

THE VOICEPRINT TECHNIQUE: ITS STRUCTURE AND RELIABILITY

*Bernard S. Kamine**

INTRODUCTION

Identification of individuals by the sound of their voices has long been an accepted courtroom practice.¹ It has been accomplished directly both in the courtroom² and extra-judicially,³ as well as indirectly with sound recordings.⁴ Voice identifications are essential to authenticating sound recordings for introduction as evidence,⁵ and are frequently the most conclusive evidence in certain types of criminal prosecutions such as those involving obscene phone calls. Until recently all voice identifications were made by the human ear, by someone familiar with the sound of the voice being identified. Although generally accepted by the courts, it has been recognized that such identifications are occasionally quite unreliable. At least one court has suggested that "a highly desirable aid to judicial determinations of truth"⁶ would be a scientific method of voice identification which is not subject to human frailties.

The voiceprint technique is reputed to be such a method.⁷ Despite a number of scientists and researchers in the speech field who have disputed the validity of the technique, two trial courts⁸ and one appellate tribunal,⁹ as well as a number of law enforcement and investigative agencies,¹⁰ have been convinced of its reliability.

* Deputy Attorney General, State of California; Member, State Bar of California, State Bar of Colorado. B.A. Univ. of Denver, 1965; LL.B. Harvard Law School, 1968. The views expressed in this article are to be attributed to the author and not to the Office of the Attorney General, State of California.

1. *E.g.*, *People v. Sica*, 112 Cal. App. 2d 574, 247 P.2d 72 (1952); *see also* 2 WIGMORE, EVIDENCE § 660 (3rd ed. 1940) and Annot., 70 A.L.R.2d 995 (1960).

2. *E.g.*, *Johnson v. Commonwealth*, 115 Pa. 369, 9 A. 78 (1887).

3. *E.g.*, *Eidson v. United States*, 272 F.2d 684 (10th Cir. 1959).

4. *See Osborn v. United States*, 385 U.S. 323, *reh. den.* 386 U.S. 938 (1966).

5. *E.g.*, *United States v. McKeever*, 169 F. Supp. 426 (S.D.N.Y. 1958).

6. *State v. Cary*, 49 N.J. 343, 351, 230 A.2d 384, 388 (1967).

7. L. Kersta, *Voiceprint Identification*, 196 NATURE 1253 (1962) [hereinafter cited as Kersta, NATURE].

8. *People v. Straehle* (Westchester County, N.Y.) in N.Y. Times, Apr. 12, 1966, at 1, col. 2. and *People v. King* (Los Angeles County, Cal.) L.A. Times, West Magazine, Mar. 26, 1967, at 12.

9. *United States v. Wright*, 17 U.S.C.M.A. 183, 37 C.M.R. 447 (1967).

10. *See Coon, Voiceprint Identification Goes to Work*, 11 POLICE 68 (1966); Coon, *Identification by Voice*, 1964 CRIME LAB. 232.

This article will evaluate the potential of the voiceprint technique as a legal tool, the theory on which it relies, the proofs offered to support it, and the arguments advanced against it. In addition, the mechanics of the sound spectrograph, the machine used for voiceprint technique identifications, will be surveyed along with the fundamental theories of sound and acoustics which it employs. Finally, the evidentiary and constitutional issues arising with utilization of the technique will be considered. However, the peripheral involvement of wiretapping, eavesdropping and related fourth and fifth amendment questions will be excluded.

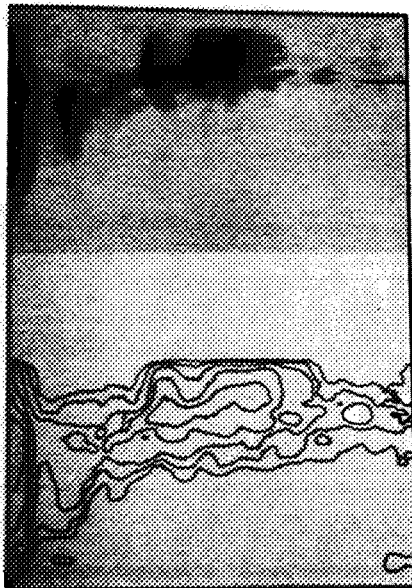


Figure 1. TWO FORMS OF VOICEPRINTS. Shown here are the two basic forms of voiceprints. The upper is the bar form, the spectrogram normally used for voiceprints, and the lower is the contour form.

I. THE VOICEPRINT TECHNIQUE

The voiceprint technique relies upon a process of pattern-matching and upon the theory of invariant speech. Pattern-matching will be considered first, in conjunction with related problems posed by the sound spectrograph,¹¹ the mechanical device used to analyze the patterns of speech.

11. See text accompanying notes 28-39 *infra*.

Pattern-Matching and the Spectrograph

Pattern-matching is a standard method of establishing identity between two objects.¹² It is employed, for example, to identify fingerprint samples,¹³ bullets fired from the same gun¹⁴ and samples of typewriting and handwriting.¹⁵ Whenever used, the features of one sample are compared with those of another, and, if a sufficient number of identical features appear on both, identity is confirmed. However, a prerequisite is that each of those features or the particular combination of those features must not appear on non-identical samples. Whether the voiceprint technique satisfies this requirement is considered in the next section.

The patterns matched by the voiceprint technique are steady-state portions of the spectrogram for the selected speech sound.¹⁶ A sound in context is influenced by both the preceding and succeeding sounds; the tongue, the jaw, and the lips move from the position for the preceding sound to that position required for the next desired sound, then move again from its position to that for the succeeding sound. The segment of the spectrogram, representing the static position of the articulators for the desired sound, is the steady-state portion.¹⁷ It must be excerpted since the transitional spectrogram segments between particular sounds vary materially due to the differences in the preceding and succeeding speech sounds.¹⁸ Syllables spoken in isolation are the equivalent of a steady-state portion.

It is important to note that the voiceprint technique can utilize more than just the steady-state portion of the spectrogram,¹⁹ and, for reasons to be given later, it is better to use identical short phrases if two samples can legally be acquired. Through use of selected excerpts, the voiceprint technique employs a pattern-matching system similar to that

12. See 2 WIGMORE, EVIDENCE §§ 410-12 (3rd ed. 1940).

13. See *e.g.*, *People v. Jennings*, 252 Ill. 534, 96 N.E. 1077 (1911).

14. See *e.g.*, *Evans v. Commonwealth*, 230 Ky. 411, 19 S.W.2d 1091 (1929).

15. A. OSBORN, QUESTIONED DOCUMENTS (2d ed. 1944).

16. L. Kersta, *Voiceprint Infallibility* (unpublished paper presented to the Acoustical Society of America meeting Nov. 7, 1962) [hereinafter cited as *Kersta. Infallibility*].

17. Interview with Lawrence G. Kersta, President of Voiceprint Laboratories, in Somerville, N.J., Jan. 30, 1968. [hereinafter cited as *Kersta interview*].

18. For an indication of the problem, compare the movement of the articulators when pronouncing *top* and *cog*, both of which have the same vowel in the middle.

19. Kersta, *Infallibility*, *supra* note 16.

used for fingerprints or ballistics. The expert will exclude the effect of the duration of the sound because of its variability,²⁰ and then will seek comparable features in both samples. The first step of the expert is to compare the relationship among the various resonance bars²¹ (see Fig. 1) at the beginning of the time axis.²² He continues this matching process along the time axis at significant points comparing the frequency/intensity patterns, widths, shapes, slopes, mean frequencies and separations of the bars, as well as the frequency/intensity patterns of incomplete resonance bars, the large dark patterns created by noise consonants, and the pattern of the vertical lines,²³ called striations.

