further research is needed . . . in which the examiners may listen to the voice as well as view the spectrograms.

ROTHMAN

1977

Examiners: 30 listeners 6 visual examiners Training duration: none

Method: Study I: Aural Study II: Visual (0 to 8kHz) Speaker population: 12

Number of words: four - 2 second speech segments Context type: random context

Temporal sequence: contemporary/ noncontemporary (1wk) Type of trial: open

Total number of trials: 5 visual 38 aural

Type of decision: same/different for each contemporary and noncontemporary
Results: 94% correct identifications were obtained for contemporary speech segments. 42% correct identifications were obtained for noncontemporary speech segments. 58.45% correct identifications were obtained when comparing different speakers. All examiners in pretest visual achieved 100% correct matching. Aural method is clearly superior to the spectrographic or 'voiceprint' method

McGLONE, HOLLIEN & HOLLIEN

1977

Examiners: 4 phoneticians Training duration: experienced

Method: visual measurement of formant fundamental frequency to obtain for Speaker population: 23 adult males

Number of words: 7 words ("I do not set the same store" Context type: fixed (normal & disguised) context

Temporal sequence: contemporary Type of trial:

Total number of trials: 46/phonetician

Type of decision: Results: A great amount of variability in the f_0 was found between normal and disguised speech. The mean bandwidth differences (f_1 , f_2 ,

f3) for the group were large and also demonstrated considerable variability. Phonetic means also differed.

HOULIHAN - Study I

1977

Examiners: 21 undergraduate students Training duration: series of lectures & discussions on phonetics, acoustics, and sound spectrography and speaker identification

Method: visual only Speaker population: 9 female, 5 male undergraduates - homogenous age and geographic background

Number of words: 9 words Context type: fixed context: 5 voice conditions (normal, lowered, falsetto, whispered and muffled)

Temporal sequence: contemporary Type of trial: open

Total number of trials: 18 matches

Type of decision: same/different Results: correct identifications: F- voice M-voice normal 100% 95% lowered 85% 95% falsetto 95% 90% whispered 5% 98% muffled 75% 100% range: 39 to 70% correct mean: 58.8% Std.D.: 8.7%

HOULIHAN - Study II

1977

Examiners: 7 students from Experimental phonetics Training duration: completion of Exp. I with feedback

Method: visual only Speaker population: 8 female, 8 male (mean age: 25.3 yrs)

Number of words: 8 words Context type: fixed context: "There's a bomb in the main post office"

Temporal sequence: contemporary Type of trial: closed

Total number of trials: 16/examiner

Type of decision: instructed to consider the sets in a particular order. All examiners considered undisguised before disguised Results: correct identifications: F-voice M-voice normal 71% 100% lowered 85% 100% falsetto 100% 67% whispered 71% 71% muffled 85% 100% The results suggest that minimally trained examiners have little difficulty with spectrographic identification

in closed, contemporary, undisguised trials. Results do not suggest that female voices are more difficult to identify than male voices.

TOSI ET AL

1979

Examiners: professional and students Training duration: IAVA certified voice examiners and 2 weeks of training, respectively

Method: aural only, visual only and aural/visual combined Speaker population: Chicano (25 female and 25 male)

Number of words: four sentences approximately 2.4 seconds in Spanish Context type: fixed context

Temporal sequence: noncontemporary Type of trial: open - randomized

Total number of trials: 600/examiner

Type of decision: same, different, no opinion. qualified percentage of self-confidence from 51 to 100% Results: Student and Professional examiners for errors of elimination and identification had a mean percentile greater for noisy samples than for quiet samples, however, professional examiners errors were due to aural only examinations whereas spectrographic/aural examinations produced 0.0% errors. The 'no opinion' option was used more by professional examiners.

REICH

1979

Examiners: 24 undergraduate students, 3 doctoral students, 3 professors of Speech and Hearing Science Training duration: brief lecture; 120 discrimination trials identical to the experiment

Method: aural only Speaker population: 40 adult males (mean age: 27.3 yrs)

Number of words: 9 words (it, is, on, you, and, the, I, to, me) Context type: fixed context

Temporal sequence: noncontemporary (2 weeks +) Type of trial: open

Total number of trials: 18 matches

Type of decision: same/different (1 to 5 certainty) Results: Both groups were able to discriminate speakers with moderately high degrees of accuracy, 92% correct for undisguised. Disguised trials ranged from 59 to 81% depending on the disguise. Recommended further research to study the combined aural/spectrographic method.

GREENWALD

1979

Examiners: 3 professional, 5 trainees (less than 2 years experience) Training duration: professionals: 8 yrs each trainees: < 2 yrs

Method: aural only, visual only and aural/visual combined Speaker population: 12 female, 12 male; American Midwest dialect

Number of words: 24 words Context type: fixed context

Temporal sequence: noncontemporary Type of trial: open

Total number of trials: 192 discrimination types Type of decision: the five IAVI alternatives

Results: Professional examiners produced no errors of false identification or elimination. 1536 decisions by all eight examiners. Effect of restricted bandwidths (240-2K, 240-2.5K, 240-3K, and 240-4K) does not increase the errors but does increase the percentage of 'no decisions'. Training of the examiner is very important on error rate. Trainees produced errors as follows: 6.1% false identification and 4.1% false elimination for all trials. However, at 240-4khz., 0.0% errors of false identification of elimination.

KOENIG - FBI SURVEY

1986

Examiners: Federal Bureau of Investigation voice identification examiners Training duration: minimum of 2 yrs experience, completion of at least 100 actual voice comparison cases, formal approval by other trained examiners

Method: combined aural/visual method Speaker population: actual criminal cases

Number of words: varied with each case Context type:

Temporal sequence: noncontemporary Type of trial: open

Total number of trials: 2000 forensic comparisons

Type of decision: very similar very dissimilar no decision (low confidence)
Results: number percent no/low conf. 1304 65.2 elimination 378 18.9
identification 318 15.9 errors false elim. 2 0.53 false id. 1 0.31
