On the Theory and Practice of Voice Identification (1979)

Pages
174

Size
5 x 9

ISBN
0309028736

Committee on Evaluation of Sound Spectrograms; Assembly of Behavioral and Social Sciences; National Research Council

On the Theory and Practice of Voice Identification

Committee on Evaluation of Sound Spectrograms
Assembly of Behavioral and Social Sciences
National Research Council

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1979

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the Councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the Committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

International Standard Book Number 0-309-02873-6

Library of Congress Catalog Card Number 79-63355

Available from

Office of Publications
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Printed in the United States of America

Committee on Evaluation of Sound Spectrograms

RICHARD H. BOLT *Chairman*	Bolt Beranek and Newman, Inc.
FRANKLIN S. COOPER	Haskins Laboratories
DAVID M. GREEN	Department of Psychology and Social Relations, Harvard University
SANDRA L. HAMLET	Department of Hearing and Speech Sciences, University of Maryland
JOHN G. McKNIGHT	Magnetic Reference Laboratory, Mountain View, California
JAMES M. PICKETT	Hearing and Speech Center, Gallaudet College
OSCAR I. TOSI	Department of Audiology and Speech Science, Michigan State University
BARBARA D. UNDERWOOD	Yale Law School, Yale University
DOUGLAS L. HOGAN	*Study Director*
WALDENA BANKS	*Secretary*

Contents

Preface

The Federal Bureau of Investigation (FBI) in March 1976 requested the National Academy of Sciences to undertake an evaluation of the use of sound spectrograms for identifying speakers from the sounds of their voices. The FBI observed that the preceding 15 years had brought about an expanding use of voice identification technology and that several kinds of related scientific experiments had been undertaken. Further, courts of law at various levels had ruled both for and against admitting evidence based on sound spectrograms, and persons who offered services in analyzing and testifying on speaker identification had established a professional organization of such practitioners. These developments had been paralleled by a widening controversy about the reliability of the technology and the admissibility of the resulting testimony.

The National Research Council in July 1976 appointed the Committee on Evaluation of Sound Spectrograms and charged it with conducting a study responsive to the request from the FBI. The Committee was organized under the Assembly of Behavioral and Social Sciences of the National Research Council. The Committee's activity to a large extent has consisted of informal, candid discussions among the Committee members, the study director, consultants, and several score persons who met with the Committee in open meetings. Further details about the Committee's activity and biographical information about the members appear in Appendix E.

The Committee during its first meeting discussed technical implications of the FBI request and agreed that

it would attempt to fulfill its responsibility by undertaking five tasks:

1. Examine talker-related characteristics of speech, their representation by sound spectrograms, and their use in the task of identifying voices.
2. Consider error rates involved in identifying voices by the use of information contained in speech sounds and study the factors that influence the error rates.
3. Suggest new or improved methods for identifying voices by the use of information contained in their speech sounds.
4. Describe the training of voice identification examiners and seek improved methods of training and testing the examiners.
5. Search the relevant scientific and legal literature, describe existing data bases for evaluating techniques of voice identification, and prepare a review paper and bibliography.

The Committee decided not to examine certain related topics that are important in their own right but are not essential to the Committee's task (see Appendix D). The excluded topics relate to such questions as: whether a tape recording is authentic or has been tampered with; what information a tape recording contains and how the information might be recovered; whether invasion of privacy has occurred in making a recording; and how to determine from a recording whether the speaker was under stress and whether the speaker was lying or telling the truth.

In making the study reported here, the Committee has attempted to maintain consistently the distinction between concepts about facts and concepts about values and to express the distinction unambiguously through the precise use of words we adopt to denote the concepts. We use the words *accuracy* and *error* to designate facts, usually of a statistical nature, that can be measured or estimated objectively. We use the words *reliability* and *acceptability* to designate value judgments as to whether information of a given accuracy or error rate is satisfactory for a particular application.

The accuracy or error rate that is judged acceptable in voice identification can vary greatly depending on the consequences of correct and incorrect identification decisions. The judgments concerning reliability and acceptability should be made by the judicial or legislative body that carries the responsibility and authority to determine

Summary

This report presents a unified discussion of technological, legal, and scientific aspects of voice identification as practiced at present and as might be improved in the future. In today's practice an examiner listens to recorded voice sounds and looks at voicegrams, which graphically represent certain features of voice sounds, in an effort to match voice samples from an unidentified person with voice samples from one or more identified persons. The technology used in transmitting, recording, reproducing, and analyzing the sounds was developed for purposes other than voice identification. The present practice is based on limited knowledge about properties of voice sounds and is conducted largely as an empirical art in which the examiner acquires skill through extensive training and experience. For a given pair of samples, the examiner typically gives one of several alternative reports, indicating either no decision or that the samples do or do not match each other with some stated level of confidence in the decision.

Courtroom cases in which witnesses have offered testimony on voice identification by this aural-visual method appeared first in 1966 and had numbered more than one hundred by the time this report was written. The judicial responses have varied widely, with rulings both admitting and rejecting voice identification evidence. To some extent, the various legal viewpoints have reflected various technical viewpoints regarding the relative accuracy to be expected of voice identification under various forensic and experimental conditions.

Scientific research in phonetics and acoustics has

produced considerable information about speech sounds as related to the speech message but relatively little information about the sounds as related to the identity of the speaker. The practice of voice identification rests on the assumption that intraspeaker variability is less than or different from interspeaker variability. However, at present the assumption is not adequately supported by scientific theory and data. Viewpoints about probable errors in identification decisions at present result mainly from various professional judgments and fragmentary experimental results rather than from objective data representative of results in forensic applications.

The Committee concludes that the technical uncertainties concerning the present practice of voice identification are so great as to require that forensic applications be approached with great caution. The Committee takes no position for or against the forensic use of the aural-visual method of voice identification, but recommends that if it is used in testimony, then the limitations of the method should be clearly and thoroughly explained to the fact finder, whether judge or jury.

Because the method is likely to continue being used to some extent in forensic applications, the Committee recommends the application of certain knowledge and techniques that are available now and could improve the method in the near future. For the same reason, the Committee points out that further improvement of the method in the more distant future could result from the use of new knowledge gained through research along lines suggested in the report.

1 The Technology and Practice of Voice Identification

As early as 1660 a witness identified a defendant by his voice.[1]

Not until 1937 did voice identification receive scientific study.[2]

Telephony introduced the possibility of voice identification at any distance.

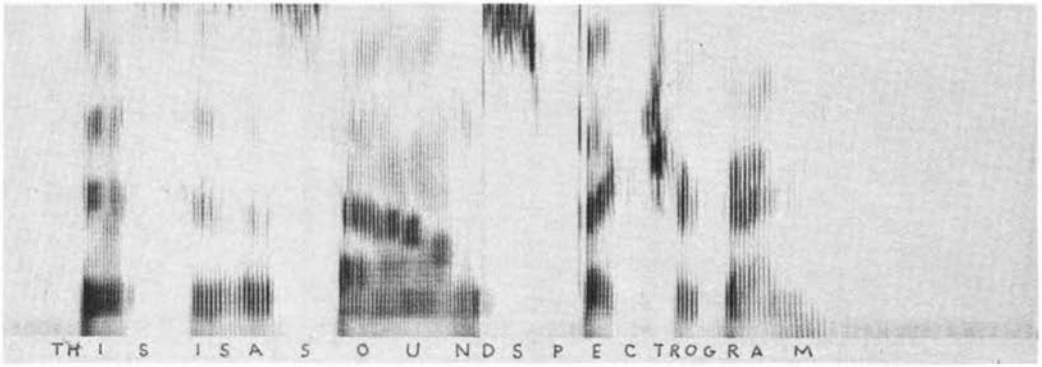
Sound recording provided the possibility of voice identification at any time.

A new instrument, the sound spectrograph, in the 1940s added the sensory capabilities of vision to those of hearing in performing voice identification.[3]

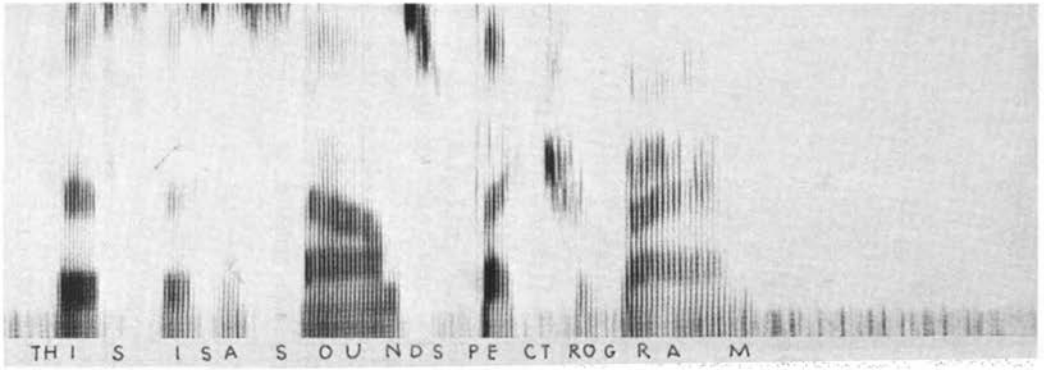
In 1966, for the first time, a court of law admitted voice identification testimony based on spectrograms of speech sounds.[4]

BACKGROUND

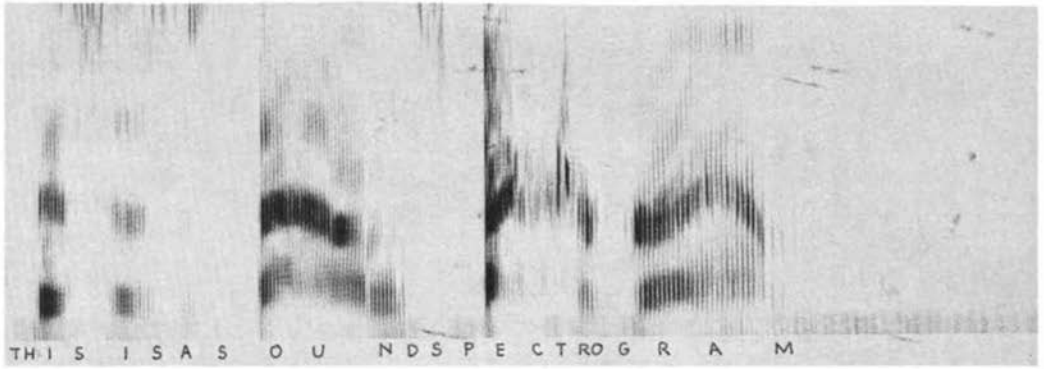
Figure 1 shows three samples of speech sound spectrograms made by a commercially available sound spectrograph. This device performs a frequency analysis of a signal and displays the frequency spectrum of that signal as a time-varying pictorial representation. In these samples the horizontal axis represents time; these samples show a segment of speech lasting about two seconds. The vertical axis represents frequency; these samples show a range of about 100 to 4000 Hz (hertz or cycles per second). The relative darkness at any point represents the relative sound level at that frequency at that time; in these



Speaker A: Day 1



Speaker A: Day 4



Speaker B: Day 1

FIGURE 1 Three samples of speech sound spectrograms.

spectrogram samples the relative darkness of the markings represents a range of about 24 decibels of sound level.

The three samples come from male speakers saying, "This is a sound spectrogram." The spectrograms on page 4 come from Speaker A; that on page 5 comes from Speaker B. These samples show that two utterances of the same speech material spoken by the same person can produce distinct differences in the spectrograms and that larger differences can appear when the same material is spoken by a different person. In some cases, the opposite can be true: the spectrogram samples obtained from two different persons can look more alike than two samples obtained from the same person. How to measure the differences and how to deal statistically with same-person and different-person spectrograms are basic questions raised in this report.

Visual displays such as the speech sound spectrograms shown here permit examination in a manner different from that afforded by listening to the sound. The eye can wander around the picture and across the time and frequency dimensions in an unconstrained way, and thereby can seek and examine small details in the physical features of the speech sounds represented graphically. In a different way, listening to the speech offers a natural and long-practiced ability to assimilate information about the meaning, nuances, dialect, and identity of the speaker. Of course neither the spectrogram examined nor the sound heard contains any voice information that was not present in the original speech signal.

When used forensically, speech spectrograms often have been called voiceprints, a term that implies that they are like fingerprints. Actually the two are fundamentally different from each other. A fingerprint is a direct graphical representation of anatomical features, the ridges in skin. The pattern of ridges for a given person remains essentially unaltered throughout that person's lifetime, and never has a case of duplication of two different fingers been discovered.[5]

In contrast to fingerprints, a speech spectrogram is an indirect representation of a complex process that produces voice sounds.[6] The process of speaking involves anatomical relationships and physiological and psychological influences that can vary from utterance to utterance spoken by the same person and can change markedly with age. Moreover, the physical system that detects, transmits, records, and prints the voice information can affect the final graphical picture in major ways. In view of these

fundamental differences between the two kinds of evidence, this report refers to spectrograms of speech sounds as voicegrams rather than voiceprints.

Further, the Committee has chosen to call the subject of this report voice identification rather than speaker identification. In principle, two different persons could have voices that are not distinguishable from each other within the limits of measurement precision available. Acoustical analysis, whether performed by listening to speech sounds or by looking at voicegrams, gives direct information about a voice but only indirect, inferential information about the identity of the person talking. The probability of correct identification of a speaker depends both upon the probability of a match between a specified voice and the voicegram it produces, and upon the probability of a match between that voice and the person to whom it is attributed.

Voice identification aided by voicegrams is a technical specialty that has become widely known in legal and investigative professions. During the past two decades the practice of voice identification has emerged as a new professional activity involving visual examination by looking at voicegrams and aural examination by listening to recorded speech sounds. The practice and technology have been applied to laboratory investigations and court proceedings concerning a variety of crimes, including kidnapping, murder, extortion, and the sale of narcotics. Courts have disagreed about whether testimony based on voicegrams should be admissible and if so under what circumstances. These and other legal aspects are discussed in Chapter 3.

EQUIPMENT AND PROCEDURES

The following description of present practice briefly describes the procedures currently recommended by the International Association of Voice Identification (IAVI), which is the only formally organized association for persons now practicing voice identification. A more detailed description of its recommended procedures is contained in Appendix A. IAVI is engaged in a continuing process of upgrading its recommended training and professional procedures in voice identification.

The present practice of voice identification as usually performed involves aural and visual comparisons of one or more known voices with an unknown voice. The aural

examination consists of listening to recordings of known and unknown voices in order to observe general similarities and differences, to screen out less useful samples, and to index the recordings that are useful for further study. The visual examination consists of analyzing and comparing the acoustic patterns of the voices as portrayed in their voicegrams. The examiner attempts to compare the acoustic patterns of identical speech elements only, such as the same phonemes, syllables, or words in each of the different voicegrams. Further, the examiner attempts to compare the speech elements within the same speech context, as is the case when the elements being compared were spoken within the same sentence or phrase, or when the elements were spoken in isolation, without any other speech sounds closely preceding or following the elements.

The equipment typically used in present practice includes tape recording and playback equipment and a sound spectrograph for making the voicegrams. Samples of known voices are recorded for use as exemplars to compare with the unknown voice sample. Sometimes the examiner uses special playback equipment that permits rapid shifting back and forth between two samples of speech for detailed comparison. Sometimes special filters are used to block out an extraneous tone, such as hum from electrical lines, or to eliminate noise in frequency ranges that do not contribute usefully to the speech sound analysis.

Either the examiner or another qualified person records the exemplars of the known voices, using procedures that have been evolved to obtain speech samples that are reasonable representatives of the known voice. The examiner or other qualified person should attempt to duplicate the physical circumstances associated with the unknown call. As soon as the voice exemplars have been recorded, they are carefully labeled and safeguarded to ensure that each exemplar is correctly identified with the voices that produced it.

Next the examiner uses the sound spectrograph to prepare the voicegrams for visual analysis. The known and unknown voice samples are played back into the spectrograph instrument. The examiners use their experience and general principles concerning contrast, darkness, and shading of the graphical images to adjust the spectrograph to yield voicegrams of "satisfactory" quality. Attempt is made to suppress extraneous noise and to enhance high-frequency information, which is believed to be of special value for voice identification. Each voicegram is labeled with the identification of the known voice or with the word "unknown." Also marked on the voicegram is the name of the

case, the date on which the print was made, and a serial number.

The examiner prepares a voicegram for examination by writing each syllable and word in the speech directly below the associated graphical marks in the voicegram. This information may be shown in normal spelling of the words, or in phonetic symbols, or both. A professional examiner is expected to perform this task with precision, because reliable voice identification requires correct association of each speech word and syllable with its corresponding pattern in the voicegram.

The actual examination usually starts with simultaneous looking at the voicegram and listening to the speech sound. The examiner concentrates on a particular segment of the speech that appears in both the unknown and the known voice sample, and switches back and forth as often as need be to arrive at a decision. If more than one known voice is involved, the procedure is repeated to compare each of the known voices with the unknown one. Sometimes comparisons among several unknown voices are made to determine whether they are likely to belong to the same person.

In carrying out the procedure, the examiner looks for unusual patterns that might be particularly important clues for identification. Especially, the examiner attempts to determine which dissimilarities in the voicegrams arise from *interspeaker* variability, indicating voices of different persons, and which dissimilarities arise from *intraspeaker* variability, indicating different utterances of a word spoken by the same person. In analyzing these clues and reaching a decision about voice identification, the examiner uses both the aural and the visual information but does not as a rule report quantitatively the relative weights given to the aural and visual parts of the examination.

Examiners are allowed to use as many samples of the voices as are available. Usually the examiners may take as much time to reach a decision as they consider necessary. After an examination is completed on a direct comparison of two voices, the examiner reports one of the following decisions:

Positive identification
Probable identification
No decision
Probable elimination
Positive elimination

Procedures for training voice identification examiners are defined only loosely at present, and they allow considerable latitude as to the trainee's background, previous training, and interaction with a mentor during the training period. Two introductory training courses are currently available. Persons taking these courses are expected to augment them by taking certain courses in speech science, by working in voice identification analysis under an appropriate mentor, and by engaging in traineeship for a period of at least two years (see Appendix A). The professional community of voice identification examiners considers a person suitably qualified to join their ranks when the person completes the training activities listed above, receives a mentor's recommendation as being proficient in voice identification, and passes an examination.

PRESENT STATUS

At the present time, the technique of voice identification is a practical methodology that is rather widely used, but that lacks a solid theoretical basis of answers to scientific questions concerning the foundations of voice identification. This disparity between practice and theory appears to be recognized by practitioners and scientists involved in the field of voice identification.

A crucial scientific question is that of speaker variability. Even though acoustical measurements can show that a person rarely if ever speaks the same word in exactly the same way twice, not much is known about the statistical description of the variability in a person's speech. Apparently even less is known about the variability of the sounds produced by different persons when they speak the same word. Without knowledge of the sources of variability, its effects on voice identification cannot be predicted or controlled with assurance. These unanswered questions about statistically valid representations of voice populations amount to a crucial shortcoming, because the very foundation of voice identification is the assumption that intraspeaker variability is less than interspeaker variability. Indeed, variability underlies essentially all aspects of the theory and use of voice identification.

The Committee recognizes that the problem of variability could be dealt with in a purely statistical manner, without consideration for the physical sources of speech sounds and their variability. The statistical approach would involve the accumulation of data from a representative sample

of speakers, words, speech sounds, biological and cultural influences, speaking environments, and other statistically significant variables. Such variables would be determined from the body of data itself, through the use of such techniques as multivariate analysis. The Committee believes, however, that a direct study of the anatomical, physiological, and cultural influences on the speech sounds and their variability would lead more efficiently to an adequate understanding of voice identification, and probably would provide a level of confidence and practical guidance that could not be reached by the statistical approach alone.

Notwithstanding the gaps in fundamental knowledge, practitioners of voice identification during a period of some 15 years have accumulated a considerable body of knowledge based on practical experience gained in forensic investigations and courtroom proceedings. Some of the practitioners have documented their experiences in voicegram analysis and testimony, and some members of the legal profession have published case histories of voice identification from the point of view of the law.[7] Several practitioners and scientists have reported laboratory studies of procedures and problems suggested by the practical applications of voice identification. The expanding literature has provided some guidance for empirical improvements in voice identification and also has served as a forum for discussion of legal aspects, including questions concerning admissibility, weight of evidence, and definition of expert, as related to voice identification based on the use of voicegrams.

This disparity between the state of development of the theory of voice identification and the state of development of the practice of voice identification is not uncommon in the evolution of a new technology. Voice identification, with its empirical advancement and its inadequacy of basic knowledge, is now at the stage of an empirical art and is moving toward the stage of an engineering practice. The final stage would be that of a fully developed technology based on science.

The development of an empirical art often starts with the emergence of a new device that suggests a new way to solve an existing problem. In this case the invention of a device for making sound spectrograms prompted their application to voice identification, even though this use was not the one for which the device originally had been designed (see note 3). As with any art, the forensic use of voicegrams has evolved mainly by trial and error. The

voicegram examiner has engaged in long periods of training, using voicegrams from known sources, and thereby has developed skill at detecting similarities and differences among speech samples that were known to come from the same speaker or from different speakers.

A practical problem for any art is that of assessing the performance that the new device or a method involving that device can deliver when used by the "very best" examiner, under the most favorable set of realistic circumstances. The result can be used, at least temporarily, as a measure of the greatest accuracy that the device or method itself can yield. A related problem is that of measuring the relative performance of other examiners. In voice identification, the sound spectrograph and the visual examination of voicegrams represent a device and a method that are used to augment the familiar method of listening to recorded samples of voices. However, the results reported to date do not appear to contain independently verifiable, empirical measures of the accuracy with which voicegram-aided identification can determine whether two samples of speech were uttered by the same person or different persons.

The engineering practice of voice identification may evolve as objective measures for assessing performance are developed empirically and as the methods of training and practice are improved. The third stage, that of a technology with a solid scientific basis, requires the parallel evolution of the science underlying voice identification. Although beginnings in this evolution have been made and several major scientific problems have been identified, the relevant information now available does not provide an adequate basis for the Committee to predict whether, and if so, when, the aural-visual process of voice identification will become a fully developed technology based solidly on science.

NOTES

1. Hulet's trial, 5 *Howell's St. Trials* 1185, 1187 (1660) (one of the trials of 29 men for high treason in the death of Charles I).
2. McGehee, F. (1937) The reliability of the identification of the human voice. *Journal of General Psychology* 17:249-271.
3. Potter, R. K. (1945) Visible Patterns of Sound. *Bell System Monograph #1368* 102:463-470. Potter, R. K.,

Kopp, G. A., and Green, H. C. (1947) *Visible Speech.* New York: D. van Nostrand (reprinted by Dover).

4. People v. Straehle, No. 9323/64 (Sup. Ct. Westchester County, 1966), noted in 12 NEW YORK L. F. 501 (1966).
5. Cummins, H., and Midlo, C. (1976) *Finger Prints, Palms and Soles: An Introduction to Dermatoglyphics.* South Berlin, Mass.: Research Publishing Co. (first published by Blakiston Co. in 1943).
6. In this report the word *voice* is used in its common meaning to include all the sounds of speech. In the science of phonetics the word *voice* is used technically to mean the sounds produced by vibration of the vocal cords, and does not include the sounds produced purely by airflow and friction.
7. For example, Jones, W. R. *Danger--Voiceprints Ahead.* 11 AMERICAN CRIMINAL LAW REVIEW 549 (1973). Greene, H. F. *Voiceprint Identification: The Case in Favor of Admissibility.* 13 AMERICAN CRIMINAL LAW REVIEW 171 (1975).

2 Scientific Aspects of Voice Identification

The practice of voice identification, as described in the preceding chapter, depends for its accuracy on the skill and judgment of the voicegram examiner. The skill and judgment, in turn, are developed mainly through extensive experience in comparing voicegrams. Clearly, so much experience would not be needed if explicit and objective criteria were available for making "match" versus "no-match" decisions. No amount of experience would suffice, however, if implicit criteria did not exist, that is, if voices did not somehow represent their speakers. The nature of these underlying relationships between voice and speaker must be discovered to provide a scientific basis for voice identification. In order to get at these relationships, we shall look again and more closely at what voicegram examiners are really doing when they use voicegrams in making a voice identification.

THE TASK OF IDENTIFYING VOICES

Listening is the initial task. Indeed, it might be the only task if the voices of the known and the unknown persons obviously differ from each other as to the sex of the speaker, dialect, or other gross characteristics, in which case, the voice identification examiner might not be called in at all. Therefore, the examiner is more likely to be called in when close similarities exist between the two voice samples, suggesting a single speaker, or perhaps when attempts at voice disguise are suspected. Thus, in real-life situations, the examiner is likely to get more "difficult" cases than "easy" cases.

What is the examiner listening for, when comparing recordings of a known and an unknown voice? The samples, taken overall, can give both general impressions and specific information about dialect, speech defects if any, and speech habits such as pauses, hesitation sounds, phrasing, inflections, and the like. For closer comparisons, using brief voice samples that are matched as to the words spoken, the examiner can listen for phonetic detail as well as for peculiarities of pronunciation.

It is for these closer comparisons that voicegrams are also used. Voicegrams provide a permanent record for scrutiny. The record shows an analysis representing the component frequencies and intensities in the time-flow of speech, patterns that the ear cannot record as such. But looking at a voicegram prompts the same questions as listening to the speech: what features characterize the speaker, what features characterize the words that were spoken, and how can these different kinds of information be separated?

For both looking and listening, much more is known about the features that characterize speech than about the features that characterize the speaker. Moreover, the inherent advantages of voicegrams are offset in large part by the fact that scientific knowledge about what to look for in the voicegram is scantier than knowledge about what to listen for in the speech sounds. This shortcoming is hardly surprising, because phonetics and dialectology were mature disciplines long before the sound spectrograph was invented. What might seem surprising is the lack of a mature scientific discipline for analyzing speech in terms that characterize the speaker, analogous to the science of phonetics for analyzing the speech sounds.

Clearly, then, one of the crucial problems--if voice identification is to progress from an art to a science-based technology--is to discover what aspects of voices point most directly and accurately to the identity of the speaker. Answers are needed for the listening task and even more so for the looking task.

Some progress is being made (see Appendix B), but the difficulty of the problem, especially with voicegrams, can be appreciated by comparing it with the related problem of discovering what aspects of speech serve to identify the phonetic elements of the message. After 30 years of research on this problem, acoustic phonetics has barely reached the stage at which it can tell a computer how to decipher a spoken message by reading the corresponding voicegram.[1]

However, knowledge about the acoustic cues for recognizing the words of a message is very helpful, albeit indirectly, in identifying voices. Such knowledge tells the examiner what similarities to ignore when comparing two voicegrams of the same word. One of the initial chores is, of course, to find and label those patterns in the voicegram that are similar because the word spoken was the same. Can the examiner then expect two patterns of the same word to be identical if, in fact, they were spoken by the same person? Actually, the two patterns are never exactly the same since even the same speaker will say the same thing variously on various occasions, but the patterns will be generally similar since the words are the same.

The examiner's task, then, is not to look for identical elements, but rather to look for similarities that would not be expected simply because the words are the same. Also, the examiner must find no differences that are greater than would be expected from a single speaker. Only then can the examiner conclude: "same voice."

Variability

The fact that voicegram patterns will differ even for the same speaker repeating the same word is crucial to voice identification. If the range of variability for a single speaker overlaps that for a different speaker who may happen to sound rather like the first, then decisions about voice identification become difficult if not highly uncertain. Thus, as already noted in the first chapter, the very foundation of voice identification is the assumption that intraspeaker variability is less than interspeaker variability. Moreover, variability *per se* underlies essentially all aspects of the theory and practice of voice identification.

Interspeaker variability of the kind just considered sets voice identification far apart from fingerprinting and may make voice identification more closely analogous to the identification of handwriting. The fact of variability raises questions about its sources. The sources of variability need to be known if the inevitable differences in voicegrams are to be explained away in attributing two non-identical patterns to the same speaker. Sources of variability certainly exist within individual speakers themselves, as already noted, but many other sources can also contribute to differences in the voicegrams.