As with fingerprint and ballistics identifications, the voiceprint technique disregards differences among the voice samples, only noting points of similarity. For positive identification the developer of the technique has suggested a minimum of 16 points of similarity.²⁴ This is analogous to fingerprints²⁵ which, in one California case, have been found to yield a presumption of positive identification if a stipulated minimum number of points can be discovered.²⁶ However, fingerprint minimums apply to comparisons between individual fingers, not all ten; whereas, the suggested minimum for the voiceprint technique applies to the entire speech sample, not just the particular compared excerpts. A critical question arises over this proposed use of an unlimited number of sounds to find a stipulated minimum number of points. With large enough speech

20. O. Tosi, An Evaluation of the Kersta Method of Voice Identification (unpublished paper being submitted for publication by the author). The author is a professor in the Dept. of Audiology and Speech Science at Michigan State University and is participating in research on the voiceprint technique for the Michigan State Police and the U.S. Dept. of Justice. [hereinafter cited as Tosi, Evaluation].

21. See text accompanying note 42 *infra*.

22. Kersta interview, *supra* note 17.

23. Tosi, Evaluation, *supra* note 20.

24. Kersta interview, *supra* note 17. No explanation is given for the selection of 16 points; no less than 20 points, though, have been found in any case so far.

When steady-state portions are excerpted from normal speech they are rarely even half the width of the spectrograms shown, hence the discovery of points of identity becomes considerably more difficult.

25. W. Turner, *Spectrogram Voice Identification*, 19 AM. JUR. PROOF OF FACTS 423, 431 (1967).

26. *People v. Collins*, 117 Cal. App. 2d 175, 183, 255 P.2d 59, 64 (1953), *appeal dismissed*, 346 U.S. 803 (1953) (disapproved and overruled, but on other grounds, *People v. Elliot*, 54 Cal. 2d 498, 503-04, 354 P.2d 225, 228-29, 6 Cal. Rptr. 753, 756-57 (1960)).

samples from two different speakers any workable minimum number of points of identity could be found,²⁷ hence they would be identified as being from the same speaker.

Two safeguards against this danger which the courts should consider are: (1) establishing a maximum limit to the number of identical sounds the expert can use to find the stipulated minimum number of points of identity, or (2) establishing a sliding scale whereby the minimum number of points required for positive identification increases with the number of identical sounds appearing in the two voice samples submitted to the expert. Until enough experience is gained with the voiceprint technique to establish what is a reasonable maximum for the number of identical sounds submitted to the expert, the first safeguard is undesirable because it may allow him to select the sounds which yield the greatest number of points. Therefore, the second safeguard is preferable. Moreover, until the voiceprint technique becomes accepted such that a specific minimum number of points can be legislated, every jury should be instructed that the weight accorded the testimony of the expert should be directly proportional to the relationship between the number of points found and the number of identical sounds in the two samples submitted to him.

A desirable variation of the safeguards, in those jurisdictions which allow the use of the technique would be to select a few phrases from the unknown voice sample and have an exemplar of these exact phrases made by the known voice. This has three advantages. First, the problem of counting identical sounds in each sample would be alleviated since all sounds would be identical. Secondly, since the influence of one sound on another would be the same in each sample, not only would excerpts be unnecessary, but the expert also would have more material to work with since comparisons of the transitional spectrogram segments can be made in the same manner as comparisons of the steady-state portions. Third, the danger of selection would be eliminated since the expert can reasonably be compelled to testify about the entire spectrograms, not merely those parts from which excerpts were made.

27. Even Kersta concedes this; however, he explains that in the cases in which he has participated there has been a sparsity, rather than an abundance, of identical sounds in the voice samples given him. Kersta interview, *supra* note 17.

II. THE SOUND SPECTROGRAPH²⁸

Another problem unique to the voiceprint technique among pattern-matching systems of identification relates to the peculiarities of the respective spectrograph; even though the voice sample may be identical the patterns may vary because of the construction and functioning of the different spectrographs utilized.

Basically, sound is the back and forth vibration of air molecules. When set in motion, one molecule collides with those surrounding it, setting them in motion, while the force of the collision propels the initial molecule back, beyond its original position. The distance traveled by the molecule is the amplitude (loudness) of the sound; the number of movements back and forth, or cycles, in a measured period of time is the frequency (pitch).

The most common device for measuring sound is the oscillograph. It indicates the frequency and amplitude of sounds by a line, or wave-form, which undulates regularly for a single-frequency pure tone, such as that produced by a tuning fork or a human whistle, but which takes on a more irregular shape for multi-frequency complex sounds. For certain studies of complex sounds, it is necessary to break them down into their component frequencies. Traditional mathematical evaluation of the oscillogram is useful for this purpose, but is inadequate for studies of very complex sounds like speech. To meet this deficiency, Bell Telephone Laboratories developed the sound spectrograph during World War II.²⁹

The heart of the spectrograph is the analyzing filter which is best explained as follows. The successive collisions of air molecules described above produce what are commonly called sound waves. When the waves strike a substance which is

28. Except where footnoted specifically herein, this section is a simplification of material in E. PULGRAM, *INTRODUCTION TO THE SPECTROGRAPHY OF SPEECH* (1959); P. LADEFOGED, *ELEMENTS OF ACOUSTIC PHONETICS* (1962), and G. Fant, *ACOUSTIC THEORY OF SPEECH PRODUCTION* 1-26, 58-62, 229-80 (1960). The first two are primers on sound, speech and the spectrograph and are excellent for a layman; the last one gives much more detailed and technical information. *See also* H. KAPLAN, *ANATOMY AND PHYSIOLOGY OF SPEECH* (1960).

29. R. POTTER, G. KOPP AND H. GREEN, *VISIBLE SPEECH* 4-5 (1947) [hereinafter cited as Potter].

unyielding they are reflected back, producing an echo. However, if the substance will vibrate at the frequency of the sound waves and if these sound waves strike it with enough force to set it in motion, the substance will resonate, producing its own sound waves at that frequency. This can be demonstrated with a number of tuning forks, two of which produce the same frequency sound. If one of the two is struck, the second will begin to vibrate on its own. The other forks will not respond similarly, because they do not vibrate at the same frequency as the sound waves striking them.

No sounds for which the spectrograph is employed have the single frequency tone of a tuning fork. Rather, they consist of complex, multi-frequency sounds which, if all the frequencies in the sound are multiples of a single frequency, are called tones; but if the frequencies are unrelated to each other, they are labeled noise. Speech involves both types, but the voiceprint technique relies primarily on spectrograms of the tones. If one of these complex sounds were produced near a bank of tuning forks, each of the forks which vibrates at the same frequency as one of the component frequencies of that sound would begin to vibrate or resonate. If a stylus were attached to all the forks to record on a moving paper (so that the time lapse would be represented) the possible response of each, the result would be a spectrogram. The bank of forks would function essentially as the analyzing filter of the spectrograph.

The earlier models of the spectrograph consisted of a magnetic tape on which the sound to be analyzed was recorded, a variable analyzing filter and a stylus which recorded the signals coming from the filter onto electrically sensitive paper. The more recent model used for the voiceprint technique has a number of refinements which, *inter alia*, produce a clearer, more precise spectrogram that is easier to read for identification purposes.³⁰

When a complex sound is analyzed by the spectrograph filter, it sends through the stylus an electric current which varies in power as does the amplitude of each particular component

30. A. Presti, *High-Speed Sound Spectrograph*, 40 J. ACOUST. SOC. AMER. 628 (1966) [hereinafter cited as Presti]. This is the machine currently being produced by the Voiceprint Laboratories, Somerville, N.J. Spectrographs produced by others may vary in detail and produce slightly different spectrograms, e.g., the Sonograph of Kay Electric Co., Pine Brook, N.J.

frequency of that complex sound. For each component frequency this current makes a dot on the facsimile paper which varies in darkness according to the power of the current, thereby indicating the relative amplitude of the particular component frequency. The typical paper has a 2:1 ratio, *i.e.*, if a certain amplitude causes the faintest possible dot, an amplitude twice that power makes the darkest possible dot.³¹ It is important to note that adjustments can be made on the filter which not only will affect the darkness of a particular dot, but may also preclude the frequency it represents from registering at all.

It also should be noted that the dot on the spectrogram represents more than a single frequency. For speech research and the voiceprint technique, the spectrograph need analyze only the component frequencies of a sound which are below 8000 cycles. In fact, for most work, those under 4000 cycles are adequate,³² since above this range appear only noise sounds of certain consonants such as the hisses made by the "s" and "f". Furthermore, producing a spectrograph filter with a single-cycle displacement, *i.e.*, the equivalent of a bank of 8000 tuning forks to indicate every frequency up to 8000 cycles, would be unwieldy, expensive and superfluous for speech research. Consequently, a scientific compromise was struck resulting in a spectrograph filter that accepts a number of frequencies at once.

To produce a complete spectrogram the sound being analyzed is played repeatedly with the filter moving to the next displacement level for each replay.³³ The result is a series of scan lines which produce a pattern similar to the scan lines on a television screen.³⁴ The frequency scale is on the vertical axis, the time scale is on the horizontal axis, and the amplitude of various component frequencies is represented by the darkness of the pattern.

31. Kersta, *NATURE*, *supra* note 7, at 1254.

32. J. Carbonell, M. Grignetti, K. Stevens, C. Williams and B. Woods, Final Report: Speaker Authentication Techniques, Project Report No. 1E6-79191-D-491-05, U.S. Army Electronics Laboratories, Ft. Monmouth, N.J. 7-8 (May 1965). (This report was prepared by Bolt, Beranek and Newman, Inc., Cambridge, Mass., and is listed in its library as BBN Report No. 1296) [hereinafter cited as Bolt, Beranek Report].

33. The spectrograph of Voiceprint Laboratories, *supra* note 13, has a filter with a 17 cycle displacement. It produces a spectrogram four inches high made up of 400 scan lines covering frequencies up to 7500 cycles. The magnetic tape accepts 2.4 seconds of sound from which a spectrogram about ten inches long is produced.

34. This refers only to the bar spectrogram. Other types include a contour

As mentioned above, the selection of filter displacement is a scientific compromise between a one cycle displacement and the oscillograph which accepts all frequencies at once. Consequently, spectrographs with different filter displacements will produce different spectrograms. Since the voiceprint technique involves a visual pattern-matching process which measures one spectrogram against another, it is evident that both samples must be produced by machines with precisely the same filter displacement.

Displacement, however, does not pose the sole problem. The cavities in the head resonate not only their natural frequency, but also resonate a cluster of subsidiary frequencies which decline in power the farther they are from the natural one. Since the usual facsimile paper has only a 2:1 ratio it is necessary to modulate the filter allowing only frequencies of a certain minimum power to register. Otherwise, the resulting spectrogram would be completely black, except for a few gray areas representing the weakest frequencies. Furthermore, the degree of the decline in power of the subsidiary frequencies varies, even for two cavities with the same natural frequency. Moreover, modulating the filter affects the relative darkness of the resonance bars on the spectrogram, and also determines their width. As a result, the two spectrograms compared must also be made on machines with filters set at the same level. It should be noted that most spectrographs are equipped to regulate the amplitude of the voice sample tape so it will not entirely exceed the capacity of the facsimile paper.³⁵

The representation of frequency along the vertical axis of the spectrogram could pose another problem. Some spectrographs are equipped to adjust the vertical scale to enlarge some frequency ranges and diminish others.³⁶ The scale used for most

spectrogram, resembling a contour map with ridges representing the increasing or decreasing amplitude. The voiceprint technique uses bar spectrograms almost exclusively. A speaker classification system, comparable to that used for fingerprints, has been suggested, using contour spectrograms. See Kersta, *Voiceprint Classification* (unpublished paper presented to the Acoustical Society of America meeting June 5, 1965) and Kersta, *Voiceprint Classification for an Extended Population* (unpublished paper presented to the Acoustical Society of America meeting June 2, 1966).

35. Presti, *supra* note 13, at 629.

36. These spectrograms were made on a Voiceprint Laboratories spectrograph. *Id.* The logarithmic adjustment has particular value for the voiceprint technique because the most significant information is contained in the lower frequency ranges. See note 15 *supra*.

voiceprint technique identifications is adjusted logarithmically to expand the lower frequency range which is more important for the voiceprint technique. The difference between this scale and an unadjusted scale is obvious to the naked eye; however, the variety of adjustments is great,³⁷ and should be checked before two spectrograms are compared.

Another concern is the range of the frequency scale. Although the human voice produces meaningful sounds which vary from about 80 cycles to about 8000, most of the equipment used to obtain voice samples accepts considerably less than that full range,³⁸ e.g., the telephone only accepts frequencies up to about 3500 cycles. This is analogous to finding only part of a fingerprint. It presents another variable for the courts to consider when determining the number of points of identity to be required and the weight to be accorded the testimony of the voiceprint technique expert.

As to malfunctioning of the spectrograph, the more advanced models have built-in checks and calibrations which will immediately indicate any malfunction to a trained operator.³⁹

III. THE ACOUSTICS OF SPEECH PRODUCTION⁴⁰

Similar to a tuning fork, a volume of air in a container may also resonate when struck by sound waves of its natural frequency. There are distinct differences from the tuning fork, however. First, the air may not merely resonate the original sound but also reinforce it by increasing its amplitude. This is demonstrated by the increase in loudness produced when the stem of a struck tuning fork is placed on a hollow box. Secondly, the air chamber will resonate not only the single tone which activates its natural frequency, but due to its imperfections as a resonator, it will also produce subsidiary sound waves clustered around that natural frequency. These subsidiary frequencies diminish in power the farther they are from the natural frequency, i.e., a subsidiary sound one cycle greater than the natural frequency is more powerful than the one two cycles greater. This effect is called damping. It is highly relevant to the

37. Presti, *supra* note 13, at 629.

38. Kersta interview, *supra* note 17.

39. *Id.*; see also Presti, *supra* note 13.

40. See note 11 *supra*.

voiceprint technique, because the pattern made by damping is governed largely by the shape of the particular resonator. For example, even though two chambers vibrate at exactly the same natural frequency, the degree of the decline in power of the subsidiary sound frequencies will differ if the shapes of the two chambers differ.

Excluding whispers and certain consonants (*e.g.*, *h*, *s*, *f*, *t* and *k*), speech originates with very complex sound produced by the vocal cords—the glottal tone. As this tone passes through the vocal tract, it is resonated, reinforced, and damped in the cavities of the nose, and throat, which are of fixed shape, and the mouth. Although the cavities formed in the mouth are predetermined to some extent by the anatomy of the individual, they are largely created by the movement of the tongue, the jaw, and the lips—the articulators. These variations account for most of the differences among speaking sounds. Since minor variations can produce widely divergent sounds,⁴¹ an individual must be careful to place his articulators in the proper position to produce the sound attributed to the meaning he wishes to convey. This is the process of learning to talk.