The coexistence of interspeaker and intraspeaker variabilities raises further questions about how examiners

cope with inherently contradictory aspects of the voicegram matching task. With what expectations do they approach the task? Will they be influenced more by points of similarity than by points of difference?

After the examiner has reached a decision that is as careful and free of bias as he can make it, what are the chances that the decision may be incorrect? It can be incorrect in two ways. The decision that two voicegrams match when they do not is called an error of false identification, and the decision that two voicegrams do not match when they do is called an error of false elimination. What are the chances of false identification and false elimination? How well can the chances of error be estimated? How easily or effectively can the chances of error be reduced?

All these questions bear on the accuracy of voice identification when practiced as an art by examiners whose main qualification is experience. The same questions serve also to define domains in which science now has, or could have, a significant role in improving voice identification. The following sections of this chapter will deal in turn with what is known, and also what needs to be known, about each of these questions.

Sources of Variability

The Speaker[2] We can improve our understanding of speaker variability by making a brief digression into the nature of spoken language. Since the primary aim of speech is communication, the speakers of a given language use a common set of words and a common set of speech sounds to identify the words. Thus, when a person speaks a word or phrase, he or she tries to produce sound patterns like those of other speakers of the dialect. But only certain aspects of these sounds need to meet this social norm and to remain the same when the same word is spoken on different occasions.

For several reasons, certain aspects of the sound pattern for any particular word may be different on different occasions. For different speakers, the vocal anatomy may be different. Regardless of the speaker, some aspects of the sounds are nonessential in that they are not used to identify words, so speakers are free to produce them in various ways. Different speakers may well develop characteristically different habits in using these nonessential aspects, or a single speaker may show considerable variation in their use from one utterance to another. This

freedom allows a speaker substantial latitude in fitting speech to a situation, to a mood, to the interpersonal relationship of the speaker and the listener, and even to a temporary emotional state and to health.

All these submessages about the speaker, including the message about identity, are merged into the complex sound stream called speech, but they do not fall neatly into separate sets of acoustic features that correspond to the various submessages. Yet recovering one of these submessages is the essence of voice identification: the task is to tease out from the tangle of sound patterns those features that correspond to the speaker's vocal anatomy and habits of forming speech sounds, since these may serve to characterize him as the speaker.

In forensic situations, the difficulty in recording suitable exemplars of a suspect's voice has already been noted; this difficulty is but one example of the ways in which circumstances, emotional states, and formal versus informal modes of speech can introduce variations that must somehow be taken into account in comparing two voices and their voicegrams.

A special class of variations occurs in a person's speech when he is attempting to disguise his voice or to mimic another speaker. Several studies have been made of mimicry.[3] They show, in general, that mimics differ in how, and how well, they impersonate voices. These studies also show that the deception is usually less easily detected by listening than by looking at voicegrams.

The alternative to recognizing and making allowance for a wide range of intraspeaker variabilities is, of course, to isolate aspects of speech that reflect most directly those things about the speaker that are distinctive and unchanging. The best candidates are those physical characteristics of vocal tract over which the speaker has least control, such as the resonant characteristics of the nasal passages, and those articulatory gestures that are so deeply ingrained as habits or are so rapid that they are no longer under voluntary control. Some of the research in these areas is discussed in this chapter in connection with automatic speaker recognition and verification and also in Appendix B.

The Message Path In the usual forensic situation, samples of the unknown voice are obtained from recordings of telephone calls. There are numerous ways in which the speaker's voice can be affected, or distorted, or contaminated by noise before it reaches the voicegram examiner for

comparison with recordings of known voices. Even the known voices are subject to some of these changes. The effect, of course, is to introduce variations into the voice samples in addition to those due to the speakers themselves.

Figure 2 shows where sources of variation are to be expected, not only in the speaker but also in the transmission path and the instrumentation through which the speech passes on its way from speaker to examiner. Also, the figure points to other problem areas that will be discussed in later sections of this chapter.

Some of the things that can affect the speech message along its path from speaker to sound recorded are the acoustics of the enclosure within which the sounds were spoken, such as room noise, echoes, and the like; the telephone or microphone that converted the signal from acoustic to electrical form; the electrical transmission path, usually telephone lines and exchanges or a radio link; and any receiving equipment preceding the sound recorded. The effects of these path elements are well understood in general, but in certain situations the effects may be understood only in part or not at all because essential information is missing. For example, the presence of an echo may be recognized as an interference, but its specific effect on voice identification may not be determinable because of a lack of information about the room in which the voice sounds were uttered. In some cases, the properties of the telephone transmission line may be unknown. In any case, the distortions, added noise, and loss of higher frequencies are often severe and can add substantially to the difficulties of making voice comparisons by ear or by voicegram.

Much the same can be said about the changes caused by recording and analyzing equipment, especially if these devices are not suited to the task or are not properly maintained and used. As a practical matter, the sound recorder is often the component that does most damage to the speech, as for example when a threatening call to a police station is recorded on a "logging" recorder at low tape speed and with correspondingly low quality. At the next stage, spectrographic analysis of the recording allows a substantial range of instrumental adjustments that affect the appearance of the voicegram. These adjustments are available to help the examiner make the best possible use of the available speech recordings, but improper use of the adjustments can further degrade the sample or even influence voicegram comparisons.

Appendix B contains further discussion of these sources

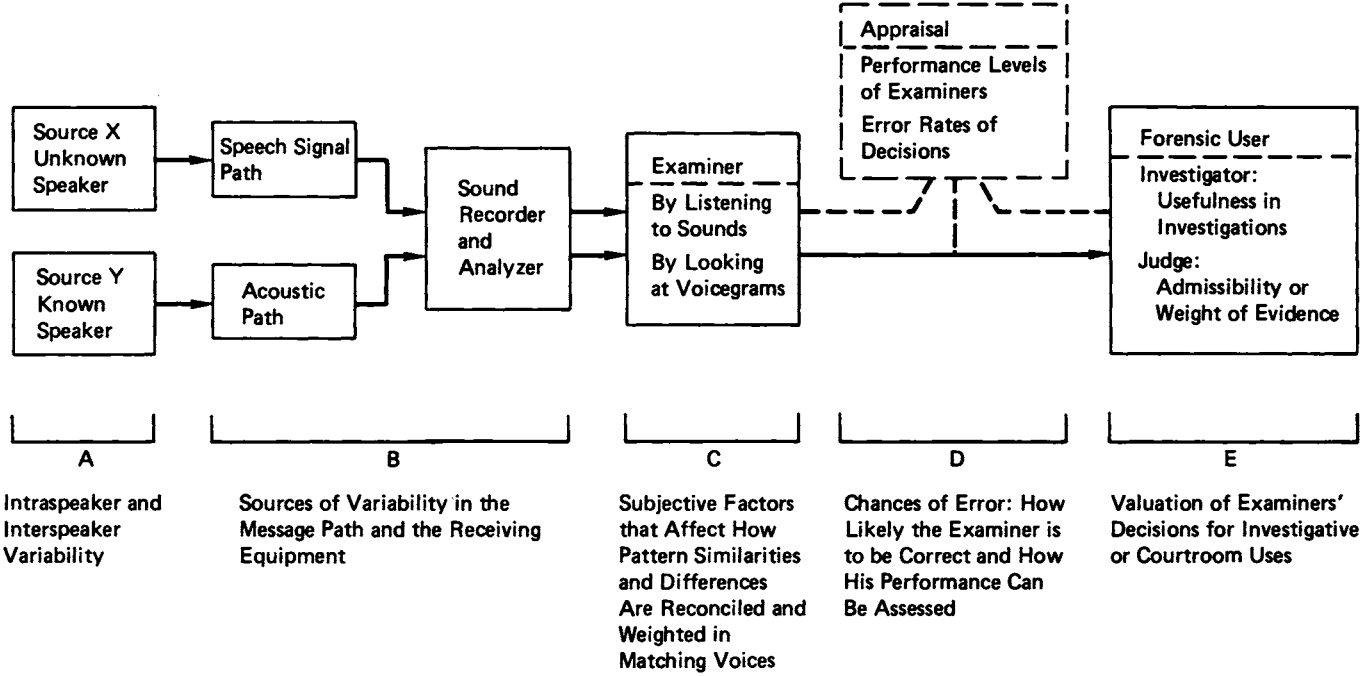


FIGURE 2 The flow of information in voice identification and the location of major problems in the technique.

of the variability in the message path and the instruments. In one of the experiments described in Appendix B, Clarke et al.[4] obtained a result that deserves mention here. The experiment, which involved listening only, measured the decrements in speaker identification and also in speech intelligibility that occurred when noise was added to the speech or when the higher frequencies of the speech were removed. Degradation of the speech signals both by noise contamination and by frequency restriction produced greater changes in speech intelligibility than in voice identification. For high-quality conditions of only slight degradation, the intelligibility score was higher than the identification score; for large amounts of degradation, the percent intelligibility was lower than the percent identification. These results suggest that the intelligibility rating of degraded speech is not a reliable indicator of its usefulness for voice identification; also, that the sound features and perceptual processes involved in the two listening tasks are probably quite different.

In general, the effect of the various sources of variability on the examiner's task is to make it substantially harder. In most cases, the cause of the speech degradations lies not in the lack of adequate technical capabilities but in the realities of practical situations.

The Examiner In the examiner's task of deciding about matches, as in the speaker's performance, we are dealing chiefly with human variability. Hence, the sources are harder to identify and the ranges of variation are harder to measure, than for the technological sources we have just been considering. Another way to put this general point about the examiner is to observe that the decisions are "subjective" and therefore likely to be variable, even though such "objective" aids as a spectrograph and voicegrams are used. This emphasizes an obvious point but an important one, namely, that the decision about a match between two voicegrams is made by the examiner and not by the spectrograph. Moreover, the decision involves a careful weighing of data that have conflicting implications and so it is not surprising that two examiners may reach different decisions by giving different weights to the similarities and differences they find in the same voice samples.

There are several factors that might lead different examiners to different conclusions about the same set of recordings. Some of the factors arise from differences in training and experience. For example, we might expect an examiner who is especially familiar with dialects and

phonetic transcription to give particular attention to the voice recordings, whereas the examiner who knows more about acoustic phonetics may concentrate on the similarities and differences in the voicegrams. Likewise, differences in familiarity with sound recording and communications technology would make some examiners more skillful than others in interpreting instrumental distortions of the speech samples. In general, overall experience with real-life cases will differ from examiner to examiner. All these factors, specific to the voice identification task, are of course imposed on the usual differences to be found within any group of persons as to intelligence, integrity, and good judgment.

In addition to these evident differences among examiners considered as a group, personal characteristics affect the decisions each would make. For example, differences exist in the self-assurance with which an examiner approaches various tasks. The differences reflect both personality and assumptions or knowledge about the specific case. Thus, some examiners may be more conservative in making a positive identification when the charge is a serious crime than when it is a minor crime. Also, in many cases, voice identification is only one part of the total evidence, and knowledge about the other evidence might influence, even if subconsciously, the examiner's view of the similarities and differences found in the voice samples.

Dealing with so many ill-defined sources of variability might seem to be impossible. Yet the problem of estimating the performance of a human decision maker in tasks somewhat similar to the matching of voicegrams has been defined scientificially and studied in depth. The underlying principles are often referred to, collectively, as statistical decision theory. Because it offers a promising alternative to the approaches used thus far in studying the accuracy of voice identification, the methods of decision theory in general, and the method known as the Receiver Operating Characteristic (ROC)[5] in particular, are presented in this report. A general description of the ROC approach appears later in this chapter and details are given in Appendix B.

THE RELATIONSHIP OF VARIABILITY TO ERROR RATES

The overall effect of variability is to increase the probability of error in the final voice identification

decision. The effect is a complex one because the different sources of variability have different quantitative effects and they differ also in the kinds of error they are most likely to cause. Speaker variability may result in speech samples that differ so much as to be judged "no match" even though they come from the same speaker. Variability in the message path or in the processing instrumentation can simply degrade the speech samples in such a way as to decrease the accuracy of matching, or the variability may distort the speech in ways that could lead directly to an erroneous decision. These kinds of variability all affect the data on which an examiner bases decisions, and further variability is introduced by the decision-making process itself. The errors resulting from variability can be of either type, errors of false identification or errors of false elimination.

The examiner's task is to detect and discount these kinds of variability as well as to find and interpret correctly the real similarities and differences between samples. Variability in the examiner's performance is bound to increase the chances of some kind of error. But errors in the decision process can go either way: the examiner can fail to detect a match between voice samples from the same speaker or can decide, incorrectly, that a match exists between samples from different speakers.

In the practical uses of voice identification, and especially in its legal uses, the crucial questions are how large the errors are likely to be in the given circumstance, how they can be estimated, and how accurately they can be estimated. Such questions will tend to resolve themselves when voice identification moves from an art toward a technology solidly based on knowledge of the features that characterize a speaker regardless of variations in voice production or transmission. At present, dependable voice features are not known and the examiner's task remains largely an empirical art. How, then, can one estimate the probabilities that a reported decision is correct or incorrect?

The usual approach has been to conduct experiments aimed at evaluating the voicegram method, either alone or by comparison with simple listening. With one exception, most of the studies were of small scale and so different from each other in experimental design as to make their results difficult to compare.

A major study using examiners and voicegrams was the Michigan State University Study (MSU).[6] This study, which attempted to simulate forensic applications, involved

variations in several source conditions: words in isolation or in the same and different contexts;[7] words recorded at the same time or with a one-month interval. A fixed path condition was used, one which approximated telephone quality. Variations that affected the receivers included the use of closed and open sets of known voices;[8] the use of six or nine words for comparison; and variation in the size of sets of knowns from 10 to 40.

Error rates for the recognition task ranged from 1 percent false identification errors in a closed-set test to 18 percent false identification and false elimination errors combined, depending on the conditions of the task. Smallest error rates occurred for matching contemporary spectrograms in small closed sets, using words spoken in isolation. Largest error rates occurred for identification of noncontemporary spectrograms in large open sets, using clue words excerpted from random context. These rates are averaged over observers.

In a study by Hazen[9] an attempt was made to determine how error rates are affected by the context in which the words were spoken. In this experiment, error rates for the two kinds of errors combined were as high as 52 percent when the samples came from different contexts. Direct comparisons between the MSU study and the Hazen study are almost impossible to make because the set sizes were different, the Hazen study used spontaneous speech whereas in the MSU study the subjects read or repeated speech, training was not identical, and not all the procedures were the same.

In another study, by Stevens et al.,[10] examination solely by listening produced results comparable with those reported in the MSU study, while examination solely by the use of voicegrams gave lower scores for correct identification. The results for the examination of voicegrams alone are not directly comparable with those of the MSU study, however, inasmuch as the subjects in the Stevens et al. study did not undergo an initial training experience.

The differences in the results of the three experiments cited above indicate that very little is really known about how well human observers perform in tasks of this kind. Moreover, only the MSU study attempted to duplicate forensic conditions.

Because forensic and laboratory conditions are so different from each other, the obvious way to get an estimate of error rates applicable to real-life conditions might seem to be to look at what has happened in court cases. However, dependable generalizations from such data are difficult to obtain, because the conditions can vary greatly from case to case and because the actual facts

about the correctness or incorrectness of voice identifications can often remain uncertain or unknown. Nevertheless, a few studies have been made (see Appendix B); the one by Smrkovski[11] offers persuasive evidence that experience and training in the use of voicegrams significantly reduces errors in voice identification.

The scientific information summarized here and reviewed in Appendix B indicates that the combined use of listening and visual examination of voicegrams as a means of discriminating between two talkers can result in far greater than chance performance. However, listening alone also can under some conditions result in greater than chance performance, and the scientific results reported to date do not provide quantitative information about the improvement in accuracy, if any, associated with the use of voicegrams. (For a discussion of the relevant scientific research, see Appendix B, p. 117-118.)

In laboratory experiments, persons with little or no initial training in voice identification have performed with false identification error rates as low as 2 percent. However, in every reported study the error rates of both false identification and false elimination have increased as the experimental conditions were changed in ways that introduced greater opportunities for variability in a person's speech signal from one recording session to the next. Further, error rates have increased when the transmission path has degraded the speech signal. Effects of degradation by introducing noise and by decreasing the frequency bandwidth were measured by Clarke et al. (see note 4) in the listening-only experiment mentioned earlier in this chapter and described in greater detail in Appendix B.

Nonetheless, experiments competently performed in the laboratory do not necessarily provide accurate estimates of error rates obtained in the forensic use of voice identification. Accurate predictions are obtainable only if the laboratory measurements and analysis correctly take into account all the factors that significantly influence the accuracy in the practical applications. A major influence is the decision process and its set of decision alternatives.

In the present practice, examiners usually report no decision when they cannot reach a match or no-match decision with confidence greater than the examiner's criterion threshold. By contrast, persons performing matching tasks in laboratory studies of voice identification usually have been required to report a decision of some kind, whether binary or scaled to more than two alternatives, on every task. The probability of deciding that two voicegrams

match when they do not match can be made smaller when the no-decision choice is allowed than when some decision must be made every time, if the examiner uses the no-decision option as an opportunity to apply a more stringent criterion. Attempts to compare laboratory results from forced-choice decisions with field results involving the no-decision option have led to some of the controversy found in the literature.[12]

As the preceding paragraph implies, a discussion of errors in voice identification is incomplete unless it includes consideration of both kinds of errors: false identification and false elimination. If an examiner tries never to report that two voicegrams match when in fact they do not, then the examiner is using the no-decision option in an attempt to minimize errors of false identification. These are the errors that the examiner will make less often than will laboratory subjects making forced-choice decisions as discussed above. But by suppressing false identification errors through the use of the no-decision option, the examiner is increasing the probability of failing to report a match when in fact a match exists The relationship of the two types of error as influenced by the human decision process is discussed below in the section on decision theory.

DECISION THEORY AND THE RECEIVER OPERATING CHARACTERISTIC IN VOICE IDENTIFICATION

The discussion thus far has dealt with the nature of the task of identifying voices and with the role of various factors, including ubiquitous variability, in influencing the probabilities of errors in an examiner's decisions. For the investigator who is working on a real-life case, or for the judge who must decide about the admissibility of evidence or consider the weight to be given to an examiner's decision, the central question is one of whether, or how much, to rely on the evidence. But the decision to rely on evidence includes value judgments that lie outside the domain of scientific method. In principle, science could provide an estimate of the probable errors involved in making a voice identification decision, but the judge, or investigator, or other user of the decision is the one who appropriately must decide how the reported degree of uncertainty would relate to the consequences of relying on the decision.

Determining whether to rely on evidence involves two distinctly different steps: first, obtaining a quantitative

estimate of the probability of error inherent in making the reported decision; second, judging whether the decision with that probability of error is acceptable for use in the particular situation involved.[13] For example, an error rate that is judged acceptable for use in resolving a dispute over a contract or will might be judged much too large to accept for use in deciding who committed a serious crime.

Improvement in the practice of voice identification will depend to a large extent on the reduction of error rates and the increase in accuracy with which the rates can be quantified. The types and probabilities of errors are influenced by both physical factors and human factors involved in the making of decisions on the basis of technical data representing the known and unknown voices in question. A unified method for analyzing error rates in relation to the influencing factors has been developed in the field of statistical decision theory.

Decision theory provides a well-established procedure, which rests upon a data plot called the Receiver Operating Characteristic (ROC), for decomposing the performance in a decision task into two independent components. The ROC curve is a graphical representation of the "power of a test" in the statistical literature. One of the components of the ROC can be thought of as a measure of the objectively determined skill of the decision maker and quality of the empirical aids and data used in making the decision. The other component can be thought of as a measure of the subjectively determined criterion by which the decision maker takes into account expectations and consequences concerning the decision. The ROC analysis is a particular application of the relation between Type I and Type II errors as commonly defined in the field of statistics.[14]

Basically, the ROC curve describes the error trade-offs that are available to a particular decision maker. Performance is characterized by various combinations of the probability of an incorrect identification and the probability of an incorrect rejection. Any decision maker can decrease the probability of one kind of error at the expense of increasing the other kind of error. In order to decrease both at once, the decision maker must improve performance by means of additional training, or more powerful aids and data, or both. Obviously, one decision maker may be better than another in the sense that for any level of incorrect identifications one has a lower rate of incorrect rejections.

For historical reasons, it is customary in the psychological literature not to plot the error trade-off itself

(probability of incorrect rejections versus probability of incorrect identifications) but an equivalent plot of probability of correct identifications versus probability of incorrect identifications. The nature of this plot is shown in Figure 3. Three curves are shown. Decision maker B is everywhere performing better than decision maker A because for each level of incorrect identification, B's probability of correct identification is higher than is A's. As the curve moves toward the upper left corner, objective performance is better. As the curve moves down and to the right toward the diagonal line D, the performance is worse. The diagonal line is the limiting case of purely chance performance: correct and incorrect identifications are equally probable. The region below and to the right of the diagonal line represents decisions that are more likely to be incorrect than if they were made by pure chance.

Curves A and B are examples of operating characteristics, functional relations that show all possible combinations of correct and incorrect decisions, on the average, that a decision maker of fixed skill using data of fixed quality can make. For more convenient use in quantitative analysis, each curve is expressed numerically by a single measure of objective performance, a measure that is proportional to the area that lies below and to the right of the curve, as shown by the shaded region in Figure 3b.

In applying the ROC process to voice identification, the objective component reflects the system producing the underlying information, the measuring instrument, and any other physical technology involved in producing the aural or visual representations of the voice samples as well as the skill of the examiner in using this information to arrive at decisions. But ultimately the examiner must make a decision as to whether two samples represent the same voice or different voices. At this point the subjective element arises. The result can be thought of as a criterion for affirming a match.

Consider the decision maker represented by the operating characteristic shown in Figure 3c. If he chooses a response criterion such that his behavior is at point *a*, then we say he has established a conservative criterion. His probability of an incorrect identification is low and as a result his probability of an incorrect rejection is necessarily high. If he chooses to respond at point *b*, then we say his criterion is lax, for while his probability of an incorrect rejection is low, that of an incorrect identification is high.

Factors that are known to affect a decision maker's

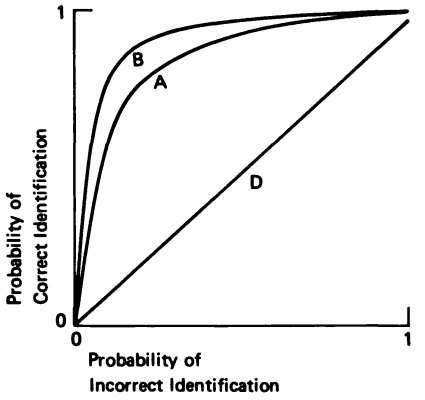


FIGURE 3a The receiver operating characteristic (ROC curve): possible performance of two different decision makers, A and B.

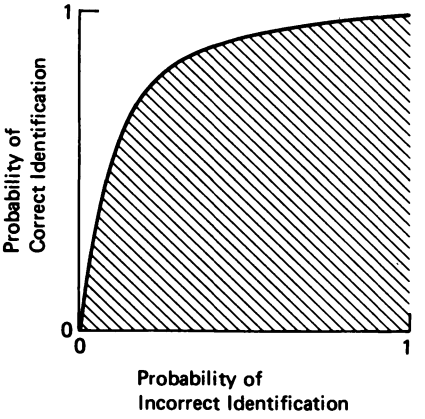


FIGURE 3b The receiver operating characteristic (ROC curve): area under the curve is a measure of how well a decision maker performs.

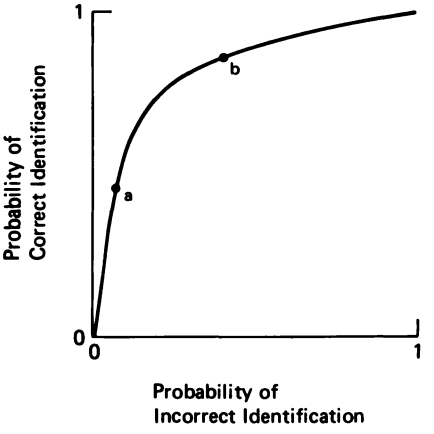


FIGURE 3c The receiver operating characteristic (ROC curve): two decision points on the same ROC curve.

criterion include his estimate of the prior probability that a match is correct and his estimate of the "payoff" matrix, which reflects the relative costs and benefits of the various correct and incorrect decisions that might be made. The choice of criterion also can be influenced by instructions to the decision maker to be more or less conservative, and quite possibly by the personality of the decision maker.

If an examiner were to work with a fixed criterion, only one point on an ROC curve would be obtained for that combination of examiner, data, and methodology. If examiners can be caused to vary their criteria by use of instructions that encourage such variation, then the form and location of the curve can be plotted quantitatively, as is illustrated in Appendix B.

For a given examiner/method/evidence situation, a variation of criterion represents movement along the characteristic curve. Thus, false identification can be greatly decreased at the expense of missing true identifications. Such a criterion variation is equivalent to placing a large penalty on false identification and giving a relatively small benefit for a correct identification.

Controlled experiments designed and analyzed in accordance with the ROC method can yield separate measures for the quality of voice identification data in aural or visual form, for the skill of the observer, and for the effectiveness of the rules and protocol used in arriving at decisions. Therefore the ROC method could be used also for rating methods of training voice identification examiners and for testing examiners themselves.

A more detailed discussion of statistical decision theory and the ROC appears in Appendix B.

STEPS TO STRENGTHEN THE SCIENTIFIC BASE OF VOICE IDENTIFICATION

Examining the task of identifying voices reveals the extent of what is still unknown in the science of voice identification. If this technique is to progress from its present status as an empirical art to a science-based technology, much more needs to be known about those aspects of speech that can be used to identify voices. Increments in scientific knowledge as well as improvements in methodology might lead to the development of voice identification as a science-based technology. Careful planning for the research would be needed in order to obtain the best

possible results with reasonable investments of time and other resources.

Here we consider some broad strategies for research on voice identification, some premises that are the foundation of voice identification but still require careful scrutiny, and some specific problem areas that are suggested by an examination of the sources of variability in voice identification.

Research strategies will, of course, be influenced by the intended forensic uses of the research results. Even though research on improving voice identification methods will go on regardless of the particular impetus provided by forensic needs, it would not be prudent simply to wait for such developments to occur. Voice identification in the forensic situation poses its own problems, and solutions developed for other applications may not be optimal; for example, techniques for improving a fully automated voice verification system for use with cooperative speakers are almost certainly not optimal when the intended use is in an interactive (human examiner + machine) analysis for use with uncooperative speakers. Nonetheless, work in voice verification may have more relevance than might be expected at first glance. For example, mimicry is a problem common to both voice verification systems and forensic applications of voice identification.

Most of the research that has been done thus far on voice identification has been aimed at an overall estimate of how well voice identification can be performed by looking at voicegrams, by listening to the speech, or both. The parameters that were manipulated in these experimental studies were mainly those that were most manageable and that seemed reasonably likely to make a difference. A major difficulty in these experiments on the methodology of voice identification is that the number of significant parameters is unmanageably large. Even the Michigan State University studies, massive as they are, had to cut many corners and did not truly approach real-life situations. Indeed, the inherent complexity of the problem probably precludes a "grand experiment" to determine once and for all the simultaneous effects of source, path, and examiner variability as all of these parameters vary over their reasonable ranges.

An alternative research approach is to conduct a series of smaller studies so planned that the experimental results will be mutually supportive and will lead to an overall understanding of the combined effects of all the parameters that are likely to be present in actual situations. Such

a strategy depends for its success on cooperation in planning experiments, in exchanging results, and in using a common data base of voice recordings so that this aspect, at least, of the total situation can be held constant across the studies. Also, some preliminary experimentation may be needed to determine which parameters are most important in avoiding errors.

The basic premises underlying a technology for voice identification likewise deserve further direct scrutiny While it is important to know how such factors as voice excitation[15] and articulatory dynamics[16] contribute to source variability, research on these topics does not go directly to the central assumption about identification by voice: that voice differences from person to person are greater than differences for any one person from time to time. Some part of the study of source variability should therefore be aimed at a direct assessment of interspeaker versus intraspeaker variations.