The size and shape of the various cavities determine the glottal tone frequencies which are resonated, reinforced and damped and the degree of each process. The resulting clusters of relatively high amplitude frequencies which are produced by each cavity are called formants and are reflected on the spectrogram by dark bands called resonance bars. They are very important both to the study of speech and to the voiceprint technique. The blackened bar along the bottom of the spectrogram represents the fundamental frequency of the glottal tone, those above represent the formants selected and produced by the various cavities in the head. It is important to note two things: (1) since the sounds from one cavity affect those from the others, the resonance bars can not be attributed solely to individual cavities; (2) since the cavities will only resonate their natural frequency and its subsidiaries, changes in the pitch of the voice will not affect the position of the resonance bars in the spectrograms, except in the unlikely case that such a change would eliminate

41. Compare the placement of the articulators when pronouncing "beat," "bit," and "bate."

completely the particular natural frequency of one of the cavities from the glottal tone.⁴²

The function of the spectrograph is comparable to that performed by the middle and inner ear which analyze the components of a complex sound wave. The brain identifies the components, thereby identifying the speech sound. Just as speech sounds of different individuals have to be very similar to enable the brain to identify them, so the spectrograms of the same speech sound by different individuals are very similar. So similar in fact, that one of the earliest uses of the spectrograph was to provide deaf people with "visual" hearing.⁴³

IV. THE THEORY OF INVARIANT SPEECH⁴⁴

The theory of invariant speech is the crux of the current controversy over the validity of voice identifications through spectrogram comparisons. Essentially, two interrelated propositions are asserted: (1) every individual speaks so uniquely that the differences between the same utterance by two people are greater than the differences between the same utterance by one person in varied contexts; (2) the spectrogram adequately portrays that uniqueness.

Adequacy of the Spectrogram

The second proposition, though not verified, is subject to less dispute. It is established that the spectrograph performs a function comparable to that of the human ear: analyzing complex speech sounds and portraying the most significant components, just as the ear analyzes the same sounds for evaluation by the brain. Therefore, it is probable that the spectrograph visually portrays many of those features which allow a listener to aurally distinguish one voice from another.⁴⁵ Moreover, the ear distorts certain sounds;⁴⁶ the spectrograph does not. Therefore, visual spectrogram comparisons should be more accurate than the ear for identification.⁴⁷

42. G. FANT, ACOUSTIC THEORY OF SPEECH PRODUCTION 20 (1960) [hereinafter cited as Fant].

43. Potter, *supra* note 12 at ix.

44. E. PULGRAM, INTRODUCTION TO THE SPECTROGRAPHY OF SPEECH (1959).

45. The spectrogram fails to distinguish as well as the ear among certain consonants, e.g., *p*, *b* and *t*. Fant, *supra* note 20, at 22-23.

46. Brief for State at 5, State v. Cary, 99 N.J. Super. 323, 239 A.2d 680 (1968) (Testimony of Dr. Tosi) (citing Record at 452-54, 520, 563).

47. 19 AM. JUR. PROOF OF FACTS, *supra* note 34, at 431.

Verification of this theory awaits the accumulation and examination of a sufficient number of different spectrograms before a finding can be made of the statistical probability of any two being identical.⁴⁸ Even the developer of the voiceprint technique concedes that this is necessary.⁴⁹ However, he contends that extrapolations from his experiments tentatively should be used to validate not only the adequacy of the spectrograph, but also the uniqueness of every individual's speech.⁵⁰ Many in the scientific community do not accept such extrapolations, though they would if they were convinced of the validity of the experiments.

The Theory

Before exploring the controversy over the experiments, the contention that the speech of every individual is unique should be examined. It will be recalled that the width of the resonance bars is largely determined by differences in the decline in power of the subsidiary frequencies produced by the various cavities, and that the location of the bars is due to the natural frequency resonated by the cavities. Prior to the development of the voiceprint technique, it was accepted that differences in the resonance bars were due to idiosyncracies in every individual's method of speaking.⁵¹ The logical extension of the proposition is that idiosyncracies in cavity formation cause the differences in the resonance bars. The shape of the cavities is determined by the anatomical characteristics of the vocal tract and by the use of the articulators. Since it is widely accepted that the anatomical characteristics are unique in each individual,⁵² this factor is the constant in the equation.

The controversy exists with the other element of the equation, the role of the articulators. The theory of invariant speech is based on the contention that positioning of the articulators

48. Note, for example, fingerprints. The FBI now has almost 100 million different sets with no repeats, therefore the statistical probability of two identical sets appearing is minute.

49. L. Kersta, *Speaker Recognition and Identification by Voiceprints*, 40 CONN. BAR J. 586, 592 (1966) [hereinafter cited as Kersta, CONN. B.J.].

50. Kersta, *Infallibility*, *supra* note 26.

51. P. Garvin and P. Ladefoged, *Speaker Identification and Message Identification in Speech Recognition*, 9 PHONETICA 193, 197 (1963) [hereinafter cited as Garvin]; see also Fant, *supra* note 20, at 21.

52. Brief for State at 2, *State v. Cary*, 99 N.J. Super. 323, 239 A.2d 680 (1968) (Testimony of Dr. Gens) (citing Record at 757).

becomes so habitual that the speaker cannot materially alter it. Hence, the articulators in tandem with the anatomical characteristics produce relatively invariant and unique speech in every person.⁵³ There is one reservation. Unless the individual is physically and linguistically mature, change is inevitable over any extended period of time; therefore, voice samples must be taken soon after one another for the theory to be applicable to children.⁵⁴

Speech is a learned process in which the infant experiments with positions for the articulators until he is satisfied that he has reproduced the intended sound; thereafter the positioning does become habitual.⁵⁵ However, these habits can be changed and often are to correct speech defects.⁵⁶ Whether an individual can deliberately alter one of these habits, either permanently or temporarily, to effect a disguise, is not certain.⁵⁷ Whether the habits are so strong that they will not involuntarily yield in some circumstances, *e.g.*, when the individual is undergoing stress,⁵⁸ or is assimilating a new regional accent, is not certain.⁵⁹ Whether the habits of one individual might compensate for the anatomical differences between himself and another as they both strive to produce the same speech sound is uncertain;⁶⁰ however, given the number of variables involved,⁶¹ it is unlikely. To what extent factors such as laryngitis, colds, and dentures alter the spectrogram is not certain.⁶² Resolving these uncertainties should

53. Kersta, *NATURE*, *supra* note 7, at 1254-55.

54. Kersta interview, *supra* note 27.

55. Brief for State, *supra* note 52 (citing Record at 756).

56. See Garvin, *supra* note 51 at 194.

57. Some Kersta experiments indicate this is not possible. Kersta, *CONN. B.J. supra* note 49, at 589; See text accompanying notes 63-91 *infra*.

58. Dr. Kenneth Stevens indicates that experiments on stress are currently being conducted at Bolt, Beranek and Newman, Inc., Cambridge, Mass., and the results so far indicate greater differences in spectrograms for one individual in stress situations than among different people in non-stress situations. Interview with Dr. Kenneth Stevens, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass., Mar. 12, 1968. [hereinafter cited as Stevens interview].

59. P. Ladefoged and R. Vanderslice, The "Voiceprint" Mystique (unpublished paper presented to the Acoustical Society of America meeting, Nov. 14, 1967) [hereinafter cited as Ladefoged, Mystique].

60. *Id.*; see also the testimony of Dr. Joos. Brief for State at 29, *People v. King*, 266 Adv. Cal. App. 466, 72 Cal. Rptr. 478 (1968) (citing Record at 944-45); testimony of Dr. Ladefoged, *Id.* at 34 (citing Record at 2290-91); testimony of Dr. Fromkin, *Id.* at 40-41 (citing Record at 2543-47); testimony of R. Vanderslice, *Id.* at 42 (citing Record at 2569-70).

61. Brief for State, *supra* note 52 (citing Record at 758).

62. J. Kress, *Voiceprints and the Law* n.15, p. 33 (unpublished paper written for the

be accomplished through further study and a larger accumulation of spectrograms.