A second assumption implicit in present methodology is that examining voicegrams, as distinct from only listening to the voice, contributes materially to the quality of the judgment as to whether two voice samples were spoken by the same person. A related question, whether inspecting voicegrams or listening to voice sounds provides the more trustworthy information, has been studied experimentally but with inconclusive results. Perhaps the underlying question to which attention should be given is whether, or to what extent, the two procedures yield different and independent information about speaker identity.

Speaker variability could be investigated in ways that would yield information relevant to the basic aspects of speech production, to the nature of voice variability, and to the design of improved displays of information about speech sounds. For example, a set of experiments could be designed to analyze separately the contributions that the vocal tract and the vocal excitation make to aural-visual determination of identity. Such experiments could be carried out by synthesizing speech with vocal tract information copied from one person and with vocal excitation information copied from another person. This research could use resynthesis of speech samples with differing excitations and tract characteristics and could evaluate the effects of each combination on the identification performance.[17]

Other projects could investigate special aspects of speech, such as the influence of the articulatory dynamics of the motion of the jaw, tongue, and other parts of the

anatomy that affect the vocal tract. Relevant influences would include the physiology of speech and the speaking habits of different people. Mimicry also merits considerable research of a detailed acoustic, physiological, and anatomical sort, because a better understanding of how mimics perform might elucidate features that are resistant to control and therefore would be consistent indicators of speaker identity.

The variability of the message path is reasonably well understood, but more research is needed to determine quantitatively the variability introduced by each element in the path, such as a recording room, a telephone, a recording instrument, and so on (see the preceding section in this chapter on the message path). New information, combined with recently developed techniques in signal processing, might lead to significant advances. For example, work is now being done on methods of reducing noise in recorded signals and in compensating for some forms of spectral distortion. Such compensation would be especially helpful to automated methods of voice identification.

Other possible improvements might result from research involving the application of principles already well understood. For example, when the recording of samples can be controlled by an examiner, efforts can be made to use recording equipment of high quality that is well maintained and properly adjusted. Research programs as well as the forensic practice of voice identification would benefit from a specially prepared handbook that provides expert guidance in the selection, maintenance, and adjustment of recording and analyzing instruments used in the study and practice of voice identification.

Research on examiner variability may lead to several opportunities for refining and improving voice identification. The opportunities could relate both to aiding the human examiner and to developing new instruments and automated methods. For example, using applications of the decision theory methodology discussed in this chapter might lead to improved methods for training and evaluating examiners, and it also might guide the development of new instruments designed specifically for use in making voice identifications alone or in collaboration with a human observer.

Methods for classifying voices will arise out of multidimensional analysis and clustering methods. This work can draw on both the standard mathematical methods and on work that has been done in determining subjective distances along many dimensions of voice characterization.[18]

Automated methods of voice identification are being developed and are yielding information of benefit to the strengthening of the aural-visual method.[19] For practical reasons, much of the effort on automated systems has been directed to the verification of voices, for example, the voices of employees seeking entrance to a controlled-access area. Verification involves only two voice samples, one of a reference sample on file and the other of a claimant to that standard, whereas identification can involve any number of voices. In general, voice verification is the simpler task because the recording conditions are controlled, the spoken words are pre-selected, the speaker generally is cooperative, and repeated trials are easy to obtain.

The engineering methods that are being developed for voice verification mostly use computer algorithms that are based partially on an understanding of underlying principles and partially on pragmatic experimentation. The processing and analysis of the voice samples has focused largely on parameters for which analytical techniques are well developed and for which the acoustical theory of speech production is most advanced. Thus, most use has been made of pitch contours, formant frequencies and their relative amplitudes and bandwidths, and the analysis of vowels and nasal consonants. The precision with which comparisons along such dimensions can be made by automated techniques exceeds that of the human listener, although this improvement in precision is undercut to a considerable extent by the inherent variability introduced by the speaker. Nevertheless, some preliminary experiments on attempted mimicry (to circumvent a verification system) have shown that a human judge can be misled even when a computer decision process can give the correct answer. Often, however, automated processing and decision techniques have been unduly sensitive to irrelevant information (e.g., telephone channel characteristics) that a human listener can easily ignore.

Experimental work on voice verification is gaining momentum and demonstrating adequate accuracy for some uses. A recent experimental effort[20] using automated methods is the Semi-Automatic Speaker Identification System (SASIS).[21] Experimenters used only the steady-state portion of ten phonemes. The phonemes are selected by an operator with computer assistance; however, the matching and decision processes are fully automatic. Using data from 250 talkers comprising over 35,000 phonetic events, identification scores that were 97 percent correct were produced when the

10 phonemes were used for comparison. This result held across some linguistic classes, though not all were investigated. However, additional experiments using telephone-quality speech showed that the SASIS method was very sensitive to telephone channel variation.

The results with SASIS and other experiments on voice verification (see Appendix B) indicate that automated methods have practical promise. Such methods are attractive for several reasons: they promise to be fast and cost-effective, results are reproducible, and the methods are explicit and objective even when the theoretical basis is insecure.

Contributions from the engineering development of automated methods to voice identification for forensic uses are likely to come in several ways. One example is that reliable algorithms will make it possible to classify voices in much the same way that fingerprints are now classified, so that a recording of an unknown voice could be compared expeditiously with other voices that are on file. Moreover, computer matching would provide an added element of objectivity into voice identification decisions. Another contribution will come from what is learned pragmatically about features that are useful in identifying voices. This should be useful in guiding basic research; also, the speed of computer methods will be helpful in testing research results across many voice samples.

NOTES

1. Klatt, D. H., and Stevens, K. N. (1973) On the automatic recognition of continuous speech: implications from a spectrogram-reading experiment. *IEEE Transactions on Audio and Electroacoustics* AU-21: 210-216.
2. Stevens, K. N. (1972) Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. Pp. 207-232 in *Proceedings, Seventh International Congress of Phonetic Sciences* (Montreal 1971). The Hague: Mouton.
3. Hall, M. E. (1975) Spectrographic Analysis of Interspeaker and Intraspeaker Variabilities of Professional Mimicry. Thesis submitted to Michigan State University. Rosenberg, A. E. (1972) Listener performance in a speaker-verification task with deliberate imposters. Abstract. *Journal of the Acoustical Society of America* 51:132.

4. Clarke, F. R., Becker, R. W., and Nixon, J. C. (1966) Characteristics that determine speaker recognition. Technical Report ESD-TR-66-636, Bedford, Mass.: Electronic Systems Division, Air Force Systems Command, U.S. Air Force.
5. Swets, J. A. (1973) The relative operating characteristic in psychology. *Science* (Dec. 7) 182:990-1000.
6. Tosi, O., Oyer, H. J., Lashbrook, W. B., Pedrey, C., and Nicol, J. (1971) Voice identification through acoustic spectrography. Report prepared under contract N170-004 for the Michigan Department of State Police. East Lansing: Michigan State University. Tosi, O., Oyer, H. J., Lashbrook, W. B., Pedrey, C., Nicol, J., and Nash, E. (1972) Experiment on voice identification. *Journal of the Acoustical Society of America* 51:2030-2043.
7. Same context means the same word spoken in the same sentence; different context means the same word spoken in a different sentence.
8. In a closed set, the unknown is guaranteed to be the same as one of the known voices. In an open set, the unknown may or may not be the same as one of the known voices.
9. Hazen, B. (1973) Effects of differing phonetic contexts on spectrographic speaker identification. *Journal of the Acoustical Society of America* 54: 650-660.
10. Stevens, K. N., Williams, C. E., Carbonell, J. R., and Woods, B. (1968) Speaker authentication and identification: a comparison of spectrographic and auditory presentations of speech material. *Journal of the Acoustical Society of America* 44:1596-1607.
11. Smrkovski, L. (1976) Study of speaker identification by aural and visual examination of non-contemporary speech samples. *Journal of the Association of Official Analytical Chemists* 59:927-931.
12. Bolt, R. H., Cooper, F. S., David, E., Jr., Denes, P. B., Pickett, J. M., and Stevens, K. N. (1973) Speaker identification by speech spectrograms: some further observations. *Journal of the Acoustical Society of America* 54:531-534. Black, J. W., Lashbrook, W. B., Nash, E., Oyer, H. J., Pedrey, C., Tosi, O. I., and Truby, H. (1973) Reply to "Speaker identification by speech spectrograms: some further observations." *Journal of the Acoustical Society of America* 54:535-537.

13. Lowrance, W. W. (1976) *Of Acceptable Risk*. Los Altos, Cal.: William Kaufmann, Inc.
14. Wald, A. (1947) *Sequential Analysis*. New York: John Wiley.
15. Voice excitation refers to the principal source of sound for speaking, i.e., the time-varying flow of air through the vocal cords.
16. Articulatory dynamics refers to the motion of the parts of the anatomy (jaw, tongue, etc.) that affect the vocal tract shape.
17. Miller, J. E. (1964) Decapitation and recapitation: a study of voice quality. *Journal of the Acoustical Society of America* 36:2002. Wood, C. C. (1977) Source/vocal tract influences on speaker discrimination. Proceedings of the Ninth International Congress on Acoustics. Madrid. July.
18. Holmgren, G. L. (1963) Speaker recognition. Report AFCRL-63-119. Bedford, Mass.: Air Force Cambridge Research Laboratories, Office of Aerospace Research. Voiers, W. D. (1964) Perceptual bases of speaker identity. *Journal of the Acoustical Society of America* 36:1065-1073. Voiers, W. D. (1965) Performance evaluation of speech processing devices II. The role of individual differences. Report AFCRL-66-24. Bedford, Mass.: Air Force Cambridge Research Laboratories, Office of Aerospace Research.
19. Atal, S. (1976) Automatic recognition of speakers from their voices. *Proceedings of the Institute of Electrical and Electronic Engineers* 64(4):460-475. Rosenberg, A. E. (1976) Automatic speaker verification: a review. *Proceedings of the Institute of Electrical and Electronic Engineers* 64:475-487.
20. This work was carried out largely by Rockwell International under a contract with the Law Enforcement Assistance Administration of the Department of Justice. The development was based on prior work by Texas Instruments Company and Stanford Research Institute.
21. Paul, J. E., Rabinowitz, A. S., Riganati, J. P., and Richardson, J. M. (December 1974) Semi-automatic speaker identification system (SASIS) analytical studies. Unpublished Summary. C74-1184/501 Prepared for the Aerospace Corporation.

3 Forensic Aspects of Voice Identification

INTRODUCTION

From the forensic point of view voicegram comparison is considered to be a scientific approach to voice identification, even though from a scientific point of view the underlying principles and the degree of accuracy of the technique of voicegram examination are not convincingly established. The use of voicegram evidence in court is governed by the standards that govern all types of expert testimony and scientific evidence. This chapter summarizes the different legal standards that various state and federal courts have adopted in deciding whether to admit voicegram evidence and, for comparison, the treatment of nonscientific voice identification evidence. The chapter considers the extent to which aural-visual voice identification may satisfy these legal standards.[1]

Forensic Voice Identification in General

Although visual comparison of voicegrams is the first scientific approach to voice identification to be used in the courts, aural identification of voices has long played a part in legal proceedings.[2] Voice identification is a possibility when words are spoken in connection with some matter that turns out to have legal importance, when the identity of the speaker is important and unknown, and when the words were heard by a witness or recorded. When an unknown speaker has committed a crime and other evidence of identity is sparse, then voice identification may seem

particularly important.[3] Prosecutors have attempted, not always successfully, to introduce evidence of voice identification in prosecutions for a variety of crimes, including extortion,[4] bomb threats,[5] kidnapping,[6] robbery,[7] murder,[8] and the sale of narcotics.[9]

Typically, in cases involving aural voice identification, the witness compares the suspect's voice with his memory of the voice of the criminal, often heard only briefly at the scene of the crime. If a tape recording is available, the voices of the suspect and the criminal may be heard in rapid succession and the comparison can be based on short-term rather than long-term memory. Sometimes the tape is treated electronically in an attempt to improve clarity.[10] The witness may have prior familiarity with the voice of the criminal, but this is not required.

No special rules have been developed to help judges and juries evaluate the testimony of someone who makes a voice identification based on his own listening. The fact finder, that is, either the judge or the jury, is generally entitled to consider the testimony, notwithstanding objections based on such factors as the uncertainty of the identification, unfavorable listening conditions (noise, stress, or shock), and other challenges to the likely accuracy of the identification. Such factors are to be considered by the fact finder in determining the weight of the evidence and do not affect its admissibility.[11] Special rules govern the use of tape recordings in court, but they relate only to the problem of ensuring that the tape accurately reproduces the sounds heard by the witness and not to the problem of making a comparative judgment.[12]

Aural-Visual Voice Identification in the Courts

Unlike the technique of voice identification based on listening, the use of voicegrams has been treated by courts as a new scientific technique that must satisfy certain conditions before it can be used as the basis of courtroom testimony. Scientific evidence is subject to special screening before it can be presented to the fact finder, because juries and judges are thought to have limited competence to evaluate such evidence and might therefore tend to give it more weight than it deserves. Testimony based on voicegrams falls in this category because the technique is unfamiliar to the juries and judges who serve as fact finders in trials and not easily understood by them.

Testimony based on voicegrams was first admitted into evidence in 1966 in *People v. Straehle*, a perjury

prosecution that resulted in a hung jury.[13] The expert was Lawrence Kersta, whose testimony was based on spectrographic methods he had developed in previous voice identification experiments.[14]

Between 1966 and 1971 several courts refused to admit testimony based on voicegrams,[15] perhaps because of criticism of the design of the Kersta studies. Only one appellate court ruled favorably on voicegram evidence during that period.[16]

In 1971, the results of a larger study conducted at Michigan State University became available.[17] This study corrected two major design defects of the Kersta study and attempted to study the effect of some of the variables encountered in field applications. The results of the study were quite favorable, and many courts subsequently authorized the use of voicegram testimony.[18] Several subsequent studies and publications, however, expressed reservations about the Michigan State study and its conclusions about the validity of the voicegram technique in the forensic context,[19] and court approval has been less than unanimous since 1974.[20]

AURAL-VISUAL VOICE IDENTIFICATION: STANDARDS FOR ADMISSION

The rules governing admissibility of scientific expert testimony are similar, but not identical, in the various state and federal courts. In every state some rule operates to prevent court use of scientific evidence that is likely to seem more impressive than it should to a lay fact finder. The function of the legal rule is not only, or even primarily, to screen out worthless evidence based on bad science. The rule also serves to screen out evidence that has some scientific basis and persuasive force, on the theory that the strengths and weaknesses of the evidence are difficult for a lay fact finder to assess, that the risk is great that the evidence will be overvalued, and that the evidence is not sufficiently valuable to justify taking that risk. The judge makes the threshold decision to admit or exclude a particular item of scientific evidence. If the judge admits the evidence, then it can be considered by the fact finder, who in a jury trial is the jury, and who otherwise is a judge.[21]

The Frye *Test*

The rule that most jurisdictions continue to acknowledge was stated in *Frye v. United States*[22] by the United States Court of Appeals for the District of Columbia. That court in 1923 excluded the results of an early form of lie detector test, stating that while scientific expertise should generally be admitted if it is based on "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."[23]

When confronted with voicegram evidence, the courts have often, though not always, applied *Frye* but have differed in what they believe the test requires. This divergence of opinion is partly due to the ambiguity inherent in the three elements of the test: "particular field in which it belongs," "the thing from which the deduction is made," and "general acceptance." The divergence also may be due to the efforts of some courts to strain the interpretation of one or more of the three elements in order to preserve the appearance of applying *Frye* while actually applying a more liberal standard that admits a greater amount of scientific evidence.

Particular Field Some courts have stated that the field in which acceptance is required is defined by scientists with broad theoretical knowledge;[24] others have stated that "the requirement of the *Frye* rule of general acceptability is satisfied...if the principle is generally accepted by those who would be expected to be familiar with its use."[25]

The language of *Frye* seems to require acceptance of the underlying theory and not just of the technique itself. It seems to follow, then, that only those knowledgeable about theory should be qualified to testify as to acceptance. It is less clear in whose domain theoretical knowledge lies. A technician is not necessarily barred from being an expert on theoretical matters, but many courts stress the need for an advanced degree, a position at a university, or membership in scientific associations. For example, although one police officer had taken courses in speech science and examined over 180,000 voicegrams, the California Supreme Court expressed "[s]ubstantial doubt... whether [he] possessed the necessary academic qualifications which would have enabled him to express a competent opinion on the issue of [general acceptance].... This area may be one in which only another scientist, in

regular communication with other colleagues in the field, is competent to express such an opinion."[26]

A California appellate court held the testimony of an engineer who had worked on voicegrams at the Bell Telephone Laboratories to be inadequate support for admissibility, saying in part that "engineering abilities must not be confused with or made a substitute for learning and training in the fields of anatomy, medicine, physiology, psychology, phonetics, or linguistics."[27]

The Pennsylvania Supreme Court took a different position, accepting the qualifications of an experienced police officer as an expert witness. That court, however, rejected the voicegram testimony on the ground that there was disagreement in the scientific community, saying that:

> [w]e do not question the sincerity of Lieutenant Nash's testimony and we respect his considerable expertise in the area of spectrography. But his opinion, alone, will not suffice to permit the introduction of such scientific evidence into a court of law. Admissibility of the evidence depends upon the *general* acceptance of its validity by those scientists active in the field to which the evidence belongs....[28]

Perhaps the most demanding position on the qualification of experts has been taken by the Supreme Court of Michigan, requiring not only education and expertise but also impartiality with respect to the type of evidence in question.[29] The court rejected the testimony of one witness who was "an experienced police officer but not a scientist" and held that both the police officer and the scientist who directed the Michigan State study (another witness) lacked the necessary impartiality because their "reputations and careers have been built on their voiceprint work...."[30]

Principle from Which the Deduction Is Made If *Frye* requires general acceptance of an underlying scientific principle or explanatory theory, then the voicegram technique probably fails the test. As noted in Chapter 2, the theoretical principles underlying voice identification have yet to be formulated and tested. Courts have not in general confronted this problem. Instead, some courts have referred to general acceptance of the procedure. For example, one court held that the burden of the proponent is to demonstrate "that the scientific principles...were

beyond the experimental...stage *or* that the *procedure* was sufficiently established to have gained general acceptance in the particular scientific field in which it belongs" [emphasis added].[31] Under this test, the specification of the procedure becomes critical: how much extrapolation to untested situations should be permitted? If the case involves variables untested by experiment, the *Frye* test may not be satisfied. Alternatively, perhaps in such a situation *Frye* is satisfied, and the voicegram evidence should be presented to the fact finder along with an explanation that the case involves untested variables that cast doubt on the voicegram evidence.

General Acceptance The concept of acceptance involves two components. First, one may accept that under certain conditions, a forensic technique will have a certain quantifiable level of empirically established accuracy. Second, one may accept that at a given level of accuracy the technique is suitable for introduction as evidence.

Courts have in the past treated the requirement of general acceptance as a unitary matter and have paid considerable attention to the degree of consensus that exists in the scientific community. Disagreement within that community has generally been regarded as an obstacle to the admission of scientific evidence,[32] though courts have not required absolute unanimity of scientific opinion,[33] and indeed some courts have admitted scientific evidence in the face of substantial scientific controversy. Some courts have implied that general acceptance can be established by the testimony of a single witness,[34] while others have required a larger number.[35]

The Committee believes that consensus in the scientific community is obtainable and should be sought in establishing the expected level of accuracy in the use of voicegram evidence. Moreover, it is important to assess the expected accuracy of the technique in the forensic context in which the technique will be used.

One study has claimed that the technique of aural-visual voice identification is more accurate under forensic conditions than under experimental conditions.[36] The argument is that examiners can reduce their rate of false identification from 6 percent to 2 percent if they are allowed to offer no opinion, as they are under forensic conditions. Several courts have invoked this 2-percent error figure,[37] and several courts have cited testimony that the technique is more reliable in the field than in the laboratory.[38]

Other courts have suspected that forensic results would be less accurate than the results of a controlled experiment.[39]

These opposite conclusions by courts have resulted from a resolvable disagreement within the scientific and technical communities. We believe that a consensus resolution of the issue is provided by the following pair of statements. (1) If in moving from experimental to forensic conditions, the decision rule is changed from not allowing to allowing a "no decision" response, *and if all other conditions are held constant*, then a lower rate of false identification is to be expected. (2) Conversely, if in moving from laboratory to forensic conditions the decision rule is held constant *and if the voice identification process is made more complex*, as by changing from closed to open sets of suspects, from contemporary to noncontemporary exemplars, from no disguise to disguise possibility, and so on, then a higher error rate is to be expected. However, the amounts by which the error rates would change and the degree to which opposite errors would compensate each other are not determinable on the basis of scientific data available at present.[40]

Even after the expected level of accuracy has been established, the question will remain whether the technique is acceptable for courtroom use. The Committee believes that answering this question does not lie within the realm of science, and that therefore the *Frye* test should not be read to call for a scientific consensus on acceptability. That judgment should be made by the courts, taking into account not only the expected accuracy of the technique but also the feasibility of explaining to the fact finder the strengths and limitations of the technique and of present knowledge about it.

Other Formulations

Several courts have admitted voicegram testimony in circumstances that seem to provide special safeguards in addition to whatever protection may be afforded by some variant of the *Frye* test. For example, some courts have emphasized that extensive expert cross-examination or rebuttal may have reduced the risk of overvaluation by the jury.[41] Others have emphasized that voicegram evidence was used to corroborate other kinds of identification evidence.[42] Another court has admitted the testimony with special cautionary instructions to the jury.[43]

Some courts have admitted voicegram testimony in informal

proceedings not involving the issue of guilt or innocence. Examples include a probation revocation hearing[44] and a petition for habeas corpus.[45]

None of these supplementary factors--rebuttal testimony, corroboration, cautionary instructions, or use in collateral proceedings--has been made a prerequisite for admission by courts considering voicegrams, but of course any combination of the above, in addition to or in place of the *Frye* test, could be made mandatory.

The Trend

Whether a trend is developing, either toward or away from admission of voicegram evidence, is difficult to discern. Many courts admitted voicegrams in the early 1970s as a result of the Michigan State University study. Once a technique has been admitted by a large number of courts, new evidence of lack of general acceptance seems to have less impact.[46] In fact, some courts seem to adopt an approach to the *Frye* test that emphasizes previous court decisions, considering general acceptance not only by scientists but also by courts.[47]

Nevertheless, the clear trend of the early 1970s favoring admission is no longer in effect. The highest courts of California, Maryland, Michigan, and Pennsylvania have recently ruled against admitting voicegrams,[48] while the United States Court of Appeals for the Second Circuit and the Maine Supreme Court have ruled in favor of admitting them.[49]

The Federal Rules of Evidence, adopted in 1975, may have an effect on admission, but this possibility remains unclear. Recently, decisions in the Sixth Circuit and the Ninth Circuit involving other forms of scientific evidence have held that general acceptance of an "underlying explanatory theory" is still required in criminal cases, despite the fact that Rule 702 of the Federal Rules requires only that the evidence be relevant and that it "assist the trier of fact."[50] Hence *Frye* appears to have continuing vitality in federal as well as state courts.

An examination of the reported decisions to date suggests that some courts, applying the *Frye* test strictly, have found that voicegram evidence fails to satisfy that test. Other courts have either construed the *Frye* test more freely or have abandoned *Frye*, and have found evidence based on voicegrams admissible. There are alternatives to the strict *Frye* test besides straining to construe it expansively. Some of these are considered below.

FORENSIC TECHNIQUES FOR DEALING WITH EVIDENCE OF A CONTROVERSIAL CHARACTER

The principal argument against the use at trial of voicegram evidence is that such evidence will be overvalued by the fact finder. While one way to deal with that danger is to exclude the evidence, other techniques are available for dealing with problematic evidence. The section that follows outlines some of these techniques.

Exclude Voicegram Evidence from Jury Trials but Admit It in Nonjury Trials

Rules of evidence are frequently relaxed in cases that are tried without a jury.[51] It has been proposed, for example, that the rule excluding hearsay evidence be abolished in nonjury trials and the question of hearsay reliability be left to the judge.[52] About one-third of the trials at which voicegram evidence was admitted were nonjury trials, and perhaps this reasoning was part of the basis for the decision to admit.

A rule admitting voicegram evidence in nonjury trials would make sense only if a judge were less likely than a jury to believe that scientific evidence is infallible or if, through education and experience, a judge were more likely than a jury to understand the arguments made by experts and hence more likely to assign the proper weight to the evidence. As a practical matter, this difference in competence between judges and juries may not exist.[53]

Moreover, a rule limiting voicegram evidence to nonjury trials might provide a strong inducement for a defendant to insist on his right to a jury trial. In practice, such a rule would be similar to a rule requiring consent of both parties to the use of the voicegram evidence, because a nonjury trial requires the consent of the defendant and sometimes of the prosecutor.[54] The alternative of a simple consent requirement, without regard to the choice between a jury or a nonjury trial, is discussed below.

Admit Voicegram Evidence Only with a Cautionary Instruction to the Jury

Several courts that have admitted voicegram evidence have based the decision in part on the fact that an instruction was read to the jury to correct possible prejudicial effects.[55] The cautionary instructions used in the cases reported to date have been very general. In one case, for

example, the jurors were told (as summarized by the Court of Appeals) that "the spectrograms were only a basis for Lt. Nash's opinion and that they could disregard his testimony if they decided that his opinion was not based on adequate education or experience or that his 'professed science of voice-print identification' was not sufficiently reliable, accurate, and dependable. The court [further said] that they need not accept his opinion if they believed the reasons supporting it were unsound or if contradictory evidence cast doubt on it."[56] This type of instruction does not comment on specific evidence but rather leaves the job of evaluating both the qualifications of the voicegram examiner and the reliability of the technique entirely to the jury. Such an instruction may serve to prevent the jury from regarding the evidence as conclusive, but it does not help them to identify particular shortcomings. This type of instruction leaves to cross-examination and rebuttal witnesses the task of calling attention to the reduced accuracy associated with such conditions as random context and non-contemporaneity,[57] spontaneous vs. read speech,[58] female voices,[59] and voice disguise.[60]

Another possibility would be to give a more precise cautionary instruction. In one case involving aural voice identification, the judge refused to instruct the jury about the findings of a study to the effect that the accuracy of aural voice identification decreases rapidly as a function of the time interval between the perception of the criminal's voice and the perception of the suspect's voice.[61] But standard cautionary instructions might be made mandatory, if it were possible to develop generally acceptable language. Given the present uncertainty about the accuracy of voicegram identification, however, it is doubtful that adequate standard instructions could be drafted.

Finally, a third possibility would be to avoid particularity but to strengthen the cautionary language. A court could instruct a jury that the technique of aural-visual voice identification is highly controversial, and that they should be particularly careful in deciding whether to accept it. Again, such an instruction fails to assist the jurors in their task, but attempts to impress on them its importance.

The use of cautionary instructions may be ineffective for at least two reasons. First, it may be impossible to develop standard instructions that are appropriate for every case and unrealistic to expect judges to develop suitable instructions for particular cases.[62] Second, a jury may be

unable or unwilling to understand and apply a complicated cautionary instruction, either general or specific.[63]

Admit Voicegram Evidence If and Only If Both Parties Have Agreed in Advance

Polygraph evidence, which is inadmissible in most courts, is being admitted more and more frequently if the parties have entered into a formal, written stipulation to admit the results.[64] This practice is based in large part on the theory that the parties involved are best situated to know their own interests. The method of stipulation seems reasonably well suited to deal with the possibility of biased witnesses. The method seems less suitable, however, to deal with the possibility that competent and impartial witnesses will leave the fact finder confused.