V. THE IDENTIFICATION EXPERIMENTS AND SCIENTIFIC OPINION

The yardstick against which the experiments must be measured is a variety of aural voice identification experiments in which listeners have achieved accuracy rates of 81 percent,⁶³ 84 percent,⁶⁴ and from 88 to 94 percent.⁶⁵

Experiments with Spectrographic Voice Identifications

The first investigation of the possibility of spectrographic identification of voices was conducted during World War II by Charles Grey at Bell Telephone Laboratories. He sought to identify enemy radio operators so the units to which they were attached could be located.⁶⁶ His findings, while promising but inconclusive, were turned over to Lawrence G. Kersta, who during the early 1960's developed the voiceprint technique in response to the need of law enforcement agencies to identify telephone bomb-threat callers.⁶⁷

Kersta conducted two types of experiments with high school girls who had received a week's training in the features of the spectrogram which he thought were relevant for speaker identification.⁶⁸ In the first experiment the subjects, working in pairs, were given spectrograms of 20, 36 and 48 isolated utterances of five, nine and twelve different speakers, respectively. Placing the utterances in piles of four each per speaker, the girls achieved average accuracy rates of 99.2 percent. Repeating the same tests with excerpts of sounds out of contextual passages, the average accuracy rate was 99 percent. The second experiment required matching one spectrogram with another spectrogram by the same speaker from a series of nine to 15

seminar Science and Law, Columbia University School of Law, September 1967) (citing interview of Kersta by Joe Pyne on WNEW-TV, N.Y., Mar. 12, 1967).

63. P. Bricker and S. Pruzansky, *Effects of Stimulus Context and Duration on Talker Identification*, 40 J. ACOUST. SOC. AMER. 1441 (1966).

64. I. Pollack, J. Pickett and W. Sumbly, *On the Identification of Speakers by Voice*, 26 J. ACOUST. SOC. AMER. 403 (1954).

65. Bolt, Beranek Report, *supra* note 15.

66. Brief for State at 6, *supra* note 46 (testimony C. Greg) (citing Record at 16-27).

67. Kersta interview, *supra* note 17.

68. Kersta, NATURE, *supra* note 7.

69. Tosi, Evaluation, *supra* note 20.

samples. On this test the subjects achieved an average accuracy rate of 99 percent. In both experiments it was found that some pairs of girls did better than others and that for all the subjects some sounds were more difficult than others. These experiments utilized a library of spectrograms of twelve different voices. For subsequent experiments this library was expanded to 123 voices. All samples were made over telephone lines;⁶⁹ hence only frequencies below 3400 cycles appeared.

In subsequent experiments, the same high school girls were given exemplars of one voice containing five different excerpts and were asked to match each exemplar with another five-excerpt sample of the same voice out of 50 samples.⁷⁰ Again, the accuracy rate averaged 99 percent.

Further investigation by Kersta included comparing samples of the same voice when whispered or muffled and when the speaker's nose was held, and comparing the normal voice and two of the dummy-voices of a professional ventriloquist. In each instance he concluded that the use of certain selected clues would provide identification despite these attempts to disguise the voice.⁷¹ Unfortunately, exactly what clues are used by the voiceprint technique have never been revealed publicly by Kersta.⁷²

A course which he offers in the technique to "qualified" students⁷³ utilizes the original library of 123 voices. After a week of training in one such course,⁷⁴ the students, five fingerprint experts, were tested individually and in two-man teams following the same procedure used in the 50-voice experiment with the high school girls. The results were an average accuracy rate of 93.46 percent.

70. Kersta, *Infallibility*, *supra* note 16.

71. Kersta, *CONN. B.J.*, *supra* note 49.

72. *But see* text at note 20, *supra*, where Dr. Tosi, who has done some work with Kersta, reveals these clues.

73. The course is a two week, in residence training program in the Voiceprint Laboratories, Somerville, N.J. Subsequently, a correspondence course and mandatory field work by the student are required. After two years of this continuing education and when a level of proficiency is reached which Kersta considers adequate, the student is deemed an "expert" in the technique. Eight students have completed the course. Three are with a federal agency, two with the Michigan State Police, and one each with the police departments of Philadelphia, Pa., Nassau County, N.Y. and Somerset County, N.J.

74. Tosi, *Evaluation*, *supra* note 20.

In a recent hearing involving the reliability of the voiceprint technique,⁷⁵ an in-court experiment was conducted in which one of the original high school girls correctly matched two spectrograms of short, identical telephone messages from each of five different speakers.

A number of identifications for law enforcement authorities have been made by Kersta personally.⁷⁶ One of his early identifications established the innocence of an accused telephone-threat caller.⁷⁷ Kersta has also testified to positive identification in three trials,⁷⁸ but in one of them the suspect had incriminated himself in the exemplar tapes from which Kersta made his identification.⁷⁹

Creighton University also tested Kersta's skill in a pilot study on the feasibility of voiceprint technique identifications for infants.⁸⁰ Evidently, fingerprints and footprints are not completely reliable to prevent confusion between parent and baby. The cries at birth of eight infants, including a pair of identical twins, were recorded. Four days later, in a different order, cries of seven of the eight, again including the twins, were recorded. Both recordings were sent to Kersta who correctly matched all seven of the fourth-day cries with the corresponding seven of the original eight. Interestingly, the spectrograms of the twins were so different that Kersta was not aware of the existence of twins until he was informed, after he had made the comparisons. A complete experiment with 150 infants is planned to verify the results of this pilot study.

Contemporaneously with Kersta's early experiments, Bolt, Beranek and Newman, Inc., Cambridge, Massachusetts, was investigating the potential of spectrogram voice identifications as compared to aural procedures.⁸¹ Two experiments were

75. 99 N.J. Super. 323, 239 A.2d 680 (1968).

76. Kersta testified to over 100 such identifications. Brief for State at 7, 99 N.J. Super. 323, 239 A.2d 680.

77. Kersta, CONN. B.J. *supra* note 49, at 593.

78. *United States v. Wright*, 17 U.S.C.M.A. 183, 87 C.M.R. 447 (1967); *People v. King* (Los Angeles County, Cal.) in *L.A. Times, West Magazine*, Mar. 26, 1967, at 12; *People v. Strachle* (Westchester County, N.Y.) in *N.Y. Times*, Apr. 12, 1966, at 1, col. 2.

79. Ladefoged, *Mystique*, *supra* note 59, indicates that in the *King* case one of the exemplar tapes sent to Kersta included a confession and the other had incriminating remarks on it.

80. Letter from B. Zumpano, Obstetrics and Gynecology Dept., Creighton University, Omaha, Nebr., Feb. 12, 1968.

81. Bolt, Beranek Report, *supra* note 32.

conducted. The first involved matching-to-sample tests comparable to those conducted by Kersta. The second was an authentication experiment in which the subjects were provided with eight "known voice" spectrograms and instructed to determine whether each in a series of 24 other spectrograms was one of the "known voices." The composition of the 24 spectrogram group was twelve "known" and twelve "unknown" voices in some tests, and 21 "known" and three "unknown" in other tests.

The best results for the matching-to-sample experiment was an average accuracy rate of 79 percent.⁸² The results for the authentication experiment varied from a low average accuracy rate of 53 percent for correctly identifying "unknown" voices to a high of 90 percent for identifying "known" voices.⁸³

The wide discrepancy in results between these matching-to-sample experiments and those of Kersta, as well as the suggested high inaccuracy rate for positive authentication of spectrogram samples, is perplexing. It cannot be due to the small differences in the spectrograms. The Bolt, Beranek and Newman spectrograms pictured frequencies between 250 and 3250 cycles, whereas Kersta's included all those ranging to 3500 cycles. Also, the differences in the subjects, high school girls for Kersta and college students for Bolt, does not justify the discrepancy.