Admit Voicegram Evidence If and Only If Other Evidence Corroborates the Identification

At common law a corroborating witness was required for conviction of perjury.[65] Several states require corroborative evidence of various kinds to sustain some or all convictions for rape.[66] A requirement of corroboration could take various forms. It could demand aural voice identification by someone familiar with the defendant, or simply other evidence tending to establish the defendant's guilt. Corroboration could be required in all cases, or only in serious crimes, or whenever the voice in question has features for which voicegram evidence has not been validated, such as the voice of a female or a disguised voice.[67]

However, to the extent that other evidence has influenced the voice identification examiner in reaching his decision, the other evidence cannot be considered to provide independent corroboration. Moreover, a requirement of corroboration has only limited utility in dealing with the danger that voicegram evidence may tend to be greatly overvalued by the fact finder. For if the corroborative evidence is not itself strong, the voicegram evidence still may be determinative for the jury and the danger of overvaluation still will be great.

Admit Voicegram Evidence If and Only If Opposing Experts are Scheduled to Testify, or If Some Other Method is Used to Promote Ventilation of the Issues

A striking fact about the trials involving voicegram evidence to date is the very large proportion in which the only experts testifying were those called by the state.[68] Commentators have underscored the imbalance between the state and largely indigent defendants in the area of expert testimony and investigation.[69] One way to reduce the danger that a jury will overvalue the evidence is to ensure that they hear the testimony of an expert who disputes the validity of the technique itself, and who can call attention to particular limitations of the technique in the case at hand.[70] In this manner, reducing the danger of overvaluation might be accomplished by the testimony of several expert witnesses with different views of the technique, or perhaps by a neutral expert, other than the person presenting the voicegram evidence, who is able to present the full range of relevant information. Unfortunately, the pool of available experts is still relatively small. And of course doubts will still exist about whether a battle of experts can successfully eliminate potential overvaluation of the evidence. Nevertheless, a rule requiring the testimony of adverse experts as a prerequisite to admissibility seems like a promising, though costly, approach to the problem.

NOTES

1. Comparison of voicegrams is just one possible scientific approach to voice identification. Other scientific techniques may be expected to appear with increasing frequency in court as expertise increases.
2. Such aural identification is usually a matter for testimony by lay witnesses and not by scientific experts. Courts have seldom confronted the question whether the comparison of voices by ear alone can be the subject of expert voice identification testimony. For a case allowing such expert testimony, by a person who had never met the defendant or spoken with him but made a comparison between disputed tapes and a taped voice exemplar, see United States v. Armedo-Sarmiento, 545 F.2d 785 (2nd Cir. 1976), *cert. denied*, 430 U.S. 917 (1977).
3. This is not to minimize the utility of voice identification in civil proceedings. *See, e.g.*, LeRoy v.

Sabena Belgian World Airlines, 344 F.2d 266, 274 (2d Cir.), *cert. denied*, 382 U.S. 878 (1965) (voice recording authenticated as evidence of liability for plane crash); *in re* Roth's Estate, 15 Ohio Op. 2d 234, 170 N.E.2d 313 (1960) (voice recording authenticated as evidence of gifts made by decedent prior to death).

4. People v. Kelly, 17 Cal. 3d 24, 549 P.2d 1240, 130 Cal. Rptr. 144 (1976); Reed v. State, ___ Md. ___, 391 A.2d 364 (1978); State v. Andretta, 61 N.J. 544, 296 A.2d 644 (1972) (requiring defendants to submit to voicegram test but refusing to determine admissibility, which was left to discretion of trial judge).
5. United States v. Otero-Hernandez, 418 F. Supp. 572 (M.D. Fla. 1976) (aural lineup of five voices).
6. Commonwealth v. Lykus, 367 Mass. 191, 327 N.E.2d 671 (1975)
7. State v. Herbert, 63 Kan. 516, 66 P. 235 (1901); People v. Sullivan, 290 Mich. 414, 287 N.W. 567 (1939).
8. Commonwealth v. Topa, 471 Pa. 223, 369 A.2d 1277 (1977) (voicegram evidence held inadmissible).
9. Typically in these cases an informer makes a purchase while an agent listens to or records the transaction by means of a wiretap or a concealed radio transmitter. The state's case generally consists of the testimony of the informer, often an individual whose previous convictions impair his credibility, corroborated by the testimony of the agent and perhaps a tape recording. The agent asserts that, upon arresting the defendant, he recognized his voice to be that of the unknown speaker in the radio transmission. *See, e.g.*, United States v. Walker, 320 F.2d 472 (6th Cir.), *cert. denied*, 375 U.S. 934 (1963); United States v. Sansone, 231 F.2d 887 (2d Cir.), *cert. denied*, 351 U.S. 987 (1956); *see also* United States v. Williams, 583 F.2d 1194 (2d Cir. 1978) (voicegram evidence admitted); People v. Tobey, 401 Mich. 141, 257 N.W. 2d 537 (1977) (voicegram evidence admitted at trial, but held reversible error on appeal).
10. United States v. Madda, 345 F.2d 400, 402-03 (7th Cir. 1965).
11. *E.g.*, Massey v. State, 160 Tex. Crim. 49, 53-54, 266 S.W.2d 880, 883 (1954).
12. *E.g.*, United States v. Biggins, 551 F.2d 64 (5th Cir. 1977); United States v. McMillan, 508 F.2d 101 (8th Cir. 1974), *cert. denied*, 421 U.S. 916 (1975). *See*

generally Conrad, *Magnetic Recordings in the Courts*, 40 VA. L. REV. 23, 28-35 (1954).

13. No. 9323/64 (Sup. Ct. Westchester County, 1966), noted in 12 NEW YORK L.F. 501 (1966). The hung jury is reported at *New York Times*, April 17, 1966, at 77, col. 3.
14. His testimony is discussed in 12 NEW YORK L.F. 501, 510-17 (1966). It was based on his article, "Voiceprint Identification," *Nature* 196:1253 (1962).
15. *See, e.g.*, State v. Cary, 56 N.J. 16, 264 A.2d 209 (1970).
16. United States v. Wright, 17 C.M.A. 183, 187-89, 37 C.M.R. 447, 441-53 (1967).
17. Tosi, I., Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J., and Nash, E., "Experiment on Voice Identification," *Journal of the Acoustical Society of America* 51: 2030 (1972).
18. See cases cited in Greene, *Voiceprint Identification: The Case in Favor of Admissibility*, 13 AM. CRIM. L. REV. 171, 184-185 & nn. 66-67 (1975). Greene reports that up to 1975, voicegram evidence had been admitted by 14 of the 15 federal trial judges that had ruled on the issue, and 35 of the 37 state courts that had ruled on it *Id.*
19. Bolt, R. H., Cooper, F. S., David, E. E., Jr., Denes, P. B., Pickett, J. M., and Stevens, K. N., "Speaker Identification by Speech Spectrograms: Some Further Observations," *Journal of the Acoustical Society of America* 54:531-534 (1973); Black, J. W., Lashbrook, W., Nash, E., Oyer, H. J., Pedrey, C., Tosi, O. I., and Truby, H., "Reply to 'Speaker Identification by Speech Spectrograms: Some Further Observations,'" *Journal of the Acoustical Society of America* 54:535-537 (1973); Hazen, B. M., "Effects of Differing Phonetic Contexts on Spectrographic Speaker Identification," *Journal of the Acoustical Society of America* 54:650-660 (1973); Hollien, H., "The Peculiar Case of 'Voiceprints,'" *Journal of the Acoustical Society of America* 56:210-213 (1974); Reich, A. R., Moll, K. L., and Curtis, J. F. "Effects of Selected Vocal Disguises on Spectrographic Speaker Identification," *Journal of the Acoustical Society of America* 60:919 (1976).
20. Admitting voicegram evidence: United States v. Williams, 583 F.2d 1194 (2d Cir. 1978); United States v. Baller, 519 F.2d 463 (4th Cir.), *cert. denied*, 423 U.S. 1019 (1973); United States v. Franks, 511 F.2d 25, 32-34 (6th Cir.), *cert. denied*, 422 U.S.

1042, 1048 (1975); State v. Williams, 388 A.2d 500 (Me. 1978); Commonwealth v. Lykus, 367 Mass. 191, 327 N.E.2d 671 (1975); *cf.* People v. Rogers, 86 Misc. 2d 868, 385 N.Y.S.2d 228 (Sup. Ct. 1976) (ordering defendant to furnish voice exemplar for voicegram analysis); State v. Olderman, 44 Ohio App.2d 130, 336 N.E.2d 442 (Ct. App. 1975) (same).

Rejecting voicegram evidence: United States v. McDaniel, 538 F.2d 408, 412-14 (D.C. Cir. 1976); United States v. Addison, 498 F.2d 741 (D.C. Cir. 1974); People v. Kelly, 17 Cal. 3d 24, 549 P.2d 1240, 130 Cal. Rptr. 144 (1976); Reed v. State, ___ Md. ___, 391 A.2d 364 (1978); People v. Tobey, 401 Mich. 141, 257 N.W.2d 537 (1977); Commonwealth v. Topa, 471 Pa. 223, 369 A.2d 1277 (1977).

21. In a nonjury trial, the judge wears two hats in this matter: he decides whether to admit the evidence, and, if it is admitted, he considers it along with other evidence in an effort to decide the case. In effect, he must decide whether to trust himself with the evidence. Some critics have questioned whether this bifurcation of roles is sensible, suggesting that the screening function should operate only in jury trials. See TAN (text accompanying notes) 50-51 *infra*.
22. 54 App. D.C. 46, 293 F. 1013 (1923).
23. *Id.* at 47, 293 F. at 1014.
24. United States v. Addison, 498 F.2d 741, 743-45 (D.C. Cir. 1974); People v. Kelly, 17 Cal. 3d 24, 30-32, 38-40, 549 P.2d 1240, 1244-45, 1249-50, 130 Cal. Rptr. 144, 148-49, 153-55 (1976); People v. Tobey, 401 Mich. 141, 145-48, 257 N.W. 2d 537, 538-40 (1977) (disinterested scientists).
25. Commonwealth v. Lykus, 367 Mass. 191, 203, 327 N.E.2d 671, 677 (1975).
26. People v. Kelly, 17 Cal. 3d 24, 38-39, 549 P.2d 1240, 1249-50, 130 Cal. Rptr. 144, 153-54 (1976).
27. People v. King, 266 Cal. App. 2d 437, 458, 72 Cal. Rptr. 478, 491 (1968), *quoted with approval in* People v. Kelly, 17 Cal. 3d at 39-40, 549 P.2d at 1250, 130 Cal. Rptr. at 154.
28. Commonwealth v. Topa, 471 Pa. 223, 231, 369 A.2d 1277, 1281 (1977).
29. People v. Tobey, 401 Mich. 141, 257 N.W.2d 537 (1977), *citing* People v. Barbara, 400 Mich. 352, 358, 376, 255 N.W.2d 171, 172-173, 180 (1977) (polygraph case reaffirming *Frye*).
30. *Id.* at 146, 257 N.W.2d at 539.

31. People v. Law, 40 Cal. App. 3d 69, 84, 114 Cal. Rptr. 708, 718 (1974) (emphasis added). It should be noted, however, that the California Supreme Court in its subsequent discussion of voicegrams focused on the need for acceptance of underlying principles rather than procedures. People v. Kelly, 17 Cal. 3d 24, 549 P.2d 1240, 130 Cal. Rptr. 144 (1976).
32. *See* Commonwealth v. Topa, 471 Pa. 223, 231, 369 A.2d 1277, 1281 (1977).
33. United States v. Stifel, 433 F.2d 431, 438 (6th Cir. 1970), *cert. denied*, 401 U.S. 994 (1971) (neutron activation analysis); Commonwealth v. Lykus, 367 Mass. 191, 198, 327 N.E. 2d 671, 675 (1975).
34. *E.g.*, People v. Rogers, 86 Misc. 2d 868, 873-74, 385 N.Y.S.2d 228, 232 (Sup. Ct. 1976).
35. People v. Kelly, 17 Cal. 3d 24, 37-38, 549 P.2d 1240, 1248-49, 130 Cal. Rptr. 144, 152-53 (1976).
36. Tosi et al. (1972) 5 pp. 2041-42 (see note 17) and Black et al. (1973) (see note 19). The Tosi experiment tested only the technique of visual comparison without aural comparison. The claim of the Tosi et al. study, then, was that the error rate could be improved in several ways: by adding aural comparisons, by allowing no-decision responses, and by several other devices as well.
37. People v. Rogers, 86 Misc. 2d 868, 880, 385 N.Y.S. 2d 228, 236 (Sup. Ct. 1976); Commonwealth v. Lykus, 367 Mass. 191, 201-02, 327 N.E.2d 671, 676-77 (1975).
38. Hodo v. Superior Ct., 30 Cal. App. 3d 778, 782-83, 106 Cal. Rptr. 547, 548-49 (1973) (voicegram testimony ultimately not admitted); Commonwealth v. Lykus, 367 Mass. 191, 201-02, 327 N.E.2d 671, 676-77 (1975).
39. People v. Kelly, 17 Cal. 3d 24, 35-36, 549 P.2d 1240, 1247-48, 130 Cal. Rptr. 144, 151-52 (1976) (citing comments by Tosi).
40. Bolt et al. (1973) and Black et al. (1973) (see note 19).
41. United States v. Baller, 519 F.2d 463, 466-67 (4th Cir.), *cert. denied*, 423 U.S. 1019 (1975); United States v. Franks, 511 F.2d 24, 33 (6th Cir.), *cert. denied*, 422 U.S. 1042, 1048 (1975).
42. Alea v. State, 265 So. 2d 96, 98 (Fla. Dist. Ct. App. 1972); Worley v. State, 263 So. 2d 613 (Fla. Dist. Ct. App. 1972) (expressly refusing to decide whether "voiceprint identification, standing alone, would be sufficient to sustain the identification and conviction of the defendant," *id.* at 615); State *ex rel.*

Trimble v. Hedman, 291 Minn. 442, 457, 192 N.W.2d 432, 441 (1972) (admissible to corroborate identification by means of ear); *see also*, Commonwealth v. Lykus, 367 Mass. 191, 327 N.E.2d 671 (1975) (court did not explicitly limit voicegram to use for corroboration).

43. United States v. Baller, 519 F.2d 463, 467 (4th Cir.), *cert. denied*, 423 U.S. 1019 (1975).
44. United States v. Sample, 378 F. Supp. 44, 51-54 (E.D. Pa. 1974).
45. State *ex rel.* Trimble v. Hedman, 291 Minn. 442, 192 N.W.2d 432 (1972).
46. See note 19.
47. *See, e.g.*, Reed v. State, 35 Md. App. 472, 483, 372 A.2d 243, 251 (Ct. Spec. App. 1977), *rev'd*, ___ Md. ___, 391 A.2d 364 (1978):

> An examination of the cases...will reveal that spectrographic analysis evidence is sanctioned in five States,...two federal circuits,...and by the United States District Court for the Eastern District of Pennsylvania. We believe, in the light of the decisions from those jurisdictions, that the *Frye* test has been met....

48. *See, e.g.*, People v. Kelly, 17 Cal. 3d 24, 549 P.2d 1240, 130 Cal. Rptr. 144 (1976); Reed v. State, ___ Md. ___, 391 A.2d 364 (1978); People v. Tobey, 401 Mich. 141, 257 N.W.2d 537 (1977); and Commonwealth v. Topa, 471 Pa. 223, 369 A.2d 1277 (1977).
49. United States v. Williams, 583 F.2d 1194 (2d Cir. 1978); State v. Williams, 388 A.2d 500, 504 (Me. 1978) (expressly holding *Frye* rule to be inapplicable but requiring "a showing...which satisfies the [trial judge] that the proffered evidence is sufficiently reliable to be relevant." [citation omitted]).
50. *See* United States v. Kilgus, 571 F.2d 508 (9th Cir. 1978) (Forward Looking Infrared System does not meet the *Frye* requirements when used to distinguish among night-flying planes of the same model); United States v. Brown, 557 F.2d 541, 554-59 (6th Cir. 1977) (Ion Microprobic Analysis for comparison of hair samples does not meet the requirements of *Frye*).

In the federal courts, the use of scientific evidence is governed by several provisions of the Federal Rules of Evidence. Rule 707 provides that "[i]f scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience,

training, or education, may testify thereto in the form of an opinion or otherwise."

Rule 703 provides that "[t]he facts or data in the particular case upon which an expert bases an opinion or inference may be those perceived by or made known to him at or before the hearing. If of a type reasonably relied upon by experts in the particular field informing opinions or inferences upon the subject, the facts or data need not be admissible in evidence."

Rule 403 provides that "evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence."

51. COMMITTEE ON RULES OF PRACTICE AND PROCEDURE, JUDICIAL CONFERENCE OF THE UNITED STATES, RULES OF EVIDENCE: A PRELIMINARY REPORT ON THE ADVISABILITY AND FEASIBILITY OF DEVELOPING UNIFORM RULES OF EVIDENCE FOR THE UNITED STATES DISTRICT COURTS, p. 4. Washington, D.C.: U.S. Government Printing Office (1962).
52. Davis, *Hearsay in Nonjury Cases*, 83 HARV. L. REV. 1362 (1970). *See also* Levin & Cohen, *The Exclusionary Rules in Nonjury Criminal Cases*, 119 U. PA. L. REV. 905, 925-31 (1971). *But see* Note, *Improper Evidence in Nonjury Trials: Basis for Reversal?*, 79 HARV. L. REV. 407 (1965) (suggesting that jury rules should apply to all trials unless rules for nonjury trials are drafted).
53. Note, *The Emergence of the Polygraph at Trial*, 73 COL. L. REV. 1120, 1131 (1973) (citing case arguing that jury is capable of sophisticated analysis).
54. In federal trials, a defendant can waive the right to jury trial only with the approval of the court and the consent of the prosecutor. Fed. R. Crim. P. 23(a), upheld against constitutional attack in Singer v. United States, 380 U.S. 24 (1965). Some states do not allow waivers; some limit waivers to crimes less severe than felonies. A large number of states allow a defendant to waive a jury trial without any approval; some require only the approval of the court; some only the consent of the prosecutor; some require both. *See* 51 CORNELL L.Q. 339, 342-43 & nn. 19-26 (1966).
55. *E.g.*, United States v. Baller, 519 F.2d 463, 467 (4th Cir.), *cert. denied*, 423 U.S. 1019 (1975); *cf.* People v. Rogers, 86 Misc. 2d 868, 881-882, 385 N.Y.S. 2d 228, 237 (1976) (requirement imposed in ruling on

future admissibility of voicegram evidence).

56. United States v. Baller, 519 F.2d 463, 467 (4th Cir.), *cert. denied*, 423 U.S. 1019 (1975). *See* 1 E. DEVITT AND C. BLACKMAR, FEDERAL JURY PRACTICE AND INSTRUCTIONS § 15.22 (3d ed. 1977).
57. *See* Tosi et al. (1972), pp. 2037-41 (see note 17).
58. *See* Hazen (1973), p. 659 (see note 19).
59. State *ex rel.* Trimble v. Hedman, 291 Minn. 442, 456, 192 N.W.2d 432, 440 (1971) (testimony of Ladefoged). *Contra*, *id.* (testimony of Tosi).
60. Reich et al. (see note 19).
61. United States v. Moia, 251 F.2d 255, 258 (2d Cir. 1958):

> Since the request was not for the admission of evidence, but for a mandatory instruction that such was the fact, the judge would need to find that it could not fairly be disputed before giving such an instruction.

citing 9 J. WIGMORE, EVIDENCE § 2568a (3d Ed. 1940).
62. It is perhaps unrealistic to expect judges to be able to discern, unassisted, errors in reasoning or informational lacunae when an expert testifies on only one side.
63. The available empirical studies are inconclusive on the question of the extent to which jurors are influenced by cautionary instructions. *Compare* R. SIMON, THE JURY AND THE DEFENSE OF INSANITY 213-20 (1967) (instructions affect outcome) *with* Broeder, *The University of Chicago Jury Project*, 38 NEB. L. REV. 744, 753-755 (1959) (some instructions do not affect outcome) *and* L.S.E. Jury Project, *Juries and the Rules of Evidence*, 1973 CRIM. L. REV. 208, 221-22 (mixed results).
64. See cases collected in Annot., 53 A.L.R.3d 1005 (1973).
65. 7 J. WIGMORE, EVIDENCE § 2040 at 359-60 (Chadbourn rev. 1978). *See also* Harnon, *The Need for Corroboration of Accomplice Testimony*, 6 ISRAEL L. REV. 81 (1971).
66. *See* Note, *The Rape Corroboration Requirement: Repeal Not Reform*, 81 YALE L. J. 1365, 1367-68 & nn. 13-18 (1972).
67. Arguably, such a feature could always be found, *e.g.*, the defendant was under severe stress.
68. In 25 to 30 cases around the country from 1971 to 1975, two prosecution witnesses were accepted as experts on aural-visual voice identification, and "[i]n approximately 80 percent of these cases, their

testimony was unchallenged and/or uncontradicted by other experts." Thomas, K., *Voiceprint--Myth or Miracle*, in SCIENTIFIC AND EXPERT EVIDENCE IN CRIMINAL ADVOCACY 273, 321 (J. Cederbaums & S. Arnold eds. 1975).

69. *See, e.g.*, Note, *The Indigent's Right to an Adequate Defense: Expert and Investigational Assistance in Criminal Proceedings*, 55 CORNELL L. REV. 632 (1970).

70. Rule 706 of the Federal Rules of Evidence allows for court-appointed expert witnesses. See Travis, *Impartial Expert Testimony under the Federal Rules of Evidence: A French Perspective*, 8 INT'L LAW. 492 (1974) (suggesting modifications of procedure under Rule 706). It might be argued that the jury should hear from a panel of impartial experts instead of, or in addition to, hearing from several different experts with conflicting points of view. *See* Martin, *The Proposed "Science Court"*, 75 MICH. L. REV. 1058 (1977) (debate surrounding idea of "science court"). However, when there is no real scientific consensus, it may be difficult to find an impartial expert who can effectively present all sides of the controversy to the jury. For that reason, it is usually necessary to rely on several different experts despite the more cumbersome nature of that method.

4 Findings, Conclusions, and Recommendations

That persons sometimes can identify each other by listening to the sounds of their voices is a common experience, not a matter of doubt. What is in doubt is the degree of accuracy with which identifications can be made under all sorts of conditions, especially in forensic situations, and the relative usefulness of voicegrams as a supplement to careful listening.

Review of the theory and practice of voice identification has made the Committee aware both of present limitations and of future possibilities. The Committee has seen experimental evidence that voice identification by aural-visual methods can be made under laboratory conditions with quite high accuracy, with error rates as low as 1 or 2 percent in a controlled nonforensic experiment. This observation and other evidence suggest that the practice of voice identification could develop into a mature endeavor built on scientific understanding.

At the same time, the Committee has seen a substantial lack of agreement among speech scientists concerning estimates of accuracy for voice identifications made under forensic conditions. The presently available experimental evidence about error rates consists of results from a relatively small number of separate, uncoordinated experiments. These results alone cannot provide estimates of error rates that are valid over the range of conditions usually met in practice.

The science and the practice of voice identification are presently in an ambiguous state, as are their relationships with the law. In the evolving science of voice

identification, much remains unknown: the basic characteristics that distinguish one voice from another, the distribution of these characteristics within large populations, the susceptibility of voices to voluntary control, as in mimicry, and much more. Likewise the present practice of voice identification lacks a solid basis for its operating procedures, for its training methods, and for its assertions of accuracy. Not surprisingly, courts and investigative agencies have had mixed experiences with voice identification and have not yet found clearly established principles to guide their evaluation and their acceptance or rejection of voice identification evidence. Because both the science and the current practice are relatively immature, early application to forensic problems has led to some confusion and controversy.

The Committee has sought to reduce the uncertainty and confusion by making an objective assessment of voice identification as it is practiced now and as it might be performed in the future. The Committee found that it could go only part way in this endeavor. Assessing the present status was relatively straightforward, and the results are given in this report. As to the future, however, we could make only a general judgment; a specific appraisal must await the generation of new information through further research and development.

FINDINGS

The following paragraphs summarize the principal findings of the Committee.

Human observers, of course, can obtain some information about the identity of a person both by listening to the sounds of the person's speech and by looking at sound spectrograms (voicegrams) of the speech. The aural and visual observations apparently do not provide identical information, but the degree of difference has not been established.

Voicegrams differ from fingerprints in a fundamental way, in that different utterances of a given word by a given speaker are not acoustically invariant whereas the anatomical ridges in the skin are topologically invariant. The variability among different utterances of a given word by a given speaker, at least for most speakers, seems to be less than the variability among the utterances of a given word by different speakers, but the statistical relations between the intraspeaker variability and the interspeaker variability have not been established.

The degree of accuracy, and the corresponding error rates, of aural-visual voice identification vary widely from case to case, depending upon several conditions including the properties of the voices involved, the conditions under which the voice samples were made, the characteristics of the equipment used, the skill of the examiner making the judgments, and the examiner's knowledge about the case. Estimates of error rates now available pertain to only a few of the many combinations of conditions encountered in real-life situations. These estimates do not constitute a generally adequate basis for a judicial or legislative body to use in making judgments concerning the reliability and acceptability of aural-visual voice identification in forensic applications.

Regarding the classification of voices, no usable method appears to exist at present. However, if an effective method of classifying voices were to be developed, it would assist greatly in identifying voices. The continuing development of automated methods for matching voices as discussed in Chapter 2 might lead eventually to a usable way of classifying voices.

Finally, the Committee finds sufficient interest in the potential forensic value of a more accurate process of aural-visual voice identification to justify further efforts toward its improvement.

CONCLUSIONS

Practice

The Committee concludes that some improvement in the practice of aural-visual voice identification could be achieved in the near term by applying knowledge and techniques that are available now.

Research

The Committee concludes that the full development of voice identification by both aural-visual and automated methods can be attained only through a longer-term program of research and development leading to a science-based technology of voice identification. The concluding section of Chapter 2 gives a general characterization of the kinds of research that should be included in a long-term program of research.

A broad range of research is needed to gain understanding

of those basic dimensions of speech sounds that characterize individual persons and thereby lead to the identification of voices. The research should be conducted in a coordinated program of selective experiments designed so that the results can be integrated into a unified body of scientific information to provide an overall understanding of voice identification. Selective experimentation sharply pointed at the relevant dimensions will serve better than a "grand experiment" that attempts to cover all the variables. Some preliminary experimentation may be needed in order to explore the relative effects that different variables have on the accuracy and error rates of voice identification, and thereby to determine which dimensions are most relevant.

The four main categories of topics for research should be: the origins and characteristics of variability; the relations between intraspeaker and interspeaker variability; the relations between aural and visual examination; and the potential of developments in automated methods of voice identification to make contributions to the understanding and improvement of identification performed by humans. Topics concerning variability correspond to the parts of the voice identification system in which the variability originates: the main parts are the speaker, the message path, and the examiner (see Chapter 2 for further detail).

Automated methods, now developing at an accelerating pace, build on and contribute to the basic science. They must, therefore, be considered in any broad research plan. Automated methods offer promising possibilities for a cost-effective method of voice classification. Data obtained from experiments on automated methods can provide new information to assist in developing a fuller understanding of features that contribute to voice identification.

An important initial step in developing research plans will be the development of a standard data base of voice samples that are representative of the relevant populations and of the characteristics encountered in voice identification. While building the data base would be started early in any research program, the data base would continue to grow in a way that phases each new acquisition of data with the need for those data.

A standard data base will be an essential element in the overall coordination of research programs, because it will make possible a useful degree of internal consistency across various projects as to the statistical properties of the data each project uses. The same data base will have continuing value as an adjunct to the training,

evaluation, and work of voice identification examiners in the future.

Forensic Use

The decision about whether to use the aural-visual method of voice identification for forensic purposes depends on the answers to several subsidiary questions. First, it is necessary to have some measure of the error rate associated with the technique. Second, it is necessary to decide whether in principle that error rate is acceptably low for use in the particular case or type of case. Third, it is necessary to decide whether, in practice, the nature of the error rate and the possible sources of error can be explained adequately to the lay fact finder, whether judge or jury, who will decide the case.

As to the first question, determining statistically valid error rates for aural-visual voice identification is possible at the present time only for controlled laboratory conditions. The laboratory results may provide some guidance as to the likely error rates for similar examination tasks in the forensic setting, but objectively justified error rates are virtually impossible to determine for most of the forensic experiences reported to date. An important step toward clarifying the nature of the error rates in voice identification will be taken by applying statistical decision theory as suggested in recommendation 3. In the long run, results of future research should provide a basis for obtaining realistic estimates of error rates expected in field applications.

As to the second question, determining the acceptability of a particular error rate for a particular forensic application is a value question and not a question of scientific or technical fact. It can be answered properly not by this Committee and not by the technical examiner, but only by the judicial or legislative body charged with regulating the proceeding in question. One and the same error rate can be judged very differently as to acceptability in different situations. For example, evidence that might be judged acceptable for use in a civil dispute might be judged unacceptable for use in adjudicating a serious crime.

The third question is whether the error rate and the possible sources of error can be understandably explained to the fact finder, who may be a judge or a jury. The fact finder must decide, in the end, whether to rely on the identification decision reported by the technical examiner.

In resolving that question, the fact finder can intelligently use the expert's advice only if he understands the inherent limitations of that advice in the case at hand. Therefore any presentation of voicegram evidence should be accompanied by a clear and thorough explanation of the limits of present knowledge about the accuracy of the technique. Such an explanation under present circumstances may be impossible to achieve or at least unwieldy, or it may be very costly. Here too, then, is a value judgment about whether the benefits of presenting voicegram evidence justify the costs associated with an adequate presentation. This value judgment, like the one discussed above, is properly made by the judicial or legislative body charged with the regulation and administration of the judicial system, and not by an examiner or this Committee.

RECOMMENDATIONS

The findings and conclusions reported on the preceding pages have led the Committee to make four recommendations. The first recommendation relates to the long-range acquisition of a basic and comprehensive understanding of the scientific aspects of voice identification. The other recommendations call for actions that could lead promptly to improvements in the present practice of voice identification and its use in forensic applications.

Recommendation 1:

Scientific Understanding of Voice Identification

We recommend that a mechanism be established to stimulate, guide, and coordinate a broad national program of scientific research on the processes of speech generation, transmission, and analysis as they pertain to the practice of voice identification.

The concluding section of Chapter 2 in this report discusses several steps required to strengthen the scientific base of voice identification. The steps include research performed in controlled laboratory experiments, in case studies of forensic experiences, and in theoretical analyses of hypotheses and objective data.

The tasks undertaken by the Committee were intended to yield an assessment of the present practice and a general discussion of potential improvements, but not to specify in detail the research and other steps toward strengthening

the scientific base of voice identification. Specifying the needed steps in detail is the task for the mechanism recommended here. The mechanism might take the form of a working group together with institutional connections, review processes, and staff as appropriate to the performance of this task.

Such a working group could oversee the implementation of the recommendations that follow, serve as a communication center, and guide the long-term development of voice identification. The Committee believes that translating these tasks into specific programs, projects, and budgets can best be undertaken by a full-time working group of modest size with appropriate personnel and support. The working group should be given general guidance by a small advisory committee of persons representing the legal, scientific, technological, and professional practice aspects of voice identification.

The Committee believes that the working group might appropriately receive support from agencies that have operational activities in voice identification. Such agencies have a natural interest in keeping informed about the developments in voice identification methodology and in having those developments guided in forensically useful, scientifically sound, and technologically feasible ways.

In addition to providing leadership toward carrying out this first recommendation, including the role of fostering implementation of recommendations that follow, the working group or other mechanism set up for this purpose could perform an urgently needed communication function. The Committee believes that further development of voice identification into a widely useful and acceptable process will require close interaction and interchange of information among people involved in legal, investigative, scientific, and technological aspects of voice identification. No one profession or discipline alone can provide the kind of common meeting ground that is required to foster candid, balanced consideration of all facets of the subject. The kind of meeting ground needed could be developed and fostered in the form of symposia or discussion groups organized and administered by the working group.

Recommendation 2:

Certification of Voice Identification Examiners

We recommend that a national mechanism be established

to develop objective standards and methods for testing the performance of voice identification examiners and to certify their competence as examiners.

Certification is a concept that is gaining attention in technologically based disciplines, including several that pertain to forensic applications. A certifying board is organized as an independent, not-for-profit corporation. It usually includes public interest members as well as professional specialists in the subjects involved. An important characteristic of a certification board is its national scope, so that there are uniform standards across the country as opposed to state or local licensing. We are not suggesting that certification of examiners could resolve the question whether voicegram evidence should be used for any particular purpose. Rather, the suggestion is that if voicegram evidence is on other criteria found suitable for use, then it should be developed and presented by examiners whose competence is certified to meet national standards.

The Committee sees a need for a national certification procedure that will command the confidence of the courts, the scientific community, and the public. We believe that the need can be met through (a) the mechanism of a certifying board or other organization with broad representation, including voice identification practitioners, scientists, lawyers, and public members; (b) the development and use of testing procedures to ensure acceptable levels of skill and training as the basis for certification; and (c) recertification at regular intervals.

An existing organization, the International Association of Voice Identification (IAVI), was established to perform some of these functions. However, the Committee believes that IAVI as presently constituted does not possess the broad base of representation usually considered appropriate and perhaps essential for a national certifying board.

The goals of a certifying board in voice identification, as in other evolving technologies and practices, would be to establish a profile of a qualified specialist and to design certification requirements that match the profile. As the technology and practice evolve, the profile may be expected to change.

The certification board should keep itself informed about continuing advances in the practice and training of voice identification examiners and should be available to offer advice on these matters as appropriate. The activities of the certification board should be reviewed from

time to time by advisory groups and sponsoring organizations, with a view to ensuring that the board does not acquire unduly concentrated power in determining directions of research, education, and procedures in voice identification.

Certification of examiners could contribute to making the present practice of voice identification more generally acceptable for forensic uses. A primary need for acceptability is a reasonable degree of accuracy in real-life situations and dependable knowledge about the accuracy to be expected. Certification of the skill and judgment of practitioners might improve both accuracy and knowledge about it.

Recommendation 3:

Improvements in Methods and Training

We recommend that practitioners of aural-visual voice identification make full use of certain available knowledge and techniques that could improve the voice identification method.

The application of relevant scientific and engineering techniques could yield improvements in the equipment and procedures used in recording speech sounds for identification purposes, in making exemplars of known speakers' voices, in training voice identification examiners, in performing identification tasks, and in evaluating identification results. The improvements could increase the amount of information obtained from voice samples, lower the error rates encountered in forensic applications, and lessen the confusion and controversy regarding the methods and results of aural-visual voice identification.

The recording and playback equipment used for voice identification purposes should be of high engineering quality closer to the quality of professional sound recording systems than to the quality of lower-priced audio equipment available in consumer markets. Increased errors in voice identification, including both false identification and false elimination, can result from restricted frequency bandwidth, background noise, inferior acoustics in the recording space, and other degraders of sound system quality.

Training in the performance of voice identification should include more extensive instruction in related scientific disciplines than is usually included at present.

The courses usually offered in speech sciences provide useful introductory material but should be broadened. For example, a knowledge of dialectology would show how shifts in vowel color could produce important differences between voices being examined by listening. Detailed knowledge of phonetics and practical skill in making phonetic transcriptions could provide important guidance for marking voicegram elements correctly. Examiners would be helped directly by a knowledge of acoustic phonetics, which deals with relations between phonemes and the voicegram patterns that serve mainly to carry phonetic information about the words spoken.

Principles and techniques of statistics should be applied more thoroughly to the voice identification process, especially in order to protect against incorrect interpretation of resulting decisions. An identification reportedly based on aural-visual information should be influenced as little as possible by information of other kinds, such as information that an examiner might receive through direct contact with suspects. A qualified person other than the examiner should make the exemplars, and the examination should involve several known voices, not just a single one, for comparison with the unknown voice in the manner of the traditional "lineup" of suspects.

Examiners should divide the voicegram patterns into as many distinguishable elements as possible, should rate corresponding elements in different voicegrams as being similar or different, and should take into account the number of similar and different elements per unit of duration of the speech samples. This procedure could increase the amount of information extracted from the sample and would point explicitly to contradictory implications that call for interpretation. Further, the results might help the examiner explain that uncertainty is inherent in all human decisions about the matching of complex data and that each decision should be qualified as to the degree of confidence with which the decision is reached.

Particularly important improvements could result from the use of statistical decision theory in the form of the Receiver Operating Characteristic (ROC). As now practiced, voice identification is largely an art in which long periods of training and practice lead an examiner to develop skill in matching voices. A practical problem that such an art faces is that of determining how well it is performing and how well it could perform under the best set of conditions attainable. An objective measure of the art's basic ability to discriminate between matches and

non-matches is needed, a measure that is independent of any particular examiner's criteria and biases in making decisions. The ROC provides such a measure, and also provides a means for assessing the relative competence of individual practitioners and the effects of training and experience.

Recommendation 4:

Forensic Testimony on Voice Identification

> *We recommend that if evidence on voice identification is admitted in court--and we take no position on admissibility--then the inherent limitations in the method and in the performance of examiners should be explained to the fact finder, whether the judge or the jury, in order to protect against overvaluation of such evidence.*

The Committee puts great importance on communicating exactly what it means and does not mean by the words *inherent limitations*. We do not mean that examiners using the present practice cannot correctly identify voices. We do not mean that the errors are "too great" or that the reliability is "too low."

What we do mean by inherent limitations relates to the probabilities of error in a given situation and to the degree of confidence with which the probabilities of error can be estimated. All human decisions based on complex data of the sort encountered in voice identification involve some amount of error on a statistical basis. In voice identification, a given decision can incorrectly match two voices that in fact are different, and can incorrectly reject a match of two voices that in fact are the same. The probabilities of making errors of the two kinds depend on the quality of the technique and the data, on the objective skill of the examiner, and on the subjective expectations and consequences of any given decision as judged by the examiner. Voice identification testimony should include information about both kinds of errors, and should remind the fact finder that both kinds of error are inherent in voice identification decisions.

Further, the testimony should explain that up to the present time, error rates for voice identification have been measured for only a limited number of experimental conditions. All the scientific results and forensic experiences to date, taken together, do not constitute an

adequate objective basis for determining the error rates to be expected for voice identification testimony given in forensic cases generally. Error rates reported in specific cases cannot be much more than informed guesses based on practical experience combined with fragmentary results from scientific experiments. Therefore the inherent limitations basically are limitations in information and understanding.

These limitations bear directly upon the problem of overvaluation of technical evidence. The equipment and procedures used in preparing and analyzing voicegrams involve specialized technology that can appear mysterious and overly impressive to the usual fact finder. So voicegram evidence, in common with other kinds of technical evidence, incurs a risk of being thought more powerful and less fallible than it really is. In addition, voicegram evidence, unlike some other kinds of technical evidence, does not at this stage of its development stand on a thorough foundation of quantitative information describing its capabilities in forensic practice.

This added aspect, the shortage of information, makes the protection against overvaluation especially important yet very difficult to achieve. The Committee concludes that this protection can best be achieved through the testimony of opposing experts, or perhaps through the testimony of an expert appointed by the court to explain the limitations of voicegram evidence. The Committee cannot make the judgment whether the cost of providing elaborate explanations is worth the benefit of admitting the evidence. Therefore the Committee takes no position for or against admissibility; the Committee recommends only that if voicegram testimony is to be admitted in a given case, the court should be assured that the capabilities and limitations of the method will be explained thoroughly.

Implementation of Recommendations

The successful implementation of the Committee's recommendations will require the cooperation of all who are involved in one way or another in voice identification. Those involved in the practice, law, and science must all share in this task, and it is to them that our recommendations are principally addressed. Operationally, we are calling upon the law enforcement community, the legal and judicial community, and the scientific and technological community to participate. We note that members of the International Association of Voice Identification are found in both the first and third communities.

The Committee believes that agencies involved in law enforcement might appropriately take leadership in initiating the working group or other mechanism proposed in our first recommendation. Such a mechanism could encourage both the development of a mature technology of voice identification and the long-term research required to give that technology its needed scientific foundation.

APPENDIX

A

Current Procedures in Voice Identification

This appendix describes the procedures that the International Association of Voice Identification (IAVI) currently recommends for use in voice identification tasks. The description is based on discussions with members of IAVI, which at present uses only the method of voice identification described, even though other methods are possible.

DESCRIPTION OF THE VOICE IDENTIFICATION TASK

Voice identification, as performed by members of IAVI, consists of the aural and visual comparison of one or more known voices with a questioned or unknown voice. One of five alternative decisions may be reached after each examination is completed. The aural evaluation consists of listening to recordings of known and unknown voices to determine similarities and differences. Such factors as pitch, rate of speech, accent, articulation, and pathologies are evaluated. Listening also may reveal speaking styles that are deliberately stilted and inconsistent, characteristics that imply attempts at disguise. The visual evaluation consists of examining and comparing the acoustic patterns of the speakers' voices as portrayed in their voicegrams. The examiner must compare the spectrographic patterns of similar or like phonetic elements only.

Some of the features considered by the examiner in the evaluation of the spectrographic patterns displayed include: mean frequency of vowel formants, formant bandwidths, gaps, vertical striations, durations, contours and intensity of

the formants, inter-formant energy, and patterns related to consonants and transitional energy between sounds.

The aural and visual aspects of the comparison are of equal importance. Percentage figures that represent the weight each contributes to the decision-making process cannot be assigned.

The voicegram examiner reaches one of the following conclusions:

Positive identification: an unknown voice is the same as a known voice.

Probable identification: an unknown voice and a known voice are probably the same. A qualitative estimate of the probability may be offered.

Positive elimination: an unknown voice is different from the known voices analyzed.

Probable elimination: an unknown voice and the known voices analyzed are probably different. A qualitative estimate of the probability may be offered.

No identification or elimination decision rendered: For a variety of reasons, including too noisy or distorted samples, no opinion on identification or elimination decisions can be rendered. The reason for this conclusion may be an insufficient number of samples or the samples may be of too poor a quality to be usable.

The examiner may request as many samples as he feels necessary before arriving at a conclusion and he may take as much time as he desires in arriving at a conclusion.

OBTAINING RECORDED MATERIAL

Recording Equipment

IAVI recommends that a good quality cassette or open reel tape recorder be used for recording voices. Circumstances surrounding the investigation may dictate AC or DC operation. Although 110 volt-AC operation is recommended, batteries also can be used. The investigator should be sure the batteries are fresh and periodic battery checks should be made to ensure continuous dependable operation. Mini-cassette recorders are not recommended.

Good quality recording tape should be used. Cassette tapes come in different lengths, yielding durations that range from fifteen minutes to two hours. Cassette tapes, capable of recording more than ninety minutes (C-90),

should not be used since the tape is very thin and often breaks or becomes tangled in the recorder.

Recording Techniques

While several methods can be used to record telephone calls, the use of inductive pickups is recommended. There are several types of inductive transducers on the market; however, the donut-shaped coil, which fits tightly over the earpiece of the receiver, appears to give the best results. The suction cup pickup is also capable of producing good quality recordings if attached properly. Telephone conversations have been recorded by holding a microphone close to the earpiece; this method is not recommended. Another method of recording phone calls, which may cause problems, is the direct-wire hookup to the telephone system. Telephone company assistance should be obtained when using this method.

Recording Unknown Voices Frequently in recording unknown voices, an experienced recording technician is not available because incriminating recordings must be made at unexpected times by a likely recipient of an incriminating call; therefore, the investigator should instruct the recipient carefully in the operation of the recording device. The recipient need record only the pertinent calls and should be requested to eliminate as much background noise as possible, e.g., radio, television, air conditioners, fans, or other noise-generating devices. Placing a recorder too close to a fluorescent light may also cause problems.

The unknown call is also often made to a police station. Many police stations record all incoming calls on 24-hour tape recorders. In that case the call in question should be copied using patch cords from the 24-hour recorder to a good quality cassette or open reel tape recorder. The quality of 24-hour tape recordings is often poor but if the machine has been properly maintained they are usable for purposes of identification.

The investigating officer should index and label the tape at the time the recording equipment is set up, recording name, the time, the date, the location, and the phone number of that location. Several calls can be recorded on the same tape. Using a new tape for each conversation is not necessary. After the recording has been made, the investigator should mark and identify it. The tape should then be placed in its original container and preserved until it can be submitted for analysis.

Recording Known Voices Ideally in recording known voices the investigator should attempt to duplicate the physical circumstances associated with the unknown call. These efforts should include recording the known voice with the same recording device used to record the unknown voice. Similarly, if the questioned conversation was recorded over a telephone, it is desirable to obtain the known exemplar over the telephone system using the same pickup device. If there is space on the questioned recording tape, the known exemplar can be placed on the same tape. If the telephone is to be used in obtaining the known exemplar, a recording directly into a tape recorder should be obtained simultaneously.

Several methods may be employed to obtain a recording of the voice of a suspect. The investigator may call on the phone and record the conversation. The investigator also may record the suspect's voice during an interview with the knowledge and consent of the suspect. The investigator also may record the suspect's voice without his knowledge by employing a hidden microphone; however, exemplars obtained with a hidden microphone often prove to be of insufficient quality for comparison purposes.

In situations in which the suspect refuses to give a voice exemplar, a court order may be sought requiring the defendant to speak for the purpose of voice comparison analysis. The courts have held that requiring the accused to submit voice exemplars for the purpose of identification does not violate constitutional rights.

While being recorded the suspect should be asked to identify himself and the investigator should give his own name and other pertinent information. Since the examiner must make use of the like or similar sounds in the comparison process, the suspect should utter the same words or phrases that exist in the questioned call. Two or three repetitions of the same utterances are desirable. This procedure will expedite the analysis and reduce the number of problems encountered in the examination. If a conversation is excessively long, the examiner may decide to use selected segments for analysis. This procedure will vary with the quality of the questioned recording.

The investigator should attempt to obtain a clear, intelligible recording of the suspect's normal conversational speech. As with the questioned recording, the known exemplar should be recorded in a quiet atmosphere. Both the suspect and the investigator should become familiar with the text before the exemplar is recorded. If it is obvious to the investigator that the suspect is attempting

to disguise his voice or alter his speech, the investigator should require repetitions of the phrase. A better speech sample is usually obtained if the suspect repeats the questioned phrases after someone else, than if the suspect reads from a transcript.

After the known voice exemplar is obtained, the investigator should mark and identify the tape, place it in its original container, and safeguard it until it can be sent for analysis.

ANALYSIS OF RECORDED MATERIAL

Prior to the preparation of voicegrams or recorded excerpts for comparison, the examiner should listen carefully to all of the recordings received from the investigator and check the transcripts to detect any discrepancies. If the examiner did not receive a transcript of the unknown voices, he should prepare one himself.

The recordings received by the examiner should first be dubbed onto 1.5 mil tape on open reels at 7.5 ips. in full track format. The recorded material is then ready for spectrographic analysis. Settings of the spectrograph for voice identification purposes should be: broad band filtering, linear expanded scale (60-4000 Hz.), and high shaping (12 dB/octave pre-emphasis). A voicegram of a segment of the speech to be analyzed should be produced and examined. If noise is observed in particular bands, attempts can be made to filter it without losing speech information.* In addition, a voicegram should be produced using the "flat shaping" setting. This voicegram should be compared with the previous one to determine if nonstandard equalization

---

*Filters cannot separate noise components in the same frequency location as speech components, and if the noise is not in the location of speech components, the eye can see and ignore it on the voicegram. The noise may be in the same location as the speech but different in temporal structure; the eye can see this and thus separate the noise and the speech; a filter cannot do this and will remove as much speech as noise.

These comments relate to conventional passive filters. Newer techniques using adaptive filtering offer a more powerful way to separate speech signals from noise under some conditions.

should be employed. In the event that both the known and unknown recordings contain fricatives with high frequency energy, voicegrams exhibiting frequencies up to 7 KHz. should be made for those portions of the recordings.

Preparation of Voicegrams

After each voicegram is made it should be labeled with the name of the speaker or with the word "unknown," as the case may be. The name of the case, the date, and the voicegram number should also appear on each voicegram. Below each voicegram should appear a written representation of the speech information contained on that voicegram. A phonetic transcription or normal spelling or both, one above the other, may be used for this purpose. The symbols should appear directly below the spectrographic pattern that corresponds to their oral manifestations. This "targeting" is crucial and an examiner is expected to carry out this task correctly.

Preparation of Recorded Material

There are a number of ways to prepare the recordings of the known and unknown voices so as to allow the examiner to make, as it were, side-by-side aural comparisons of the two voices. One may use multitrack tape cartridges or tape loops, two separate tape recorders, or appropriately re-recorded segments of each voice. The purpose of this procedure is to enable the examiner to compare voice samples by listening to them as closely in time as possible. Virtually any method that enables the examiner to listen to the same phrase or sentence spoken by each voice within a few seconds of each other will accomplish the desired end. Care should be taken not to degrade the quality of the original recordings in preparing the aural comparison material.

VOICE EXAMINATION

With these materials, comparison tape recordings and voicegrams from the unknown and known voices, the examiner proceeds with the examination of voices for purposes of identification or elimination.

The examiner plays back and listens to each voice alternately, as often and as long as required. The examiner

exercises critical listening by comparing perceptual features from the two voices, such as: melody pattern, pitch, quality, respiratory grouping of words, and any peculiar common features. In many cases a difference in pitch is observed between the unknown voice and the known voice. Often the incriminating recording appears to present a different pitch than the known recording exemplars. If the examiner believes that the difference is due to the incriminating recording's being at the wrong speed, he may play it back at a different speed to attempt to compensate. Justifying such speed change is likely to be difficult.

If there is more than one known voice, the procedure is repeated, comparing each of the suspected voices with the unknown one. Sometimes comparisons are made among several unknown voices to determine whether or not they belong to the same person. Eventually a panel of listeners also may be used to judge on similarities or dissimilarities of the compared voices.

In the visual phase of the examination the examiner aligns the same phonetic elements from the unknown and known voicegrams and compares them for similarities and differences in the following features: mean frequencies and apparent bandwidths (clarity) of formants, rates of change of formant frequencies, levels of components between formants, type of vertical striations and distances between them, spectral distributions of fricatives and plosives, gaps of plosives, and voice onset times of vowels following plosives. Peculiar spectral patterns are very important clues if found in the same frequency-time coordinates of both the unknown and known voicegrams; for example, the chances that a distinctive spike of energy at the same coordinates of a word could be produced by different persons seem remote.

Dissimilarities found in both sets of voicegrams can be attributed to either intraspeaker or interspeaker variation. The experience and subjective judgment of an examiner are used to determine whether the differences are due to intraspeaker or interspeaker variability, thus lending credibility toward identification or elimination respectively. Only in cases beyond doubt--subjectively speaking--may the examiner render a positive decision. In all other instances a probable identification, elimination, or no opinion should be given.

In summary, the examiner listens to the voices and visually examines the voicegrams simultaneously before yielding a decision. In the later phases of the examination

he may detect differences in pitch, phonetic variations, etc. between the compared speakers that may explain some spectrographic differences. An examination should take as long as the examiner desires. The examination need not be completed during a single session, but should be interrupted if in the judgment of the examiner this option is needed to overcome tiredness or boredom. An examiner may be advised to request an independent opinion from a colleague from time to time.

EXAMINER TRAINING

At the present time, the procedures for training examiners are only loosely defined and they allow for considerable latitude with regard to a trainee's educational background, attendance at voice identification training seminars, and the degree of interaction between the trainee and the mentor over the course of the training period. IAVI has recently formed a committee to study examiner training and qualification procedures.

In order to become a trainee one should attend a voice identification training workshop or take equivalent courses. At the present time two such programs are available. One is provided by Michigan State University (MSU) and the other by Voice Identification Incorporated (VII). The MSU program has lasted four weeks, while that offered by VII has lasted two weeks.

At the present time, the following additional requirements must be fulfilled by IAVI trainees:

Have a baccalaureate degree and/or appropriate courses in the speech sciences. This requirement is important to elevation to full membership in the IAVI. Courses should include phonetics, linguistics, speech communications, and communications electronics.

Become actively engaged in voice identification analysis tasks.

Report at regular intervals (at least quarterly) on case work to a full member or committee as designated by the Board of Directors of IAVI.

Submit testimony from the work supervisor that the trainee has actively engaged in traineeship for a period not less than two years.

Be recommended by the appointed IAVI member or committee as being proficient in voice identification warranting membership consideration.

Pass an examination prepared and administered by the Board of Directors or designated committee. This examination will test the trainee's knowledge of voice identification in the areas of theory and practice.

APPENDIX

B

Scientific Issues in Voice Identification

In this appendix we attempt to give an overall survey of scientific information that bears on the subject of voice identification. The appendix begins with a brief discussion of the production of the speech signal, and is followed by discussions of: the sources of variability in a received speech signal; statistical ways to assess the performance of scientific research in which voicegrams and/or listening comparisons were carried out by human observers; and work in automated techniques. The appendix concludes with a brief annotation of several major survey articles and an extensive bibliography.

THE SPEECH SIGNAL[1]

Speech is a time-varying acoustic signal resulting from one or more forms of acoustic excitation being dynamically changed by the articulation or movement of the vocal tract (which is comprised of the mouth and nasal cavity). The principal sources of excitation are puffs of air coming through the glottis, turbulent air being forced through a constriction (as in the "s" sound), and puffs of air suddenly released from blockages other than the glottis (as in pressure release from a lip blockage to form a "p" sound). The glottis--the space between the vocal cords--tends to vibrate at a controllable frequency when air is forced through the vocal cords from the lungs. The frequency created is called the pitch or fundamental frequency[2] (often F_0 in the scientific literature). Sounds caused by

turbulent air are called fricatives; those caused by released blockage are called plosives.

Glottal excitation (often called voiced excitation because of the fact that it consists of puffs or pulses of air and not a sinusoidal excitation) contains the harmonics of the fundamental frequency. In the case of voiced excitation, the fundamental frequency and its harmonics are modified in amplitude by the resonances of the vocal tract. In voiced sounds, especially vowels, each of the resonances of the vocal tract can be observed readily by examining the frequency spectrum of the sounds. The increased amplitude components of the vocal tract resonances are called formants (often numbered F_1, F_2, F_3, etc. in the literature).

The frequency spectrum of a complex acoustic signal may be examined by means of a display called a voicegram, which is made by a device called a sound spectrograph. A spectrograph uses effectively many analyzing filters to examine the amplitude or the energy present at a given frequency at a given time. Figures B-1 and B-2 are voicegrams of the phrase "I *can* see you," with the word *can* emphasized. An uncertainty exists when making a frequency/time examination of a signal. We may either measure frequency more accurately at the expense of time accuracy--the result of using a "narrow band" analyzing filter--or we may measure time more accurately at the expense of frequency accuracy--the result of using a "wide band" filter. Figure B-1 displays a narrow band analysis and Figure B-2 displays a wide band analysis. Note that in Figure B-1, during the vowel or voiced sounds, the fundamental frequency and the harmonics are well displayed. In Figure B-2 the vertical lines or striations that occur during the vowel sounds represent the time delineation of the individual puffs of air through the glottis. The broad bands of energy in the vowel areas represent the formants. Each formant region tends to contain several pitch harmonics. Formant behavior is more readily discerned in the wide band voicegram (Figure B-2).

When fricative or turbulent excitation occurs, vocal tract cavities behind the excitation act as anti-resonances (or energy absorbers). Thus, in the "s" sound in Figures B-1 and B-2 we may observe the effect of an anti-resonance and a resonance as shown. The excitation for the "s" is unstructured or "noise like." The "k" sound in *can* is a plosive.

Other properties of speech that can be noted in voicegrams are the prosodic features of stress and intonation.