However, two factors could account for the variation in results: the difference in choice of sounds and the complete lack of any kind of training for the Bolt subjects. In some tests Bolt used entire words instead of excerpted sounds, while in others they utilized complete phrases. However, better identification results were reported for the spectrograms with phrases than for those with single words;⁸⁴ therefore, this would militate against the use of excerpts. Thus, the one week of training which Kersta gave his subjects must account for the difference between his results of 99 percent and 93.46 percent, and the 79 percent achieved by the Bolt subjects. Interestingly, the Bolt subjects indicated that they utilized many of the clues employed by Kersta in his voiceprint analysis for their identifications; however, they also used some of the rejected features.⁸⁵

82. *Id.* at 58.

83. *Id.* at 64.

84. *Id.* at 43.

85. *Id.* at 62.

Recently an attempt to replicate Kersta's experiments was made at Case-Western University. The subjects were given one-word exemplars for each of five voices.⁸⁶ They were asked to match these exemplars, one at a time, with 15 other spectrograms (three from each voice). This was done first with two isolated words, then with the same words taken from various contexts. The average accuracy rates reported were 78.4 percent for the isolated words but only 37.3 percent for the words from contextual passages.⁸⁷

The experimenters suggested that the most likely reason for the discrepancy between their results and Kersta's was the difference in matching procedures.⁸⁸ However, Kersta's second series of tests with the pairs of high school girls and the training tests with the fingerprint experts were essentially the same as this experiment. Therefore, the more likely explanation for the discrepancy in results was the difference in method of excerpting sounds from context and the training procedures. The latter explanation was expressly rejected by the Case-Western experimenters.⁸⁹

The sounds taken out of context for this Case-Western experiment were not merely the steady-state portions, but also included the transitional segments of the spectrogram reflecting the influence of preceding and succeeding sounds. Apparently, the subjects were given no specific instructions concerning the differences which appear in those areas of the spectrogram even though the particular sounds are the same.

In addition, since the clues used by the voiceprint technique were unknown to them, the experimenters had to select what they thought to be the appropriate clues. As a result, they excluded some clues which are used by the technique and included others which had been rejected.⁹⁰ Although all the subjects had had previous experience with the spectrograph, the two hours of specific instruction given them on identification clues was apparently insufficient.

86. M. Young and R. Campbell, *Effects of Context on Talker Identification*, 42 J. ACOUST. SOC. AMER. 1250 (1967) [hereinafter cited as Young, Effects].

87. *Id.* at 1252.

88. *Id.* at 1254.

89. *Id.*

90. *Id.* at 1251.

In experiments conducted at the University of California, Los Angeles, it was demonstrated that the dissimilarities among various spectrograms of the same speaker are generally as numerous as the similarities among different speakers.⁹¹ The experimenters contended "that it is not possible even to estimate the chances of two voices being indistinguishable since a given individual's voice is not invariant (sic)." For this reason, it was concluded that the voiceprint technique, which disregards dissimilarities among spectrograms for the purposes of identifying the voice with the exemplar is liable to result in incorrect identifications. However, the possibility that some method could be devised to prevent mistakes was not precluded by their findings, and, significantly, the clues used by the voiceprint technique to find points of identity were unknown to the experimenters.

Scientific Opinion About the Voiceprint Technique

Scientific opinion is currently divided between two groups, each claiming that the other is not qualified to judge the merits of the voiceprint technique.⁹² Favoring the technique are acoustics engineers and sound specialists; those opposing it are linguists and speech specialists. In fact, a knowledge of both acoustics and linguistics, as well as elementary physiology, anatomy and statistics, probably is requisite for making an appropriate evaluation.

The critical question concerning the technique is whether the shapes of the cavities, due to articulation and the anatomical differences among people, vary more between two individuals than between two pronunciations by the same individual. The acoustics of speech production are portrayed by the spectrograph, and therefore an analysis of a sufficient number of spectrograms would tend to statistically answer this question. However, whether two individuals are able to duplicate the acoustical phenomena of each other's speech can only be

91. Ladefoged, *Mystique*, *supra* note 59.

92. Compare the testimony of Kersta in Brief for State at 19, *People v. King*, 266 Adv. Cal. App. 466, 72 Cal. Rptr. 478 (1968) (citing Record at 1563-64), *M. Rettinger*, *Id.* at 23 (citing Record at 1593) and *Fred East*, *Id.* at 24-25 (citing Record 1262-63, 1293-94) with the testimony of Dr. Gerstman, *Id.* at 39 (citing Record at 2524-25), Dr. Clarke, *Id.* at 27 (citing Record at 2177-80) and Dr. Ladefoged, *Id.* at 33-34 (citing Record at 2276-77).

determined by studying anatomy and the physiology of speech production. Furthermore, at least a minimal knowledge of phonetics is necessary to select proper sound excerpts from the spectrogram.

Clearly, the requisite modicum of acoustic theory must have been mastered by anyone experimenting with the spectrograph. Since an acoustician need know nothing about linguistics or the physiology of speech, the opinions of the speech specialists probably should be given more credence.

Both proponent and opponent alike agree that before the voiceprint technique can be deemed conclusively accurate more voice samples must be made and compared. It is also agreed that variations in spectrograms due to machines of different manufacturers, though sometimes significant, do not have any relevance to determining the validity of the technique *per se*. Nevertheless, opponents of the technique have mounted a three-pronged offensive. First, the developer's experiments are attacked not only as irrelevant to the positive identification of voices, but also for his failure to publish adequate reports which would allow others to replicate them. Secondly, it is contended that even if the theory of invariant speech were proven, the spectrograph is not sophisticated enough to provide reliable information for identifications. Finally, the developer's ability to make positive identifications utilizing the technique is disputed.

As mentioned above, no one has been able to replicate the experiments upon which the validity of the technique is based. This fact alone is sufficient to generate doubt in the minds of most of the opponents.⁹³ This problem is aggravated not only by the developer's failure to publish the specific clues of the technique, but also his tendency to oversimplify the process when testifying in trials. As a result, other scientists reading the trial records and following the procedures outlined in the record conclude that it is impossible to correctly identify voices using the methods of the technique.⁹⁴ The only scientists who have complete descriptions of the technique are those conducting experiments at Michigan State University. However, until a

93. Letter from Dr. Frank Clarke, Senior Research Psychologist, Sensory Sciences Research Center, Stanford Research Institute, Menlo Park, Cal., Feb. 12, 1968.

94. Ladefoged, *Mystique*, *supra* note 59.

sufficient number of experiments are completed, they refuse to testify to the accuracy of the technique.⁹⁵ With respect to the operation of the spectrograph, the leading Bell Telephone experts, who were at least in close proximity to the early experiments, have refused to testify concerning the technique.⁹⁶

Aside from the lack of replication, the experiments on which the claims to reliability of the technique are based are doubtful authority for the proposition that absolute, positive identifications can be made.⁹⁷ Matching-to-sample experiments, like those involving the high school girls and fingerprint experts, afford many extraneous clues to the correct answers.⁹⁸ The "unknown" to be ascertained in these experiments is merely to which "category" the sample belongs;⁹⁹ the judgment is based on the relative likeness of the sample to the exemplar, not its absolute sameness.¹⁰⁰ What is needed are authentication experiments comparable to those of Bolt, Beranek and Newman. The Creighton University infant birth cries experiment was also of that type. However, this experiment is only a harbinger of the possibilities of positive identification with the voiceprint technique, particularly since it did not probe the issue of whether articulation might compensate for, rather than accentuate, differences in anatomy.