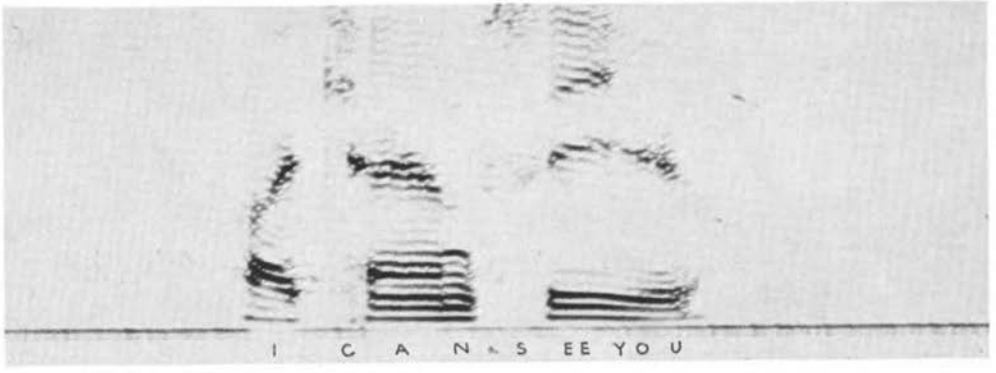


FIGURE B-1 Voicegram displaying a narrow band analysis.

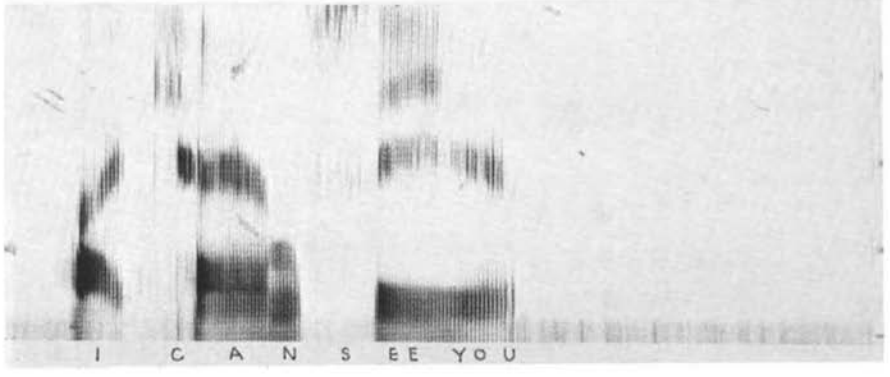


FIGURE B-2 Voicegram displaying a wide band analysis.

These features are manifested by changes in both energy and pitch. (When we emphasize words, we increase energy and pitch.)

This brief discussion of speech generation and the spectrographic representation of speech signals is necessarily simplistic and incomplete. Many other effects, major and minor, can occur. For example, coupling and uncoupling of the nasal cavities by movement of the velum or soft palate affect the acoustic signal. In summary, as vocal tract articulation takes place, dynamic variation of the speech signal occurs. No one ever produces two identical speech signals, even if he or she wants to.

THE TASK OF VOICE IDENTIFICATION

Research relevant to voice identification has been concerned largely with two areas, called here verification and identification. The task of verification (sometimes called recognition) is to make a true-false decision that a person is who he claims to be by comparing an unknown speech sample with a previously obtained speech sample (or with information derived from that sample) and making the decision: same or different voice. The task of identification is to examine an unknown sample of speech and then to compare it with speech samples from a (potentially) large number of persons and reach one of two conclusions: either (1) the unknown matches none of the known samples, or (2) the unknown matches one particular known sample. Although the verification and identification applications are different from each other, they share much in technique.

In reviewing the work on voice identification by listening and voice identification by visual examination of voicegrams, one finds that experimenters have generally treated them as independent tasks. The only method used widely as a forensic method of identification, however, relies on both aural and visual methods. The extent to which these two kinds of information are independent remains to be determined.

One of the earliest reported experiments in voice identification[3] was motivated by practical questions raised during a criminal proceeding. Although the forensic application of voice identification did not motivate research again for about 23 years, many voice identification experiments have been reported since the first.

In addition to aural and visual methods, automated voice recognition[4] systems have been investigated for about

17 years, for various reasons, such as: (1) to achieve some performance level for minimum cost; (2) to utilize what are believed to be totally algorithmic methods; and (3) to utilize pattern classification techniques that attempt separation and classification without an underlying model of the process by which the patterns are produced.[5]

The results of these experiments have enhanced understanding of the perceptual basis of voice identification and have increased knowledge about the acoustical characteristics of speaker identity. This work has also increased awareness of the enormous complexity of the speech signal.

Sources of Variability

Discriminating among persons by voice involves distinguishing between the inherent differences in one person's voice and the inherent differences between voices of different persons: i.e., intraspeaker variability and interspeaker variability. Implicit or explicit assumptions about these variabilities are being made when two samples are compared by a voicegram examiner, who decides which similarities and differences are significant and which are not.

Sources of variability range from factors intrinsic to the speaker, such as the inherent variability found among speaker's repetitions of the same text and the effects of psychological stress, to factors extrinsic to the speaker, such as room acoustics and distortions in recordings. These factors bear on judgments made about identity based on speech samples.

Not all the sources of variability mentioned above are routinely present in forensic situations involving voice identification. The influence of any of these factors is itself inconsistent, making evaluation of the precise contribution of any factor in a particular situation difficult, and compensation more so.

The Speaker Psychological stress may affect the acoustical characteristics of speech. The emotional state of a person is likely to be different at the time of making an incriminating communication and at the time of recording an exemplar, although some form of stress may be present in both instances. Other emotional or mental states, such as depression, may also alter voice characteristics. The way in which an individual is affected by emotion, and the exact acoustical effects on speech are not necessarily the same from speaker to speaker.[6] A psychogenic voice disorder is a more remote example of a psychological factor

that can complicate a problem of forensic voice identification,[7] for example, a person who apparently can speak only in a hoarse whisper although no physical cause is present. Even a cooperative suspect may be unable to provide an exemplar of a requested type (for example, of a "natural" production).

Speakers adjust speaking characteristics in relation to the surrounding noise environment, for example, one speaks loudly to overcome a noisy phone connection. Intrinsic sources of variability that are present at the time an unknown sample is recorded present more serious problems than those present when a known sample is recorded, since in the latter case the surrounding circumstances are known and the person can be observed.

The recording of an exemplar is inherently a formal situation, in which a person is likely to be self-conscious about his speech. To produce a voice exemplar, words are either read or repeated after a spoken example in order to ensure the linguistic similarity of the exemplar and unknown sample. The unknown sample may have been spoken in an informal style with articulatory characteristics not found in the exemplar. The attempt to obtain the same spoken words as the unknown sample may introduce stylistic differences from the person's spontaneous speech.

Research comparing performance of voicegram examiners with contemporary and noncontemporary samples has shown that identification error rates are higher for noncontemporary samples. As the elapsed time between recordings increases, so does the opportunity for any intrinsic source of variation to affect the identification.

The time at which a recording is made in relation to the speaker's cycle of sleep and wakefulness can introduce variability between utterances. General fatigue and vocal fatigue from recent lengthy talking both can affect speech characteristics. These factors have not been formally studied, however, to determine their acoustical effects.

Endocrine cyclic factors are another possible source of variability in speech.[8] For example, the pitch of the voice may be altered slightly. Variability in laryngeal source and vocal resonance due to temporal factors and health (discussed below) are probably much more common than articulatory modifications.

The elapsed time between recorded speech samples may be years, in which case the effects of aging may be a source of speech variability. The pronunciation of certain words and the use of grammar may be speaker-distinctive, but they are subject to change with linguistic exposure and

education, which of course may have taken place over a long span of time.

Persons do not produce exact acoustical replications upon repetition of the same utterance; in fact, an intentional attempt to do so would itself constitute an unnatural speaking situation. In a complex physiological process such as speech production, considerable inherent variability is unavoidable. Voicegrams of a speaker's repetitions of the same words will show more variation among themselves than will duplicated fingerprints. A spectral "match" of voices is a judgment of a gestalt similarity; a perfect match, even for isolated features, is not likely.

Allergies and respiratory infections are common sources of variability in speech characteristics. Some of the prominent features of speech are manifestations of the resonances of the vocal tract, including nasal cavity coupling; upper respiratory infections are thus capable of directly affecting these characteristics of speech. Furthermore, features of speech that have been used with some success in discriminating voices, such as nasal consonants, cannot be considered invariant when such alterations in the state of health are present.

Other aspects of health may also increase the variability of speech characteristics, in ways that are currently unknown. Speakers do not merely permit speech to be disturbed by physical factors; they may actively (albeit unconsciously) try to compensate to sound more natural and intelligible. In this way, not only the state of health but also the compensatory attempts themselves become sources of variability. Medication, intoxicants, and drugs have variable effects on speech that may depend on the degree of intoxication, addiction, or withdrawal symptoms.

Dental work can induce temporary or long-term speech alteration, for example, oral anesthesia, bridgework, or dentures.[9] Speech compensations may or may not restore the original speech characteristics.[10]

Dialectical differences among speakers are auditorily striking indices of individual speech characteristics. Semantic, syntactic, and segmental and suprasegmental aspects of speech are all involved in dialectical variation. These aspects of speech are all also superficially susceptible to mimicry and disguise. Some persons are multilingual and/or multidialectical, characteristics that imply an ability to shift from one language or dialect to another at will.

The uninitiated listener may give undue weight to dialectical similarities, tending toward the notion that

speakers of the same dialect "all sound alike." At present it is not known whether certain dialects make their speakers appear more homogeneous spectrographically, thus masking individual identity. An examiner who is unfamiliar with a particular language or dialect may not be easily able to distinguish personal from dialectical characteristics of speech. Dialectical similarities are a special example of a pervasive source of spectral similarity that complicates the voice identification task.

It is important to remember that two voicegrams are bound to contain substantial gross similarities if the same words were spoken. A well-trained examiner is aware that this source of similarity is present, although judges and juries may be overly impressed with the text-related rather than the speaker-related details of voicegrams. Ultimately, a voicegram examiner cannot separate linguistic (including dialectical) features completely from speaker-dependent features. This gray area of uncertainty is fundamental in the voice identification task.

Not all the information relevant to voice identification in a recording is contained in the message-bearing portions. Individual mannerisms that are audible may appear to point rather convincingly to identity; they are most likely to be identified by listening rather than by looking at voicegrams. Examples of such idiosyncratic detail are habitual coughing and throat clearing, type of laughter, degree and type of fluency and dysfluencies, oral clicking and gulping, voice breaks, duration of phrasal groupings and breathing patterns, habitual mispronunciations, and speech defects. The correspondence of such mannerisms in a pair of voice samples may be striking and may be given considerable weight by an examiner since speakers are not usually conscious of such mannerisms. However, such mannerisms may also be products of psychological stress or susceptible to mimicry.

In the broadest sense, vocal disguise can be attributed to any of the intrinsic sources of variability already discussed. Vocal disguise may be inadvertent, in which case "disguise" becomes essentially synonymous with "spuriously dissimilar." Intentional vocal disguises may be present in a speech sample, such as whispering or speaking in falsetto. Research on voice identification that included disguised voices indicates that the task of identification is substantially more difficult with this added variable.

The perceptions of the person recording the exemplar are a subtle source of variability introduced in a voice identification task. If the unknown sample was judged to

be spoken with an attempt at vocal disguise (or if the suspect is perceived as attempting to disguise his voice for the exemplar recording), the suspect may be "coached" to provide an "acceptable" exemplar similar to the criminal sample. Since not all vocal disguises may be so obvious as whispering, the perception that the factor of disguise is present and the degree of influence exerted to obtain a specific type of exemplar may not be recognized or adequately communicated to the examiner.

The Message Path The message path from the speaker to the listener (or the recorder) usually adds noise to the signal and often produces distortions of the signal. The noises and distortions modify the speech signal and therefore modify the sound of the speech and the appearance of the voicegram.

The message path for most voice identification signals consists of the speaker, the surrounding acoustic environment, the telephone microphone, the telephone line (sometimes including sophisticated long-distance signal-processing equipment), the receiving telephone and its associated electromagnetic or other telephone pickup, the tape recorder, the tape, the tape reproducer, and the analysis system (the sound spectrograph). Any one of these transmission elements can add noise to the signal and, in the worst circumstances, completely obscure the voicegram. Furthermore, any one of these transmission elements can change the spectrum shape, and can, in the worst circumstances, almost completely eliminate the frequencies necessary for analyzing a voicegram.

Some of the sources of variability from the receiving telephone are under the control of a law enforcement agency, so the variability can be eliminated by a suitable choice of equipment, operating procedures, and maintenance procedures. Other sources of variability are not under control; those performing the task of voice identification should be aware of their existence.

Sometimes the voicegram examiner can partially correct distortions by an empirical process of filtering and equalizing the signal based on listening to the signal. But the correction process itself is a source of variability and is open to the criticism that correction may artificially increase the similarity between the voicegrams for the unknown sample and the exemplar.

The quality of the analysis of a speech signal is greatly influenced by the difference between the speech signal level and the interfering noise level. Neither the absolute

speech signal level nor the absolute interfering noise level is of itself particularly important. Thus, the very same noise level on a telephone line may produce a poor signal-to-noise ratio[11] for a quiet speaker, talking away from the telephone microphone, and a very good speech-to-noise ratio for a loud speaker, talking close to the microphone. Each element in the message path presents a new opportunity for reducing the speech signal level or increasing the noise level or both. Once a speech signal is thus degraded, it is very difficult to recover what was obscured. The speaker and the acoustical environment are a major source of variability of the signal-to-noise level of the speech signal. The signal level is controlled by the loudness of the voice and by the distance from the speaker's mouth to the telephone microphone. The noise level is controlled by the loudness of the noises at the telephone microphone. Common sources of acoustic noise indoors are other persons' talking, noisy household appliances, and the reverberation of the speaker's voice, which may become severe in an echoey room. Common sources of acoustic noise outdoors include motor vehicles, construction equipment, etc. The acoustic noise level ranges from inaudibility to a level greater than the signal level.

The telephone line is another major source of variability of the signal-to-noise level. The signal level is controlled by the length of the telephone line, which may range from an interoffice line with a length of tens of meters, having no appreciable signal attenuation, to a local line with a length of tens of kilometers, having substantial attenuation. Long-distance circuits and some longer "local" circuits are amplified; their range of attenuation is similar to the range of local lines. In any given location the line distance between the transmitting and the receiving telephones is variable and depends on the availability of alternative telephone lines for use by the automatic switching equipment. This situation is especially applicable for longer "local" calls that are placed between different telephone exchanges.[12]

Typical electrical noise sources on the telephone line are hums, clicks and pops, dialing noises, cross-talk with other telephone lines, and background hiss. Line noise ranges from inaudibility to a level greater than the signal level.

The Output System A recording of a telephone call that is a candidate for the voice identification procedure will typically be made in one of two ways: on a logging

recorder that routinely records all incoming calls (to a police or fire station, for example) or through a pickup that is attached to a telephone on which an incriminating call is expected. Two types of telephone pickup coils are typically used: the "donut" type and the "suction cup" type. Neither type, if properly installed, will degrade the signal-to-noise level. Improper installation of the "suction cup" pickup coil, however, can pick up powerline hum from the transformer.

The tape recorder, the tape, and the tape reproducer, when properly maintained and operated, will not degrade the signal-to-noise level of the signal at the telephone pickup coil. Logging recorders, because of their typically lower quality, related magnetic head geometry, and narrow tracks must be especially carefully maintained to minimize degradation of the signal level and frequency response. When telephone conversations are recorded without technical supervision of the tape recorder, a recorder with "automatic level control" should be used to prevent loss of signal level from underrecording or the introduction of distortion from overrecording.

Reduced level of the recorded signal--usually accompanied by a loss in signal-to-noise ratio--also can result from improper maintenance, low-quality tape, or improper reproducer gain control setting. The tape recorder, tape, and reproducer system can introduce noise if the system is improperly maintained or if low-quality tape is used. Other noise sources are cross-talk from other tape channels, incompletely erased tape, and tape squeal from low-quality tape.

The sources of variability described above can result in increased error rates in the voice identification task. Differences between spectrographic patterns being compared may be attributed erroneously to individual differences rather than to variability introduced by one of the sources just described. Conversely, these differences may be discounted on the basis of factors that cause variability and thereby may not be viewed as evidence against a match.

DECISION THEORY USED IN VOICE IDENTIFICATION

From earlier developments in statistics and decision theory[13] has evolved a well-established method, called the Receiver Operating Characteristic (ROC), for measuring the performance in decision tasks involving physical instruments and human observers. The ROC method provides a way

to exhibit and analyze separately the objective and subjective processes involved in a decision.

In applying the ROC process to voice identification, the objective component includes the system producing the underlying information, the measuring instrument(s) and any other physical technology involved in producing the aural or visual representation(s) of the voice sample(s), and the demonstrable skill of the observer. When the task of an observer is to decide whether two samples represent the same voice or different voices, the subjective component of the decision process includes the examiner's estimate of the prior probability of a match as well as the examiner's estimate of the "payoff" matrix--that is, the relative costs and benefits of the correct and incorrect decisions that might be made.

Decision Criteria and Identification

Any identification process is fundamentally a decision problem. In considering the statistical nature of the problem, we realize that the decision criterion is an important determinant of decision outcomes and that it is especially important with human examiners.

We consider first the simple binary task, whether a given unknown sample matches a particular known sample, a task that we call the "simple discrimination task."

In this simple discrimination task, we have a known voice sample and a similar sample that is either the same or a different voice. The decision task can be represented by a matrix representing the possible states and the possible decisions that the examiner can make. Figure B-3 shows the states as vertical columns of the matrix and indicates the two possible states by m, $\bar{m}$: the two voices in fact either match or do not match. The two possible decisions are the horizontal rows of the matrix and indicate by M, $\bar{M}$ the two possible decisions: match or no match.

The entries of the matrix are the conditional probabilities of each decision given the two possible states. There are two correct decisions: responding "match" when a match does exist--a correct identification--and responding "no match" when a match does not exist--a correct rejection, or true elimination. There are similarly two errors.[14] These errors have various names; usually we call the decision "match" when a match does not exist an incorrect identification, and we call the decision "no match" when a match does exist an incorrect rejection or false

Decision \ State	m = sample does match	$\bar{m}$ = sample does not match
M = sample does match	$P(M\|m)$ correct identification	$P(M\|\bar{m})$ incorrect identification
$\bar{M}$ = sample does not match	$P(\bar{M}\|m)$ incorrect rejection	$P(\bar{M}\|\bar{m})$ correct rejection

FIGURE B-3 Decision matrix of a simple discrimination task.

elimination. One should note that the quantities of the matrix satisfy two linear equations, one being

$$P(M|m) + P(\bar{M}|m) = 1 \qquad \text{Equation 1a}$$

This equation simply affirms that if the true state is m, then some decision is required, either M or $\bar{M}$. The second equation,

$$P(M|\bar{m}) + P(\bar{M}|\bar{m}) = 1 \qquad \text{Equation 1b}$$

likewise affirms that if the true state is $\bar{m}$, then some decision is required, either M or $\bar{M}$. The fact that the matrix contains two (not four) independent probabilities is useful in a manner to be described below. Specifically, we shall focus on correct and incorrect identifications, $P(M|m)$ and $P(M|\bar{m})$ in our analysis.

In general, an examiner wants to make as few errors as possible and as many correct decisions as possible. Yet even this goal is an oversimplification, since some errors are more serious or costly than others. Similarly, one correct response may be more important than another. In addition, Equations 1a and 1b remind us that correct and incorrect responses are related and changing one may affect the others.

The two conditional probabilities $P(M|m)$ and $P(M|\bar{m})$ are important in evaluating how well the voicegram examiner can

perform. A judge or jury is primarily concerned with a related probability, that is, the conditional probability of a match given that the examiner has said there is a match $P(m|M)$. This quantity can be calculated: using what is often called Bayes's theorem, it is

$$P(m|M) = \frac{P(m)\ P(M|m)}{P(M)} \qquad \text{Equation 2}$$

where $P(m)$ is the probability that a match exists (the so-called *a priori* probability of a match) and $P(M)$ is the probability of an examiner's deciding in favor of a match. If one has sufficient data one can estimate $P(M)$: it is simply the total number of matches called by the examiner divided by the total number of attempts. Examiners will presumably differ in the values of $P(M)$, some being more conservative than others and hence having lower values of $P(M)$. One may argue that $P(M)$ cannot be estimated very accurately for a particular examiner or that $P(M)$ depends on unknown factors--e.g., the examiner may tend to be very cautious in certain cases. Similarly, there will be a wide range of opinion on the correct estimate of $P(m)$--with the defense and the prosecution often having widely disparate opinions.

Note, however, for fixed values of $P(m)$ and $P(M)$, increasing $P(M|m)$ does increase $P(m|M)$.

The ROC Curve

In the case of voice identification, we do not yet have a set of elemental features to determine identification as comparison of the ridges, delta patterns, and bifurcations of fingerprints determines identification. Rather, a number of different considerations influence the decision and, in many cases, we cannot point to the particular feature that determines the final decision. We may, however, think of each examination as resulting in the determination of the likelihood that a match does exist. We know that this is a statistical process and we expect the estimated likelihood of a match will tend, on the average, to be higher when a match does exist than when one does not exist.[15]

Figure B-4 is a hypothetical attempt to represent this situation by assuming that the different factors that influence a match are roughly normally distributed with equal variance, given either true state (match or no match), and that the estimated likelihood of a match on the average is greater when the sample is in fact a match than when it is not.

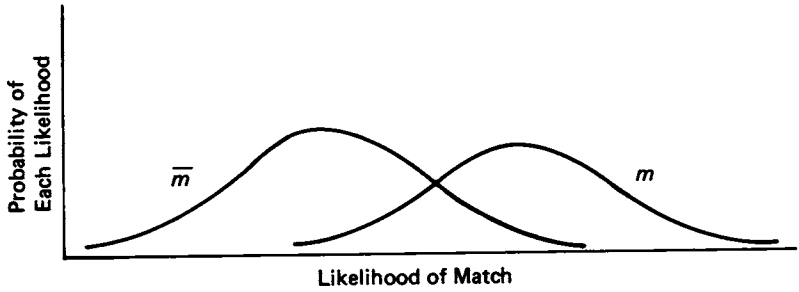


FIGURE B-4 Probability density associated with a match (m) and a nonmatch ($\bar{m}$).

The voicegram examiner must select some criterion value along the likelihood-of-match continuum and say that a match exists whenever a particular likelihood exceeds this criterion. Of course the examiner may choose to report "no decision," a choice that amounts to reporting that the criterion for a match is not exceeded in that particular observation. As one varies the criterion value, one will influence the various probabilities represented in the matrix of Figure B-3. If one varies this criterion systematically over all values, one can trace a curve like the solid curve shown in Figure B-5, in which the two key probabilities we have identified serve as the coordinates of the unit square. Note that, for any fixed situation, one can always achieve a higher probability of calling matches when they do exist--but only by increasing the incorrect identification rate.[16]

Given a certain statistical separation between the distributions of Figure B-4, one could operate at a number of different combinations of the basic probabilities, as in Figure B-5. The general form of the curve in Figure B-5 is commonly called the Receiver Operating Characteristic (ROC) curve.[17]

The importance of the ROC curve is that it provides an analytic distinction between two different but important aspects of the identification problem: (1) the discriminating capacity of the combination of evidence and examiner, and (2) the decision criterion.

The Discriminating Capacity The first aspect of the decision problem is the separation between the two statistical distributions. A better signal-to-noise ratio of the recording, better bandwidth, and even better training or

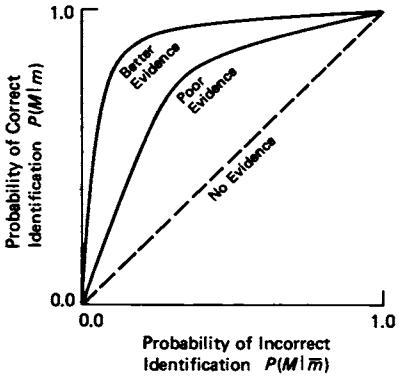


FIGURE B-5 The receiver operating characteristic (ROC) curve: possible performance of a decision maker given evidence of different qualities.

skill on the part of the examiner will lead to a higher probability of correct identification given any probability of incorrect identification. In short, an ROC curve such as that represented by the broken line in Figure B-5 will result if the discriminating capacity of the combination of evidence and examiner is improved. Note that the improvement may result from better evidence (better signal-to-noise ratio) or better ability to interpret evidence. Improvements of either kind will lead to better identification performance. At the other extreme, if the two distributions overlap completely and identically, no evidence can be obtained and the ROC curve becomes the dotted line of Figure B-5. Figure B-5 shows that changes in evidence may alter the total area under the curve. With better discriminating capacity for every probability of incorrect identification, the probability of a correct identification is increased. This improvement in the discriminating capacity should be sharply contrasted with movement along a single curve, as determined by the decision criterion. The capacity to make discriminations is completely separate from the particular decision criterion of the examiner.

The Decision Criterion The second aspect of the decision problem is the decision criterion. Selection of a particular likelihood of match as the cutoff for a positive response (see Figure B-4) is tantamount to the selection from the ROC curve of a particular pair of probabilities for correct and incorrect identifications. If one can increase the capacity of the process to discriminate, then one can increase the probability of a correct identification $P(M|m)$ for any value of incorrect identification $P(M|\bar{m})$. If the

capacity for discrimination is fixed, then one can increase $P(M|m)$ only if one is willing to tolerate a higher probability of incorrect identification, $P(M|\overline{m})$. Similarly, one can always reduce the probability of an incorrect identification if one is willing to tolerate a lower probability of calling matches when they do exist. Changing the decision criterion does not necessarily make for better decisions--only for different ones. However, the decision associated with one decision criterion may be better than that associated with another, as we shall see shortly, depending on the *a priori* probabilities of a match and on the values and costs of the various decision outcomes.

In summary, the discriminating capacity determines the area under the curve and the decision criterion determines the particular point on the curve.

Values and Costs

A major contribution of decision theory analysis is a quantitative account of how the values and costs (negative values) associated with the various decision outcomes should influence the decision criterion. In fact, if one can assign definite values and costs to the various decision outcomes and *a priori* probabilities $P(m)$, then the theory can specify the optimum decision criterion--the one that maximizes the total expected value.

To see how the optimum decision criterion is established, let us begin by defining the values and costs associated with the four possible outcomes of the decision task. Figure B-6 shows these quantities defined in the same format as the stimulus-response matrix of Figure B-3. The quantity V_1 is the value of a correct identification. The quantity C_2 represents the cost (negative value) of an incorrect identification. Similarly, V_2 is the value of a correct rejection, and C_1 is the cost of an incorrect rejection. We would like to be able to assign commensurate values and costs to outcomes--along a single dimension, for example, money--although admittedly, in practice, these quantities are often difficult to specify. If the various decision outcomes can be given definite commensurate values and costs, and one imagines that decisions are made repetitively, always with the same statistical evidence, then one can show that the optimum decision criterion is given by the value of β, as defined in the lower part of Figure B-6.

The criterion is optimum in the sense that it will maximize the overall expected value of a decision. That is,

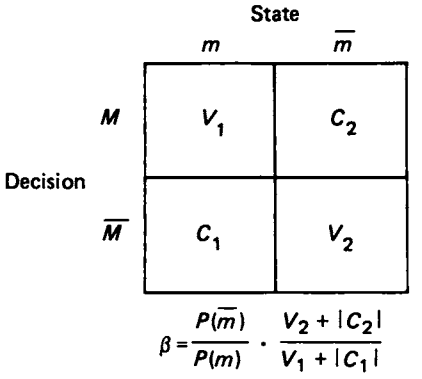


FIGURE B-6 Values and costs of decision matrix (entries are values and costs associated with all possible decision outcomes).

no other decision criterion will, on average, produce a larger yield, given the values and costs assigned to the matrix. If one assumes normal distributions of evidence (as in Figure B-4) the value of β defines a particular likelihood of match, i.e., a particular decision criterion, specifically at the point along the axis of Figure B-4 at which the ratio of the ordinate of the m distribution to the ordinate of the $\bar{m}$ distribution is equal to β. Moreover, β is equal to the slope of the ROC curve at the point yielded by that criterion.