Kersta is acknowledged as having more experience with the spectrograph than any other person working in the speech field.¹⁰¹ However, it is widely believed that he has not demonstrated his capacity to make positive voice identifications. Undoubtedly, this is partly due to his failure to publish a comprehensive paper on the voiceprint technique, and partly on the lack of information within the scientific community concerning his work for investigative and law enforcement agencies. But Kersta's failure to adequately explain the technique when testifying in court is primarily responsible for the lack of confidence in this method. The most searing denunciation of the

95. Brief for State, *supra* note 46 (testimony of Dr. Tosi).

96. Kress, Voiceprints, *supra* note 62, at 23.

97. Stevens interview, *supra* note 58.

98. Young, *Effects*, *supra* note 86, at 1254.

99. Brief for State at 36-37, *supra* note 92 (testimony of Dr. Gerstman) (citing Record at 2403-05).

100. Ladefoged, *Mystique*, *supra* note 59.

101. Stevens interview, *supra* note 58.

technique¹⁰² is based on an attempt to replicate an identification in a particular case¹⁰³ in the manner described by Kersta. The authors demonstrated that Kersta had apparently misidentified some sounds, that he had merely pointed out the position of the resonance bars which would be practically the same for everyone pronouncing those sounds, that there were wide differences in some of the spectrograms he had used when testifying, and that he had never explained the clues used to find points of identity. This should be considered an indictment of Kersta's secretiveness and credibility, but not necessarily of the voiceprint technique.

The opponents further stress that even if the theory of invariant speech is valid, which is sharply disputed,¹⁰⁴ the spectrograph has not been demonstrated to be sophisticated enough to show the invariant features.¹⁰⁵ Relying on their own experience with the spectrograph some opponents believe that it is possible to produce voice imitations which would confound the technique,¹⁰⁶ while others think that more sophisticated equipment is required to produce spectrograms which could be used for identifications.¹⁰⁷ One has suggested that only by looking at spectrograms under a microscope could reliable identification features be found.¹⁰⁸ It has further been suggested that even for someone with Kersta's experience the spectrograph is technically too unrefined for positive identifications,¹⁰⁹ though it is probably adequate for classification of speaker types.¹¹⁰ However, again it must be noted that none of these critics had access to detailed information about the voiceprint technique.

102. Ladefoged, *Mystique*, *supra* note 59.

103. *People v. King* (Los Angeles County, Cal.) in *L.A. Times, West Magazine*, Mar. 26, 1967, at 12).

104. Dr. Stevens noted that not enough research in the physiology of speech has been done to show what cavities produce exactly which marks on the spectrogram and which of those, if any, are invariant enough to provide data for positive speaker identification. Stevens interview, *supra* note 58; *accord* Ladefoged, *Mystique*, *supra* note 59; but *contra* Testimony of Dr. Gens, *supra* note 52.

105. Brief for State at 26, *supra* note 92 (testimony of Dr. Clarke) (citing Record at 2155).

106. *Id.* at 32 (testimony of Dr. Ladefoged) (citing Record at 2258-59).

107. Stevens interview, *supra* note 58.

108. Brief for State at 10, *supra* note 46 (testimony of Dr. Gerstman) (citing Record at 920-24).

109. Brief for State at 28, *supra* note 60 (testimony of Dr. Joos) (citing Record at 920-24).

110. Brief for State at 39, *supra* note 92 (testimony of Dr. Gerstman) (citing Record at 2521).

The opposition is epitomized by the resolution of the committee of the speech communication section of the Acoustical Society of America, which passed unanimously in March, 1966:

The Technical Committee on Speech Communication is concerned that "voiceprints" have been admitted as legal evidence on the basis of claims which have not yet been evaluated scientifically. The Committee invites the Executive Council to consider the matter and take appropriate action.¹¹¹

Even the most virulent opponent of the technique has admitted that within a selected population voice identifications by spectrogram are possible.¹¹² Others suggest that voiceprint technique identifications are more likely valid than not, are entitled to some measure of reliability,¹¹³ and are certainly far better than chance.¹¹⁴ Another opponent predicted that there would soon be general scientific acceptance for use of spectrograms negatively, *i.e.*, to prove the unknown voice and exemplar are not the same.¹¹⁵ However, most refuse to express an opinion, only urging that more study must be made to verify the technique before it can be accepted.¹¹⁶

VI. EVIDENTIARY PROBLEMS WITH THE VOICEPRINT TECHNIQUE

Whenever the voiceprint technique is employed in a criminal case a number of legal problems may arise. A short discussion with respect to self-incrimination, the right to privacy, due process and admissibility will suffice to illustrate the potential problems.

Self-Incrimination

*United States v. Wade*¹¹⁷ appears to have settled the controversy dealing with voice identifications. The Supreme

111. A sub-committee is presently investigating the claims made for the voiceprint technique to see if they have been substantiated. Stevens interview, *supra* note 58.

112. Brief for State at 11, *State v. Cary*, 99 N.J. Super. 323, 239 A.2d 680 (testimony of Dr. Ladefoged) (citing Record at 1216-22).

113. Brief for State at 29-30, *supra* note 60 (testimony of Dr. Joos) (citing Record at 953 and 978).

114. Testimony of Dr. Clark, *United States v. Wright*, 17 U.S.C.M.A. 183, 189, 37 C.M.R. 447 (1967).

115. Stevens interview, *supra* note 58.

116. Interview with Dr. Carl E. Williams, Research Scientist, Bolt, Beranek and Newman, Inc., Cambridge, Mass., Mar. 11, 1968.

117. *United States v. Wade*, 388 U.S. 218 (1967).

Court, Mr. Justice Brennan writing, held that compelling a defendant in a line-up to repeat the exact words used at the scene of the crime does not violate the fifth amendment. The voice itself was found to be a physical characteristic and not incriminating unless the defendant was compelled to disclose knowledge about the crime.¹¹⁸ Repeating the exact words used by the perpetrator at the scene of the crime was not considered to be such a disclosure. Four Justices dissented from this part of the opinion. Since making a voice exemplar consists merely of recording a voice demonstration, *Wade* would preclude any problem with federal self-incrimination protection.

Although the federal self-incrimination standard was applied to the states through the Fourteenth Amendment in *Malloy v. Hogan*,¹¹⁹ it established only the minimum requirement.¹²⁰ The states have been left free to grant wider protection and many have.¹²¹

Right of Privacy

In *State v. Cary*,¹²² the Supreme Court of New Jersey suggested a novel fourth amendment penumbra. As a prerequisite to issuance of an order for a voice exemplar, the court ordered a pre-trial hearing on the reliability of the voiceprint technique. The rationale was that, unless the voiceprint technique is acceptable as evidence, ordering the defendant to make an exemplar of his voice would be an "unreasonable" search, and in violation of the fourth amendment. If the technique proved unacceptable as evidence, then the state would have no interest to justify forcing a defendant to undergo even this slight indignity.

The New Jersey court relied on the holding in *Warden*,

118. *Id.* at 221-23.

119. *Malloy v. Hogan*, 378 U.S. 1 (1964).

120. *Griffin v. California*, 380 U.S. 609 (1965).

121. Most states do not depart from the *Wade* standard, e.g., *People v. Ellis*, 65 Cal. 2d 529, 421 P.2d 393, 55 Cal. Rptr. 385 (1966); *Lanford v. People*, 159 Colo. 36, 409 P.2d 829 (1966) (dictum); *Lenoir v. State*, 197 Md. 495, 80 A.2d 3 (1951); *Johnson v. Commonwealth*, 115 Pa. 369, 9 A. 78 (1887). *Contra e.g.*; *Beachem v. State*, 144 Tex. Crim. R. 272, 162 S.W.2d 706 (1942) (subsequent cases indicate waiver is not difficult and that compulsion may be requisite to violate the privilege); *State v. Taylor*, 213 S.C. 330, 49 S.E.2d 289 (1948); *State v. Freeman*, 195 Kan. 561, 408 P.2d 612 (1965). Other states have not yet had to face the problem directly, but they probably will follow the example of *Wade* unless the wording of the state constitution compels otherwise.