Consider for a moment the slope of the ROC curve. Starting at the lower left the slope is very high, and it diminishes gradually to a value near zero in the upper right-hand corner of the unit square. As one example, the value of β will have a value of one if the probability of a match, $P(m)$, is equal to the probability of no match, $P(\bar{m})$, and the values and costs are all equal. A slope of one corresponds to a point near the middle on an ROC curve. Now suppose that we hold the *a priori* probabilities fixed and vary just one of the values or costs. Specifically suppose that we make the cost of an incorrect identification, C_2, much greater than any other value or cost. Then the value of β becomes very large. This large value will force the probability of an incorrect identification, $P(M|\bar{m})$, to a very low value in order to obtain a very great slope at that point on the ROC curve. Similarly, if V_1, the value associated with a correct identification, is very

high, then β becomes very small and hence the optimum decision criterion is near the upper right on the ROC curve. In this case one accepts a high probability of an incorrect identification in order to achieve a high probability of correct identification. One can see how variations in the other value and cost, or in the *a priori* probabilities, will serve to manipulate the optimum decision criterion.

Alternative Views of the Discrimination and Decision Processes

It is particularly important to distinguish the two aspects of decisions--the capacity to discriminate and the decision criterion--when evaluating some proposal for a change in the procedure of voice identification. Suppose, for example, that some new training technique is proposed, and it is asserted that trainees in this new technique report more matches. In evaluating these results we need to know if the increase in the number of correct identifications, $P(M|m)$, is accompanied by an increase in the probability of an incorrect identification, $P(M|\bar{m})$. If so, depending on the exact quantities involved, the new technique could amount to simply moving a point along the same ROC curve, i.e., without an increase in discrimination capacity. If not, we have a more interesting and useful result--the training technique actually increases the area under the ROC curve, and has accomplished genuine improvement of performance.

This brings us to an area of potential disagreement. The issue concerns whether or not one can increase the probability of a correct identification, $P(M|m)$, without increasing the probability of an incorrect identification, $P(M|\bar{m})$, with an apparently fixed quality of evidence. According to ROC analysis, if the quality of the evidence and the skill of the examiner are fixed, then, by definition, changing the probability of a correct identification can be achieved only by a change in the probability of an incorrect identification. This is true because, according to this analysis, if the quality of the evidence and the skill of the examiner are fixed then only the decision criterion can change, and that change influences both probabilities. The choice moves along the ROC curve that is fixed for that quality of evidence and for a given skill.

A plausible counterargument is to maintain that one can change a different sort of criterion, namely the criterion for accepting a given voicegram as appropriate for

examination--the criterion for examining the evidence at all. Thus, one might impose different standards as to how great the signal-to-noise ratio must be or how much bandwidth must be available or how long the recordings must be, before one is willing to render a decision concerning the identity of the voices. This amounts to prescreening or trying to select only those cases with high quality of evidence and might well have the effect of increasing the probability of a correct identification while leaving the probability of an incorrect identification unchanged. According to ROC analysis, this prescreening would have the effect of changing the quality of the evidence, or the inherent discriminability of the two alternatives. Although there are no available data on this issue, a determination of whether or not prescreening is effective should be comparatively easy. What is needed is an independent evaluation (and, hence, an independent determination of the true state) to ascertain if prescreening changes the area under the ROC curve--a change in the quality of the evidence--or simply alters the position of a point along a single ROC curve.

SCIENTIFIC EXPERIMENTS IN VOICE IDENTIFICATION

Two kinds of experience provide knowledge about the problems inherent in voice identification as well as some indication of possible success. The first is the experience of those who have attempted the task in real-life situations. The second is that of laboratory experimenters who have sought both to make controlled experiments in which the "truth" is known so that the methodology can be verified and to determine more about the nature of the underlying problems. This section presents laboratory experiments (some of which have strong components of real-life situations), their results, and their implications.

Sample Set

A common voice identification task involves presenting a subject with a test voicegram and several labeled reference voicegrams. The subject must decide which, if any, of the reference samples was produced by the speaker who produced the test sample. This task is often referred to as an open, matching-from-sample task. If the set of reference samples is known to include a sample produced by

the speaker who produced the test sample, the task is referred to as a closed, matching-from-sample task. When the reference set includes only one sample in an open, matching-from-sample task, it is called a discrimination or same-different task. The terms *test* and *reference* are often referred to as *unknown* and *known*, respectively. Similarily, the term *subject* is often referred to as *examiner*.

Professional voicegram examiners do not demand that voicegrams from a preset number of known speakers be included with the voicegrams of the unknown speaker. Although examiners are occasionally presented with a reference set (a "lineup") of voicegrams from known speakers, they are more commonly given a recording of only one known speaker to compare with a recording of the unknown speaker. Thus, they are typically performing discrimination tasks.

The nature of the two tasks, the discrimination task and the task of matching from an open or closed reference set, is different and may affect the examiner's methods. In addition, the problems inherent in the two tasks are different and can substantially affect the examiner's *a priori* feelings. In the closed reference set, for example, the examiner tries to find the closest match, not one that meets some perceived criterion of a match.

Size of the Reference Set Smrkovski[18] has reported results for the discrimination task, but he did not investigate the effect of increasing the size of the reference set. Tosi *et al.*[19] have been the only investigators to treat the size of the reference set as a variable in a visual voice identification experiment. Their results show that as the size of the reference set was increased from 10 to 40 speakers, the false identification error rate increased by as much as a factor of two (from 4.9 percent to 9.8 percent) for some conditions.

Homogeneity In general, when subjects are asked to perform a matching-from-sample task, performance is related to the homogeneity of the reference set. The task of identification is more difficult and errors can be expected to be greater when the reference set consists of voices that are very similar to the unknown voice.

Under forensic conditions an examiner probably never will be presented with recordings of known and unknown talkers of different sex or presented with a recording of a child to compare with that of an adult. Nor is it likely that an examiner would be presented with two voices (if

apparently undisguised) that are obviously dissimilar. In effect, an aural prescreening of known (reference) samples is carried out by the persons requesting the comparison. If aural similarity of voices is correlated with spectrographic similarity, as suggested by Stevens *et al.*[20] then the professional examiner is likely to be confronted with pairs of voices of much greater than average aural and spectrographic similarity than would be expected in voices chosen at random from a population matched for only factors such as age and broad dialectal background.

The finding of Tosi regarding the size of the reference set, referred to above, may be the result of increased speaker homogeneity because of prescreening. Tosi described a 5-step procedure that was followed by the examiners in the study. Steps 1 and 2 of their procedure were: (1) to compare the voicegrams of the unknown and known voices by a rather fast scan and (2) to discard those known voicegrams that appeared subjectively to the examiner to contain no significant similarities with the unknown voicegrams. These preliminary steps usually reduced the task to a very few known voicegrams to be examined further.

As the size of the reference set was increased from 10 to 40, it appears that the likelihood of finding better matches for the unknown among the "very few... known voicegrams to be further examined" increased. If the examiners' strategy was to eliminate all but say three or four voicegrams, by performing Steps 1 and 2, then it is reasonable to expect that the homogeneity of the final few voicegrams would be greater when the size of the reference set was 40 than when it was 10.

Tosi described the 250 male speakers used in their study as a "highly homogeneous group." Homogeneous, apparently, because of similarity of age (17-27), current occupation (students at MSU), and linguistic background (native white Americans with no marked dialectal differences or speech defects). Stevens attempted to achieve a high degree of homogeneity in the reference set by selection of 24 speakers from an initial population of 59 on the basis of a prescreening. They used the approximate vocal tract length--a rough measure of the average spacing of formant frequencies--and six aural attributes as the basis for their screening. Although Stevens generally obtained much higher incorrect identification scores than did Tosi, there are no conditions sufficiently common to the tasks investigated by the two studies to infer from their results that differences in the homogeneity of speaker populations was relevant.

Sex of Speaker Most of the reported voicegram experiments have used only male speakers. Some question exists as to whether experimental findings from these studies can be extended to the female population. The higher pitch of female voices often results in a loss of continuity in the spectrographic display of formant trajectories. In general, the frequency positions of spectral prominences for voiced sounds may be less accurately displayed on standard wide-band voicegrams when the speaker's fundamental frequency exceeds 300 Hz. Furthermore, due to a smaller vocal tract configuration, female speakers typically exhibit higher formant frequency positions than do males. In situations in which the speech channel bandwidth is reduced to less than 3 KHz, this effect could cause performance of visual identification to be worse for female than for male speakers.

Both male and female speakers were used in two aural-visual identification experiments reported by Smrkovski. In the first experiment[21] he used three male and three female speakers. In a later experiment (see note 18) he used seven male and seven female speakers. Smrkovski reported that no significant difference was found in the ability of examiners with more than one month of training to identify either male or female speakers. Houlihan,[22] in visual identification experiments, confirms these results. In view of the effects of pitch and vocal tract dimensions on the spectrographic representation of a speaker's voice, however, it may be premature to dismiss sex as an irrelevant variable in aural-visual identification on the basis of experiments that considered so few speakers.

Dialect Virtually all voicegram experiments have used adult male speakers of General American English with no noticeable speech defects. In the forensic situations, suspects (known voices) are generally also adult males, but in a significant number of cases, some form of dialect or accent is involved. No attempt has been made to determine if visual identification performance is sensitive to the sometimes stylized speech patterns of speakers of dialects. Populations of speakers of very similar dialects may be sufficiently homogeneous to affect visual identification performance.

Background and Training of Examiners

For the voicegram experiments reported to date, the amount of preliminary training of naive examiners has ranged from

several hours to several weeks, but training time has not been treated as a variable. At present, most professional examiners are trained and certified as experts in aural-visual voice identification by the International Association of Voice Identification (IAVI). The selection, training, and certification procedures of these examiners are described in detail in Appendix A of this report.

Tosi reported in 1972 (see note 19) identification error rates for each of three groups of examiners: women, male and female students of criminal justice, and male undergraduate students. They found no significant differences in performance among the three groups. Although the effect of examiner background on aural-visual voice identification performance has not been investigated (except for the 1972 Tosi *et al.* study), several experimenters have observed substantial differences in performance among examiners in identical tasks. This finding parallels results reported in aural experiments and may be relevant to the design of selection procedures for professional examiners. For example, certain persons may demonstrate superior abilities at certain subjective tasks as a result of innate talent rather than of extended learning programs.

Smrkovski (see note 18) distinguished among 3 groups of examiners in reporting the performance of 12 examiners associated with IAVI. For a small number of trials (10 discriminations by each of 12 examiners), Smrkovski found that examiner performance as measured by percent correct identifications and eliminations was superior for the more experienced examiners. The least experienced examiners (four novices with about one month of training) made one false identification out of 20 actual non-matches. No false identifications were made by the other examiners (4 members with more than 2 years of training and 4 trainees with less than 2 years of training). For the 4 members and 4 trainees he observed no false eliminations; the 4 novices made a total of 5 false eliminations. These results provide at least weak evidence that the experience an examiner brings to the task improves performance.

A variable that may be significant from the standpoint of examiner training is performance feedback, and no reported study has yet treated this variable. The possibility of criterion drift over a period of time is ever present in the performance of a subjective task. This problem may be especially important in the ongoing training of professional examiners, who might mistakenly interpret corroborative evidence or jury decisions as proof of correct matches.

Speech Material

Context Prior to 1973, a controversy existed over the legal right of prosecutors to require suspects to repeat, for purposes of tape recording, the words of an incriminating message. IAVI examiners thus made aural-visual identification decisions from available speech samples from the known and unknown speakers that had at least 10 words in common. Such speech samples, however, did not have identical phonetic environments.

A number of laboratory experiments have used the phonetic environment as an experimental variable. Experimental results indicate that examiner performance is worst for sample words taken from varying environments (referred to as random context); better for sample words taken from identical environments (referred to as fixed context); and best for words spoken in isolation (a situation that is not relevant to forensic applications).

A change in the legal situation brought about a change in practice: in 1973, the Supreme Court held that there is no constitutional objection to a grand jury subpoena directing a suspect to repeat the words of an incriminating message.[24] Since 1973, then, it has been possible to obtain speech samples for comparison with at least 10 words in common from the same phonetic environment. IAVI examiners now require such samples before they will render a positive identification or elimination. This requirement does not guarantee, however, that every word in the message will be useful for spectrographic comparison. Many factors in addition to phonetic environment may cause variations in the spectrographic representation of the same word by the same speaker from one recording session to another.

Duration IAVI examiners must be able to find similarities among at least 10 words shared by the unknown and the known speech samples before making a response "match." A total duration or total syllable count requirement is not specified. Stevens *et al.* (see note 20) found that visual identification performance was almost twice as good for words of 4-6 syllables as it was for monosyllabic words. Although Hazen[25] and Young and Campbell (see note 23) attribute increased error rates for words in context versus words in isolation in part to durational effects, the effects of the phonetic environment of the sample words probably outweighed those of duration. The 1972 Tosi *et al.* study considered the effect of using 9 monosyllabic words compared to 6 and found no significant difference in

examiner performance. This result may have been affected, however, by learning effects. The 6-word trials were not performed until the examiners had completed 8 months of 9-word trials.

It may be inappropriate with regard to aural-visual identification to consider the duration of a stimulus simply in terms of syllable count or elapsed time. A professional voicegram examiner may be presented lengthy speech samples made by the unknown and known speakers, each speaking the same message. The examiner must then determine what portions of the message should be compared. There are no quantitative guidelines to assist in this determination. He must decide if the effects of the speaker's situation (see below) or the message path have rendered portions or perhaps all of the message useless for purposes of comparison. The requirement of 10 matching words, independent of the total usable speech information, is the only quantitative criterion described by IAVI.

The Speaker When speech is produced during the commission of a crime, it may be affected by the psychological state of the speaker. The speaker may be under the influence of alcohol or some other drug that might affect speech. He may be attempting to disguise his voice and he will probably be speaking spontaneously (i.e., not reading from a text).

A suspect, the known talker, is compelled either to read or to repeat phrases and sentences transcribed from an extemporaneous conversation. The resulting speech is likely to be articulated more precisely and the prosody and rate are likely to be different from those of the speech spoken by the unknown speaker. The physical surroundings, too, with their attendant noise and reverbatory characteristics, are likely to be different from those encountered by the unknown speaker. The noise and spectral characteristics of the message path used by the unknown speaker are likely to be different from those of the message path used by the known speaker. At the time each message is recorded, then, neither the psychological states of the two speakers nor the physical recording situations are likely to be comparable.

In contrast to the forensic situation, speech samples obtained in experimental settings typically involved speakers who have volunteered to be recorded. In obtaining both unknown and known samples, talkers are asked to read from a familiar (though often nonsensical) text.

All recordings are made in relatively unthreatening surroundings--usually the same for both known and unknown samples--and the speaker is probably not sick or under the influence of drugs.

Psychological State Reliable experiments that deal directly with the effect of altered psychological states on aural-visual voice identification performance await the development of a reliable method for determining a person's psychological state or of obtaining a measure of stress. The work of Hecker *et al.*[26] indicates that a speaker's spectral patterns are altered by task-induced stress. Similar findings are reported by Williams and Stevens (see note 6) in an investigation on the effects of emotions on speech. These studies indicate that intra-speaker variability is increased by varying emotional factors.

Disguise Three recently reported investigations concerning voicegram identifications have examined the effect of vocal disguise on performance. Hollien and McGlone[27] had 6 speech scientists attempt to match each of 23 disguised speakers to one of a reference set of 22 normal speakers in a visual identification task. Neither the nature of the disguises nor the context of the speech was reported. All speakers repeated the same utterance. The average error rate was 77 percent, and ranged from 68 percent to 84 percent. Because no performance measure was given for these examiners and speakers under a no-disguise condition, the effect of these disguises on the performance of voicegram examiners is difficult to assess.

Reich et al.[28] examined the effects of 5 selected disguises on the voicegram identification task. They used 4 speech scientists as examiners. These examiners were given 4 weeks of training, including numerous matching trials for which the unknown or test items were samples of disguised speech and for which the reference set was made up entirely of samples of either undisguised or disguised speech. For the experimental trials, examiners attempted to match a test sample of disguised speech to one of 15 undisguised reference samples in an open trial. The examiners also matched undisguised test samples to undisguised references. The average error rate for false identification and false elimination combined, for all examiners combined, and for the condition of undisguised speech, was about 43 percent. The authors did not report the separate values for the two different types of error, an omission that restricts the possibilities for interpreting the data. Although there were marked differences in the degree to which each disguise

affected the error rate, all disguises substantially increased the error rate, compared with the undisguised condition. Houlihan (see note 22) investigated the effects of 4 types of vocal disguise: lowered fundamental frequency, falsetto, whispered speech, and muffled speech. One of the experiments used 8 male and 8 female speakers, closed reference sets for each group of speakers, undisguised voices and the 4 disguises just listed, and 7 moderately trained phonetics students or examiners. Briefly, the results indicate poorer performance matching across disguises and to some extent within disguises.

Endres *et al.*,[29] in the earliest published paper on voice disguise, show experimental results in which persons attempting to disguise their voices shifted the frequencies of their formants by as much as 10 percent or more, sometimes above and sometimes below the frequency of the corresponding formant in their normal, undisguished speech.

Mimicry In this report disguise and mimicry are discussed separately even though they possess certain features in common. In both processes a person is speaking in such a way as to change the voice sounds and to avoid sounding like the person who is speaking. However, in mimicry a person tries to sound as much as possible like some other, particular person, whereas in disguise a person is not necessarily imitating another person and may be speaking in so grossly unnatural a way as to sound like nobody who is speaking normally. Successful disguise can hide only the identity of the person speaking, whereas mimicry can do that and at the same time, if "successful," can falsely identify someone other than the true speaker.

The literature appears to contain only two studies that are concerned explicitly with the relation of mimicry to voice identification. The first is a study conducted by W. Endres and others in Germany and reported in 1971 (see note 29). The other is an unpublished master's thesis submitted in 1975 to Michigan State University by Malcolm E. Hall.[30]

Endres *et al.* studied the spectrographic effects of mimicry by two well-known German mimics who imitated the voices of five speakers. The study shows that the mimics produced imitations in which the formant frequencies were measurably different from those in the mimic's natural voice and at the same time were measurably different from the frequencies of the corresponding formants in the voice of the person being imitated.

For example, the center frequency of the second formant of vowel /a/ spoken by one of the mimics was about 1280 Hz.

The frequency shifted to 1550 Hz when the mimic imitated a certain person, and shifted again to 1480 Hz for the natural voice of the person being imitated. These and other sample data from the Endres study are listed in Table B-1. These numerical data were obtained by measuring the positions of datum points presented graphically in Figure 5 in the paper by Endres *et al.*

In discussing their results, Endres *et al.* say that "... the imitator can change the formant positions of his voice within certain limits." They also state that "... the formant structure in the speech of the person to be imitated and in that of the imitator, in general, do not agree ...." These statements are consistent with the sample data shown in Table B-1 and with the related data given in the Endres *et al.* paper.

Hall, in his master's thesis, reports that he made 69 wide-bandwidth voicegrams of the voice of a professional mimic, including both the mimic's natural voice and his voice when imitating six different persons, and also reports that he made 88 voicegrams of the persons imitated. Hall's conclusions include the following statements: "The analysis of the data collected...suggests that interspeaker variability does exist between a mimicked disguised voice and the natural voice of the subject being mimicked." The analysis also suggests that "... the intraspeaker variabilities are minute and not significant when comparing a mimicked voice with the natural voice of the mimic."

Hall's thesis reports no formal voicegram matching experiments and no objective comparative measurements with regard to the frequencies and structure of the formants. However, for the purpose of evaluating the conclusions given above, we have been able to obtain some relevant quantitative data by making measurements on some of the voicegrams reproduced in the Hall thesis. We have measured the frequencies of the first three formants for a few combinations of vowels spoken and persons imitated. The results are shown in Table B-1 for direct comparison with some of the results of Endres *et al.* An exact comparison is not possible because the two studies do not focus on the same vowels, and because the Hall study does not provide sets of data simultaneously covering the mimic's natural voice, his imitating voice, and the natural voice of the person imitated, but contains only two of these three voice types in each set of data. This lack is indicated in the table by the designation *n.a.* for the data not available.

Even a casual inspection of the data in Table B-1 confirms Hall's conclusion that interspeaker variability

exists between a mimicking voice and the voice of the person being mimicked, and this conclusion agrees also with the results of Endres *et al.* However, the data in Table B-1 do not support Hall's conclusion that the variabilities between a mimic's natural voice and his voice when imitating someone else are "... minute and not significant...." Without discussing the value question as to what is significant, we need only point out that data in the table show more or less similar differences between intraspeaker pairs of data and interspeaker pairs of data.

In Table B-1 the intraspeaker differences between the natural voice and the imitating voice of the mimic range from 20 to 230 Hz in the data from Hall's voicegrams and from 5 to 385 Hz in the data from Endres *et al.* The interspeaker differences between the imitating voice of the mimic and the natural voice of the person imitated range from 10 to 635 Hz in the Hall data and from 45 to 430 Hz in the Endres data. For both the Endres and the Hall data given in the table, the average percentage differences lie between 10 percent and 20 percent for both the intraspeaker and the interspeaker data.

Elapsed Time The elapsed time between the recording of an unknown message and the recording of a known message may, under forensic conditions, be as hosrt as several hours or as long as several years. Tosi *et al.* (see note 19) were the only investigators to use elapsed time between recordings as a variable in a voicegram experiment. Their results indicate that for certain conditions the error rate more than doubled when noncontemporary as opposed to contemporary voicegrams were used. This result occurred for open trials with words in random context and for which the unknown speaker was not represented in the set of known speakers. Independent of other conditions, however, examiner error rates were consistently worse for noncontemporary situations than for contemporary ones. The differences in error rates were especially marked (3 to 9 times as great) for conditions in which words were spoken in isolation or in a fixed context. For words in random context the effect was less (no difference to 2.5 times greater).

Smrkovski (see note 18) reported the results of an experiment in which the samples being compared were recorded 15 months apart. In examining the results obtained by 4 trained examiners, 90 percent of the actual matches were reported as positive identifications; the remaining 10 percent were reported as either probable identifications or no decision. Of the actual non-matches, 80 percent were

TABLE B-1 Formant Measurements of Mimics

RESULTS FROM ENDRES *ET AL.*[a]

Vowel Spoken	Formant	Measured Formant Frequencies in Hertz: Natural Voice of Mimic	Voice of Mimic Imitating Other Person	Natural Voice of Other Person
Person H[b]				
/a:/	First	835	780	665
	Second	1290	1315	1155
	Third	2340	2480	2195
/i/	First	305	345	190
	Second	2285	1965	2010
	Third	2745	2360	2790
Person Li[c]				
/a:/	First	835	985	880
	Second	1290	1535	1470
	Third	2340	2185	2575
/i/	First	305	295	200
	Second	2285	2280	2335
	Third	2745	2690	2890

RESULTS FROM HALL[d]

Vowel Spoken	Formant	Measured Formant Frequencies in Hertz: Natural Voice of Mimic	Voice of Mimic Imitating Other Person	Natural Voice of Other Person
Subject 4[e]				
/e/	First	375	280	n.a.
	Second	1785	1760	n.a.
	Third	2285	2315	n.a.
Subject 5[f]				
/e/	First	375	145	n.a.
	Second	1785	2000	n.a.
	Third	2285	2420	n.a.

TABLE B-1 Continued

/ʌ/[g]	First	n.a.	615	625
	Second	n.a.	1355	1385
	Third	n.a.	2475	2510
Subject 3[h]				
/æ/	First	n.a.	635	560
	Second	n.a.	1160	1795
	Third	n.a.	2000	2405

[a]Endres, W., Bambach, W., and Flosser, G. (1971) Voice spectrograms as a function of age, voice disguise, and voice imitation. *Journal of the Acoustical Society of America* 49:1842-1848.
[b]Endres *et al.* (1971), p. 1845, Figure 5.
[c]Endres *et al.* (1971), p. 1845, Figure 5.
[d]Hall, M. E. (1975) Spectrographic Analysis of Inter-speaker Variabilities of Professional Mimicry. Thesis submitted to Michigan State University.
[e]Hall (1975), p. 20, Figure 6.
[f]Hall (1975), p. 20, Figure 6.
[g]Hall (1975), p. 21, Figure 9.
[h]Hall (1975), p. 25, Figure 18.

reported as positive eliminations; the remaining 20 percent were reported as either probable eliminations or no decision. The speech material was a sentence of 9 monosyllabic cue words recorded by cooperative speakers in their normal working environment.

Quality The noise and bandwidth characteristics of recordings obtained under forensic conditions vary a great deal. Speech-to-noise ratios of less than 5 dB are sometimes encountered. The 1972 Tosi *et al.* study reported no significant difference in performance between examiners using voicegrams made from quiet telephone recordings and those using voicegrams made from noisy telephone recordings. This noise was inserted in the speaker's environment from loudspeakers; although a noise energy level was measured at the speaker's head, the conventional speech-to-noise ratio in the composite signal was not measured. Nor was there a way to compute this ratio from the available information.

Professional examiners use voicegrams with a frequency range of 80-4000 Hz. The useful bandwidth of field

recordings is often less than 3 KHz. The effect of bandwidth limitation on visual identification performance remains unreported. Tosi *et al.* found no significant change from (1) microphone into tape recorder to (2) telephone through inductive pickup into tape recorder. The spectral characteristics of the telephone channel were not reported.

Thus, the effects of noise and spectral characteristics on visual identification performance remain essentially unknown. The frequency locations of the fourth and higher formants are relatively insensitive to message content. For this reason it has been suggested that the loss of spectral information in the range of 3-4 KHz may adversely affect visual identification performance.

Listening-Only Experiments

Many more experiments have been reported on identification by listening alone than on identification involving voicegrams. One of these studies[31] is so much larger than the others that it stands out as does the Tosi *et al.* study. There is, however, wider agreement on the ability of humans to recognize speakers by listening alone than there is on their ability to recognize speakers by looking at voicegrams. The Clarke *et al.* study examined several methods of administering tests of identification by listening and considered a wide range of message path conditions with regard to noise and bandwidth. The results reported by Clarke *et al.* can fairly be said to be representative of the results reported from most recognition-by-listening experiments to date.

Figures B-7, B-8, and B-9, taken from the Clarke *et al.* study, represent a summary of the results most relevant to the forensic application of voice identification by listening. Each figure contains an ROC-curve showing observer (examiner) performance for various experimental conditions. For all the conditions represented by these curves, the observer was asked to listen to 2 utterances and decide whether or not they were spoken by the same talker. He was also asked to indicate one of 3 degrees of confidence (very sure, quite likely, or not certain). These responses were then converted into a 6-point scale of similarity that was used to construct the ROC curves shown in the figures. The solid points in the inserts in the lower right of Figures B-7 and B-8 show the mean percentage of the observers' actual "same" and "different" responses that were correct as a function of experimental condition. The bar through each point shows the range of scores from the highest scoring observer to the lowest

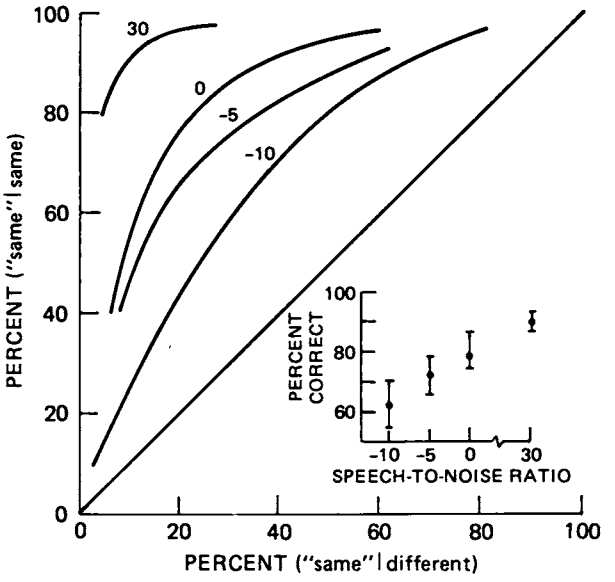


FIGURE B-7 Median ROC curves and mean percentage correct scores as a function of speech-to-noise ratio. Source: F. R. Clarke, R. W. Becker, and J. C. Nixon (1966) *Characteristics that Determine Speaker Recognition.* ESD-TR-66-636. Prepared under contract by Stanford Research Institute, Menlo Park, Cal. Bedford, Mass.: Electronic Systems Division, Air Force Systems Command, U.S. Air Force.

scoring observer. For the best conditions (30 dB speech-to-noise ratio and 90-4500 Hz bandpass), the error scores ranged from 5 percent to 15 percent. As the speech signal was degraded by adding noise or by reducing its bandwidth, the error rates increased, but even at the extreme conditions (-10 dB speech-to-noise ratio and 90-250 Hz bandpass), the median performance was better than chance and the best observers achieved scores of about 70 percent correct.