122. *State v. Cary*, 49 N.J. 343, 230 A.2d 384 (1967).

Maryland Penitentiary v. Hayden,¹²³ where Justice Brennan argued that there must be some nexus between that which is seized and the criminal activity. There is additional support for this position in *Schmerber v. California*¹²⁴ and *Terry v. Ohio*.¹²⁵ In order to determine whether a search was "reasonable" under the fourth amendment, the United States Supreme Court, in both cases, balanced the state's interest in invading the defendant's privacy against the severity of the invasion. Though both cases involved warrantless searches, the same balancing of interests test should be applicable when the state seeks a warrant or an order compelling a prisoner to give a voice exemplar. The nexus between the compelled voice exemplar for voiceprint purposes and the criminal activity is the use of the voiceprint to identify the defendant at trial. But, until the voiceprint technique is shown to be admissible in evidence in the courts of a particular jurisdiction, the state has no justifiable interest in compelling an individual to submit a voice exemplar.

Due Process

Although the method of taking a voice exemplar may raise a due process problem, it is difficult to distinguish the substance of the procedure from that used in securing handwriting exemplars, or possibly fingerprints. While identifications made via direct confrontations have been questioned due to their "highly suggestive atmosphere,"¹²⁶ submission of a known and only one unknown voice sample to a voiceprint technique expert is not tantamount to the direct confrontations condemned by the courts. Yet it is conceivable that the content of the voice samples, e.g., a confession, could be so highly suggestive that it could influence psychologically the subjective evaluation of the spectrograms by the expert.

Another possible due process problem arises when the defendant is compelled to make an exemplar of the incriminating words in an obscene phone call case. Though the exemplar would not constitute self-incrimination in most jurisdictions, it could be a re-enactment of the crime. However, since the exemplar would

123. *Warden, Maryland Penitentiary v. Hayden*, 387 U.S. 294 (1967).

124. *Schmerber v. California*, 384 U.S. 757 (1966).

125. *Terry v. Ohio*, 392 U.S. 1 (1967).

126. *Biggers v. Tennessee*, 390 U.S. 404, 408 (1968) (Douglas dissenting).

not be played to the jury, and so long as the voiceprint technique expert does not testify as to its contents when making his identification, this issue should be obviated.

Admissibility

To determine admissibility of scientific evidence, courts are confronted with two distinct problems. The first is the use of experts to evaluate the reliability and admissibility of the new scientific technique or device itself. The second is the qualification of experts who are to render opinions on the applications of judicially accepted scientific devices (*e.g.*, radar) or scientific techniques (*e.g.*, fingerprint and ballistic identifications). The test used in most jurisdictions to determine whether a new scientific innovation is sufficiently reliable to be admitted into evidence is drawn from dictum in *Frye v. United States*¹²⁷ which requires "general acceptance in the particular field."¹²⁸ This standard for admissibility has been subjected to heavy criticism by the commentators,¹²⁹ who point out that most courts use the test only to exclude scientific devices which evaluate subjective factors.¹³⁰ In California, for example, the *Frye* test is supposedly used;¹³¹ yet, in *People v. Williams*,¹³² the court admitted the Nalline test of narcotic use even though the experts testified that it could not "be truthfully said that the Nalline test has met with general acceptance by the medical profession" ¹³³

It is apparent that a more sophisticated test must be formulated for determining the admissibility of a new scientific innovation. Since the courts generally are not competent to evaluate the scientific data submitted to support the findings urged as verification of the new device or technique, they must rely upon advice of scientific experts who conduct independent

127. 293 F. 1013 (D.C. Cir. 1923).

128. *Id.* at 1014.

129. C. McCORMICK, *THE LAW OF EVIDENCE* 363 (1954); J. RICHARDSON, *MODERN SCIENTIFIC EVIDENCE* n.17 at 15, n.20 at 16, 132-34 (1961).

130. McCORMICK, *supra* note 129; RICHARDSON, *supra* note 129, at 153-55.

131. *Huntingdon v. Crowley*, 64 Cal. 2d 647, 653-54, 414 P.2d 382, 388, 51 Cal. Rptr. 254, 260 (1966).

132. *People v. Williams*, 164 Cal. App. 2d 858, 331 P.2d 251 (1958), *cited with approval* in *Huntingdon v. Crowley*, 64 Cal. 2d at 654, 414 P.2d at 388, 51 Cal. Rptr. at 260.

133. 164 Cal. App. 2d at 862, 331 P.2d at 253.

investigations and present their conclusions. When making the initial determination of admissibility, courts should not tabulate a quantitative "general scientific acceptance," but rather, should establish criteria for a qualitative evaluation of the experts who are urging the reliability of the new innovation. *Williams* and *People v. King*,¹³⁴ a recent voiceprint case, represent a new trend in this direction by the California courts.

If such a qualitative test is adopted, then the court will have to ascertain what scientific disciplines are relevant to an evaluation of the data which is urged in support of the new device or technique. Once this is accomplished, the testimony of the appropriate experts will be weighed in accordance with their expertise and background in the relevant discipline. There should be a presumption against admissibility. Thus, the proponent of the evidence not only will be required to present experts, but also to demonstrate that the discipline in which their expertise lies is the relevant one.

The voiceprint technique is particularly apropos for application of this standard. While there is a split among the experts as to the reliability of the technique, the experts represent different scientific disciplines. Instead of summarily rejecting the technique because of this split,¹³⁵ the court should apply the qualitative test to determine which of these experts have the relevant expertise, and then accept the views of those who qualify. This appears to have been the approach of the court in *People v. King*,¹³⁶ although the testimony at the trial was not directed to meeting this qualitative test.

The voiceprint technique is at an awkward juncture today. The relevant scientific disciplines are linguistics and speech, and it is the experts in these fields who are the most virulent opponents of the technique. Other than the developer himself, who is not a linguist but an acoustical engineer, there is no one in the relevant scientific disciplines ready to vouch for the technique. Even the linguists at Michigan State University and Bell Telephone Laboratories, who have received all the data, are not yet ready to testify as to the reliability of the technique for positive speaker identification. As a result, not only does the

134. *People v. King*, 266 Adv. Cal. App. 466, 72 Cal. Rptr. 478 (1968).

135. *State v. Cary*, 99 N.J. Super. 323, _____, 239 A.2d 680, 685 (1968).

136. *People v. King*, 266 Adv. Cal. App. 466, 72 Cal. Rptr. 478 (1968).

technique fail to meet the *Frye* test of general acceptance, but it also is presently unable to meet the burden under the more sophisticated qualitative test for initial admissibility. This is not to say that the voiceprint technique is not reliable; rather, it must be corroborated by at least some other scientists in the relevant scientific disciplines.

VII. CONCLUSION

The voiceprint technique has been demonstrated to have great potential for identifying individuals solely by sound spectrograms of their voices. In matching-to-sample experiments it has proven to be at least as reliable as aural identification of speakers. The few authentication experiments which have been conducted suggest a high potential for positive speaker identification. However, adequate scientific acceptance in the relevant scientific disciplines for positive identification will depend on the outcome of replication experiments now being conducted, particularly those at Michigan State University. It is likely the technique will be demonstrated reliable, in which case it will be admissible in most jurisdictions as long as enough care is exercised when securing and utilizing the voice exemplars so that the state constitutional privilege against self-incrimination and federal due process are not offended. The only problem then remaining will be adequate qualification of the expert witness.