In Figure B-9, ROC curves are used to illustrate the range of performance encountered by Clarke *et al.* for their 16 listeners in a speaker discrimination test. The large difference in performance between the best and poorest observer has been reported in several other speaker

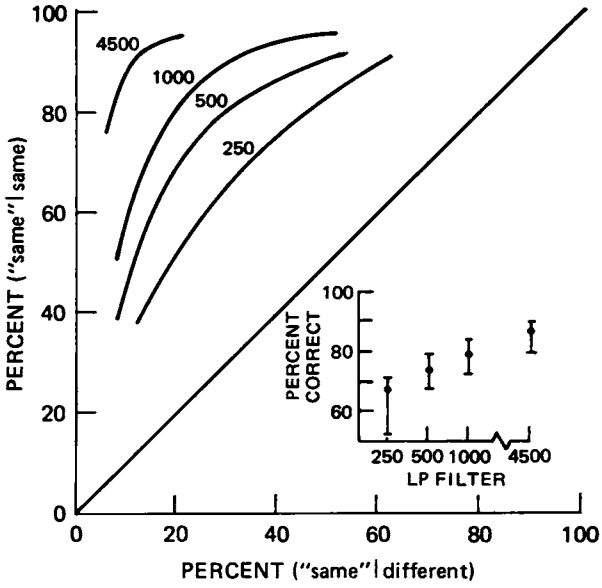


FIGURE B-8 Median ROC curves and mean percentage correct scores as a function of low-pass filtering. Source: F. R. Clarke, R. W. Becker, and J. C. Nixon (1966) *Characteristics that Determine Speaker Recognition.* ESD-TR-66-636. Prepared under contract by Stanford Research Institute, Menlo Park, Cal. Bedford, Mass.: Electronic Systems Division, Air Force Systems Command, U.S. Air Force.

recognition-by-listening studies[32] and, as mentioned earlier, is relevant to the question of professional examiner selection procedures. Also shown on Figure B-9 is the ROC curve achieved by jointly considering all responses to each item by all listeners. That is the multiple observer ROC curve. To obtain that curve, a new similarity index for each stimulus pair was obtained by summing the similarity index of each listener for that pair. Despite the inclusion of even the poorest observer responses in the multiple observer curve, there is a marked improvement in performance for the group over the best individual observer.

For the Clarke *et al.* experiments, all speech samples

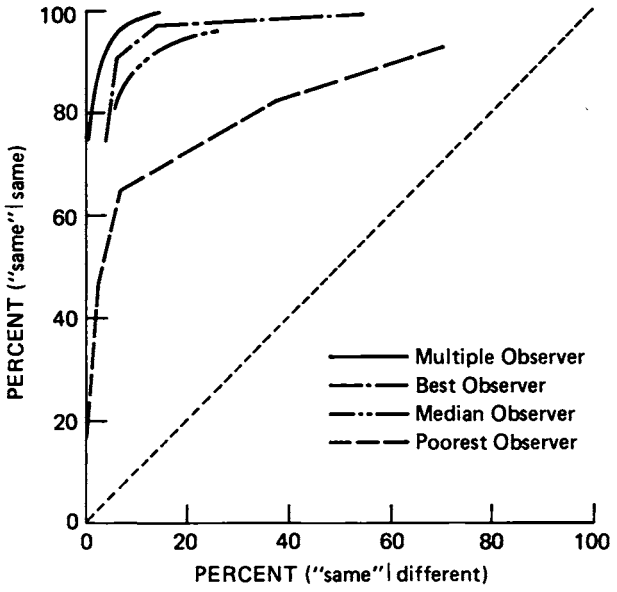


FIGURE B-9 ROC curves for same-different speaker discrimination test at .30 dB speech-to-noise ratio. Source: F. R. Clarke, R. W. Becker, and J. C. Nixon (1966) *Characteristics that Determine Speaker Recognition*. ESD-TR-66-636. Prepared under contract by Stanford Research Institute, Menlo Park, Cal. Bedford, Mass.: Electronic Systems Division, Air Force Systems Command, U.S. Air Force.

from a given speaker were recorded during the same session, but the two speech samples presented to the listeners always consisted of two different 5-syllable sentences. It may be assumed that listener performance would improve if the two speech samples were always of the same phrase. For trials in which signal degradation was present, only the second (i.e., the *B* stimulus of the *A-B* stimulus pair) of the two speech samples was degraded. This last condition is representative of the typical forensic situation in that exemplars from suspects are often recorded under noise and bandwidth conditions different from the conditions present at the time of the unknown recording.

Results reported by Clarke *et al.* indicate that noise and bandwidth conditions do not affect voice identification in the same way that they affect speech intelligibility. Figure B-10a illustrates the comparative effects of noise. The solid curve shows the effect on voice identification, which in this example extends from about 65 to 85 percent correct identification, measured on the left scale, as the average level of the speech sound changes from 10 dB below to 30 dB above the noise level, measured on the abscissa. Over the same range of speech-to-noise levels, the dashed curve shows that the speech word intelligibility extends from almost zero to more than 95 percent on the right scale.

Figure B-10b illustrates the effect of reducing the frequency range. The frequency limit above which the speech sounds are curtailed extends from 250 to 8000 Hz, measured along the abscissa. The solid and dashed curves and their scales have the same meanings as in Figure B-10a.

Figures B-10a and B-10b show that adding noise and reducing the frequency range degrade both the speech message and the identification clues. However, for both kinds of transmission impairment, voice identification is relatively more resistant to degradation than is speech intelligibility. Even when the noise level exceeds the average speech level by 10 dB, or when all frequencies above 250 Hz are cut off, a useful amount of identification information remains, whereas the speech intelligibility is almost eliminated.

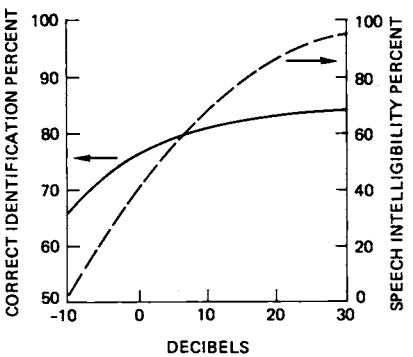


FIGURE B-10a Speech level above noise: decibels.

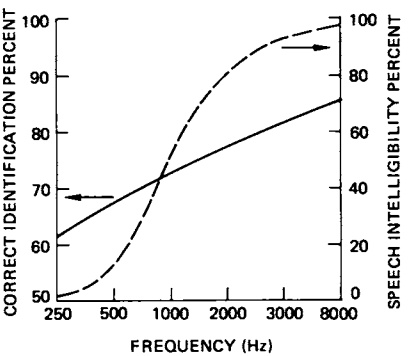


FIGURE B-10b Upper limit of frequency band: hertz.

The results described above indicate that although listening as a method of speaker recognition results in less than perfect performance, it is an extremely robust method with regard to certain degradations. Comparison of the performance scores reported by Clarke *et al.* directly with those reported by Tosi *et al.* is impossible because the two experiments share no common test formats. Comparison of the relative imperviousness to signal degradation of the listening method with that of the spectrographic approach is difficult because no experiment in speaker recognition by visual examination of voicegrams has documented signal degradation conditions in a way that would make such a comparison possible. It would appear, however, that for certain recording conditions (contemporary speech samples, 30 dB speech-to-noise ratio and 90-4000 Hz), the performance measures for observers in the Clarke *et al.* experiment and in the Tosi *et al.* experiment were comparable.

Experiments With and Without Listening

The effect of including listening in the voice identification task, thus transforming it into an aural-visual identification task, has not been adequately investigated. Stevens *et al.* (see note 20) examined the relative effectiveness of aural and visual identification when done separately by subjecting the same examiner to equivalent matching-from-sample tasks using the same speech stimuli for each set of tasks. They reported average error rates of 6 percent for aural and 21 percent for visual identification. The 6 examiners used in this study were randomly selected from 8 who had in turn been chosen from 10 volunteers on the basis of their ability to become familiar with an ensemble of 6 previously unfamiliar voices. They received a minimal amount of training in spectrographic pattern matching techniques. This lack of training makes assessment of the significance of the error scores difficult. It remains to be determined if aural identification performance is superior to visual identification when trained visual identification examiners are used.

Since the aural and visual tasks involve different modalities and mechanisms, it is plausible that an increase in accuracy can be obtained by using both aural and visual methods. The amount of possible gain will of course depend on the amount of independent information that can be provided by the two judgments. It can be argued that the two tasks are just two means of displaying the same information; on the other hand, it is clear that examination of the

visual display is not subject to the time constraints of playing back the aural signal. Although in current practice examiners use a combination of aural and visual methods, there is no current quantitative data about the independence of the information obtained from the two methods. This lack of data clearly suggests a need for research.

The Stevens *et al.* study does, however, provide the only reported performance measure for voice identification by listening for the test format in which the listener is able to hear the test item (the unknown) or any of the comparison items (the knowns) any number of times desired before making a decision. The visual equivalent of this format was used by Tosi *et al.* exclusively in their study. An error rate of about 6 percent was reported by Tosi *et al.* (visual examination of voicegrams only) across all other conditions, for the condition of 10 voices in the reference set. Approximately the same error rate was reported by Stevens *et al.* (listening only) for the condition of 8 voices in the reference set. These results would seem to support the idea that listening and visual examination of voicegrams are comparable as single-mode methods of speaker recognition.

AUTOMATED METHODS OF VOICE IDENTIFICATION

Many of the automated techniques involved in speech processing and the extraction of speech signal descriptors are used for both speaker identification and speaker verification tasks. In the verification task, however, there is no problem in finding which known sample is to be compared against the unknown sample since the purported identity is stated. In the identification task a strategy of some sort is required; the examiner either compares the unknown voice with all samples in the "library," or compares the unknown voice with a smaller number of samples that some classification scheme has selected as being likely to produce matches.

Voice verification normally involves a so-called cooperative speaker, one who is willing to provide a sample of an utterance and to try to make it linguistically the same as a previously recorded utterance. The cooperative speaker will attempt to provide "natural" prosodic and phonetic events in the speech sample. The verification system generally uses the same message path for the samples of speech to be compared.

Voice identification involves either an uncooperative speaker or a noncooperative speaker. An uncooperative speaker is one who knows a sample is being made and deliberately tries to impede the identification by not speaking in a natural or normal manner. A noncooperative or naive speaker is one who either is unaware that a voice sample is being made or is unconcerned with the use to which the sample is being put. Furthermore, since one or both of the samples are being taken without full control of the situation, linguistic, prosodic, phonetic, and message path variability may be present in the samples being compared. Further distinctions can be made in the two tasks as reported in the literature, but most of these distinctions are application-dependent and not the result of some fundamental difference between them. Forensic applications have elements of both identification and verification. In an investigation it would be useful to compare an unknown against a large set of potential suspects, with or without a classification scheme. If the investigation comes to the point of supplying evidence for or against a specific suspect, the problem is much more like that of verification, which involves a closed set.

A considerable amount of the work in automatic speaker recognition has been closely related with ongoing work in automatic speech recognition (recognition of what has been said, rather than who said it). The signal processing techniques have often been the same, although the problems have been quite different. In automatic speech recognition, variability among different speakers is critical, whereas in automatic speaker recognition, variability among different utterances is critical. Both areas involve the effects of intraspeaker variability.

Research on Automated Techniques

Work has been done using a wide variety of automated techniques, examining various characteristics under both text-dependent and text-independent conditions. The text-dependent situation is of course well suited to the speaker verification task. All of this work has involved making measurements of some kind from the speech signals and processing the measurements to arrive at an estimate of the probability that the speech did or did not come from a given speaker. This probability can be used to make decisions regarding speaker identity; the particular decision made is a function of the payoff matrix for the application in question.

Pitch One speech characteristic that has often been used is pitch, under both text-dependent and text-independent conditions.[33] Use of this single variable has been motivated by the fact that pitch variation is known to be a highly speaker-dependent phenomenon and that pitch can be measured under conditions of poor signal-to-noise ratio and a substantial amount of spectral distortion. In the text-dependent situations the statistics stabilize after a few seconds of speech, in the text-independent situations they stabilize after a minute or so of speech. These experiments, conducted on small populations of talkers (around 10), have resulted in correct identification of 96-98 percent in the closed-set (forced-choice) situation. (Note that in the closed-set or forced-choice situation every missed identification is a false identification.)

Spectral Characteristics Work has been done on other spectral characteristics of the signal. Long-term spectral distributions (applied to text-independent material) have been observed to stabilize at 30 to 60 seconds of continuous speech[34] and correlations of spectral characteristics produced about 1 percent overlap between intraspeaker and interspeaker variation for a particular 30-speaker population. Much of the work using spectral measurements has been conducted on specific phoneme samples, generally vowels and nasals. In some cases the establishment of phoneme identity has been accomplished manually and the subsequent analysis has been carried out in a fully automatic way. These manually aided systems have been called "Semi-Automatic Speaker Identification Systems."[35] The results range from perfect identification on a set of 10 talkers recorded 2 weeks apart (Pfeifer) to 1- or 2-percent error rates under forced-choice conditions in other situations. The major study in this area[36] was conducted on a very large speaker population, on a data base intended to represent the investigative and legal situation. Other work[37] indicates that useful speaker discrimination may be possible using phoneme nuclei if the phonemes have not been classified (i.e. text-independent).

A problem with spectral measurements is that they are sensitive to the effects of the message path; that is, the decision algorithms will accept message path features as well as speaker features. Methods have been suggested that should do a good job of removing message path characteristics.[38] These techniques should give good results and reduce the measurements to those of speaker difference only.

Formant Methods Other work has attempted to look at formant frequencies and formant trajectories as a means of speaker identification. This work has been motivated by the fact that the formant trajectory (like the pitch contour) appears to be speaker-dependent and to present an idiosyncratic feature of the speaker. The work in general has depended on the available methods for automatic formant extraction; these are not always well behaved, especially in noisy signals.

LPC Coefficients Still other work has used linear predictor coefficients in efforts to provide complete yet compact descriptions of the speech signal. This work, done on text-dependent material,[39] demonstrated that linear prediction processing, followed by a spectrum computation, permitted very high accuracy and also provided a means of compensating for message path effects. Another experiment indicated that linear prediction processing could extend to text-independent situations at the cost of requiring more data.

Estimates of Physiological Parameters Automated speech processing from other applications is also relevant. For example, Davis[40] tried to determine factors in speech that relate to pathological conditions of the vocal tract. The acoustic parameters that Davis used to describe laryngeal pathologies can also be used as descriptors of individual voice characteristics. These include pitch and amplitude perturbations and other measurements that could be extracted automatically from telephone-quality speech. This work, adapted to normal speech, provides features that until now have played only a very small part in methods of automated speaker identification.

Prospects and Directions

The scientific evidence indicates that the performance of trained examiners engaged in speaker identification is about as good for listening alone as it is for visual examination of voicegrams alone. Furthermore, the large difference in performance between the best and worst listeners and the marked improvement in listening performance that resulted from using a multiple (including good and bad listeners) observer decision criterion (compared to a single observer) are facts that deserve attention in the design of speaker identification methods that use human observers.

Automatic techniques of speaker recognition have much to recommend them for investigative and forensic purposes. They can be made to provide repeatable, objective results. Also, the effects and potential benefits of various alternative designs of the techniques can be evaluated readily once an adequate data base of speech and speaker samples becomes available. Relevant data that are available already are adequate to serve initial needs of a program for exploring new automatic techniques of voice identification.

NOTES

1. For a very readable, complete, and accurate description of the speech signal, see Denes, P., and Pinson, E. (1973) *The Speech Chain.* Anchor S70. Garden City, N.Y.: Anchor Press/Doubleday.
2. Strictly defined, pitch is the subjective impression of fundamental frequency; in practice the two terms are often used interchangeably.
3. McGehee, F. (1937) The reliability of the identification of the human voice. *Journal of General Psychology* 17:249-271.
4. Here the word *recognition* refers to both identification and verification.
5. Atal, B. S. (1976) Automatic recognition of speakers from their voices. *Proceedings of the Institute of Electrical and Electronic Engineers* 64:460-474.
   Rosenberg, A. E. (1976) Automatic speaker verification: a review. *Proceedings of the Institute of Electrical and Electronic Engineers* 64:475-487.
6. Williams, C. E. and Stevens, K. N. (1972) Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America* 52:1238-1250.
7. Aronson, A. E. (1973) *Psychogenic Voice Disorders.* Philadelphia: W. B. Saunders.
8. Brodnitz, S. S. (1971) Hormones and the human voice. *Bulletin of the New York Academy of Medicine* 67:183-191.
   Lucente, F. E. (1973) Endocrine problems in otolaryngology. *Annals of Otolaryngology* 82:131-137.
9. Horii, Y., House, A. S., Li, K-P., and Ringel, R. L. (1973) Acoustic characteristics of speech produced without oral sensation. *Journal of Speech and Hearing Research* 16:67-77.

10. Hamlet, S. L., and Stone, M. (1976) Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses. *Journal of Phonetics* 4:199-218.
Hamlet, S. L., Geoffrey, V. C., and Bartlett, D. M. (1976) Effect of a dental prosthesis on speaker-specific characteristics of voice. *Journal of Speech and Hearing Research* 19:639-650.
11. The term dB (abbreviation for decibel) is a unit that represents the energy level of a signal on a logarithmic scale. A 15 dB speech-to-noise ratio represents about a 30:1 energy ratio. 3 dB represents a 2:1 ratio; 10 dB represents 10:1 ratio; and 20 dB represents a 100:1 ratio. Normal telephone quality is expected to be better than a 30 dB speech-to-noise ratio.
12. Duffy, F. P., and Thatcher, T. W., Jr. (1971) Analog transmission performance on the switched telecommunications network. *Bell System Technical Journal* (April) 50(4):1311-1347.
Gresh, P. A. (1969) Physical and transmission characteristics of customer loop plant. *Bell System Technical Journal* (Dec.) 48(10):3337-3386.
13. A classic paper, which helped to establish the fundamental importance of likelihood ratio, is Neyman, J., and Pearson, E. S. (1933) On the problem of the most efficient type of statistical hypothesis. *Philosophical Transactions of the Royal Society of London* Series A: Mathematical and Physical Sciences. 231: 289.
14. The two errors, called α and β errors in statistical literature, can also be used in an analysis similar to the ROC curve. In statistics one tries to maximize what is called the power of a statistical test.
15. Technically we are using the likelihood of a match as a test statistic. Consider the evidence as a vector, X, in a multidimensional space. We compute the likelihood of that evidence

$$l(X) = \frac{f(X|m)}{f(X|\bar{m})}$$

so that the vector has been replaced by a scalar l for this binary case. Now we can think of l given m or $\bar{m}$. By construction, then, the likelihood of the likelihood is the likelihood

$$l(1) = \frac{f(1|m)}{f(1|\bar{m})} = 1$$

So higher likelihood means that there is a greater probability of a match.

16. The assumption of normality in Figure B-4 dictated the form of the curves in Figure B-5. If other assumptions are made, different curves will result. If likelihood is used as the test statistic, the curves will be monotonically increasing with monotonically decreasing slope. The greater the area under the curve, the better the statistical evidence.
17. Green, D. M., and Swets, J. A. (1966) *Signal Detection Theory and Psychophysics*. New York: John Wiley (Reprinted by Krieger, Huntington, N. Y.).
18. Smrkovski, L. L. (1976) Study of speaker identification by aural and visual examination of non-contemporary speech samples. *Journal of the Association of Official Analytical Chemists* 59:927-931.
19. Tosi, O., Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J., and Nash, E. (1972) Experiment on voice identification. *Journal of the Acoustical Society of America* 51:2030-2043.
20. Stevens, K. N., Williams, C. E., Carbonell, J. R., and Woods, B. (1968) Speaker authentication and identification: a comparison of spectrographic and auditory presentations of speech material. *Journal of the Acoustical Society of America* 44:1596-1607.
21. Smrkovski, L. L. (1975) Collaborative study of speaker identification by the voiceprint method. *Journal of the Association of Official Analytical Chemists* 58:453-456.
22. Houlihan, K. (in press) The effect of disguise on speaker identification from sound spectrograms. Presented at joint session of IPS-77 (International Phonetics Sciences Congress) and the Academy for Forensic Application of the Communications Sciences. To appear in *Current Issues in the Phonetic Sciences*, ed., H. Hollien and P. Hollien. Amsterdam: John Benjamins B.V.
23. Kersta, L. G. (1962) Voiceprint identification. *Nature* 196:1253-1257.
    Young, M. A. and Campbell, R. A. (1967) Effects of context on talker identification. *Journal of the Acoustical Society of America* 42:1250-1254.
24. United States v. Dionisio, 410 U.S. 1 (1973).
25. Hazen, B. (1973) Effects of differing phonetic

contexts on spectrographic speaker identification. *Journal of the Acoustical Society of America* 54:650-660.

26. Hecker, M. H., Stevens, K. N., von Bismarck, G., and Williams, C. E. (1968) Manifestations of task-induced stress in the acoustic speech signal. *Journal of the Acoustical Society of America* 44:993-1001.
27. Hollien, H., and McGlone, R. E. An Evaluation of the 'Voiceprint' Techniques of Speaker Identification. Proceedings of the 1976 Carnahan Crime Counter-Measures Conference. Lexington, Kentucky.
28. Reich, A. R., Moll, K. L., and Curtis, J. F. (1976) Effects of selected vocal disguises upon spectrographic speaker identification. *Journal of the Acoustical Society of America* 60:919-925.
29. Endres, W., Bambach, W., and Flosser, G. (1971) Voice spectrograms as a function of age, voice disguise, and voice imitation. *Journal of the Acoustical Society of America* 49:1842-1848.
30. Hall, M. E. (1975) Spectrographic Analysis of Interspeaker and Intraspeaker Variabilities of Professional Mimicry. Thesis submitted to Michigan State University.
31. Clarke, F. R., Becker, R. W., and Nixon, J. C. (1966) Characteristics that determine speaker recognition. Technical Report ESD-TR-66-636, Bedford, Mass.: Electronic System Division, Air Force Systems Command, U.S. Air Force.
32. Stevens *et al.* 1968 (note 20) and Bricker, P. D., and Pruzansky, S. (1966) Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America* 40:1441-1449.
33. Atal, B. S. (1972) Automatic speaker recognition based on pitch contours. *Journal of the Acoustical Society of America* 52:1687-1697.
Green, N. (1972) Automatic Speaker Recognition Using Pitch Measurements in Conversational Speech. Report 1000. May. Joint Speech Research Unit, Cheltenham, England.
Mead, K. O. (1974) Identification of Speakers from Fundamental-Frequency Contours in Conversational Speech. Report 1002. Joint Speech Research Unit, Cheltenham, England.
34. Li, K.-P., Hughes, G. W., and House, A. S. (1969) Correlation characteristics and dimensionality of speech spectra. *Journal of the Acoustical Society of America* 46:1019-1025.

35. Becker, R. W., Clarke, F. R., Poza, F., and Young, J. R. (1972) A Semiautomatic Speaker Recognition System. Final Report. Prepared for the Michigan Department of State Police by the Aerospace Corp.
Pfeifer, L. L. (1974) Inverse Filter for Speaker Identification. Final Report. Speech Communications Research Laboratory RADC-TR-74-214. Rome, New York: Rome Air Development Center, Air Force Systems Command, Griffiss Air Force Base.
36. Broderick, P. K., Paul, J. E., and Rennick, R. J. (no date) Semi-Automatic Speaker Identification System. Unpublished report prepared by the Aerospace Corp.
37. Kashyap, R. L. (1976) Speaker recognition from an unknown utterance and speaker-speech interaction. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 24(6):481-488.
38. Sambur, M. R. (1976) Speaker recognition using orthogonal linear prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-24(4): 283-289.
Itakura, F. (1975) Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-23(1):67-72.
39. Atal, B. S. (1974) Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *Journal of the Acoustical Society of America* 55:1304-1312.
40. Davis, S. B. (September 1976) Computer Evaluation of Laryngeal Pathology Based on Inverse Filtering of Speech. Monograph No. 13, Santa Barbara, Cal.: Speech Communications Research Laboratory, Inc.

BIBLIOGRAPHY

Several publications provide broad surveys of the topics discussed in Appendix B. These publications are cited and described below.

Stevens, K. N. (1972) Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. Pp. 207-232 in *Proceedings*, Seventh International Congress of Phonetic Sciences (Montreal 1971). The Hague: Mouton.

This paper provides an introduction to the acoustic theory of speech production and describes the various factors that cause variations, with special emphasis on physiological causes of variability.

Hecker, M. L. (1971) *Speaker Recognition: An interpretive Survey of the Literature*. ASHA Monograph No. 16. Washington, D. C.: American Speech and Hearing Association.

This monograph describes the first phase of a study funded by the Law Enforcement Assistance Administration by the Michigan State Police. The second phase resulted in the major study by Tosi et al., which is frequently cited in Appendix B. This monograph provides a tutorial introduction to the acoustic theory of speech production, discusses interspeaker and intraspeaker variability, and reports on the status as of 1970 of experiments in voice identification by listening, voicegram comparison and automatic methods. (Note: This survey preceded the Tosi et al. study.)

Bricker, P. D., and Pruzansky, S. (1976) Speaker recognition. Pp. 295-326 in Norman J. Lass, ed., *Contemporary Issues in Experimental Phonetics*. New York: Academic Press.

This publication concentrates on experiments in listening and on development of a conceptual framework for understanding the results of the experiments. Visual and automatic voice identification experiments are treated more briefly.

Atal, B. S. (1976) Automatic recognition of speakers from their voices. *Proceedings of the Institute of Electrical and Electronic Engineers* 64(4):460-475.

This work provides a survey and methoodology overview of work in automated voice identification.

Rosenberg, A. E. (1976) Automatic speaker verification: a review. *Proceedings of the Institute of Electrical and Electronic Engineers* 64*(4)*:475-487.

This paper is another look at the current state of automated work and techniques with emphasis on verification only.

Abberton, E. (1974) Listener identification of speakers from larynx frequency. Eighth International Congress on Acoustics. London. *Proceedings:* 1:273.

Aerospace Corporation. Speaker Identification Program Final Report. Prepared for the National Institute of Law Enforcement and Criminal Justice, Law Enforcement Assistance Administration, U.S. Dept. of Justice.

Agnello, J. G. Measurements and analysis of visible speech. Pp. 379-397 in Sadanand Singh, ed., *Measurement Procedures in Speech, Hearing, and Language*. Baltimore: University Park Press.

Alpert, M., Kurtzberg, R. L., Pilot, M., and Friedhoff, A. J. (1963) Comparison of the spectra of the voices of twins. Abstract. *Journal of the Acoustical Society of America* 35:1877.

American Bar Association Section of Criminal Justice (1976) *Voiceprint Identification: Admissible Evidence?*. Washington, D.C.: American Bar Association.

Atal, B. S. (1972) Automatic speaker recognition based on pitch contours. *Journal of the Acoustical Society of America* 52:1687-1697.

Atal, B. S. (1974) Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *Journal of the Acoustical Society of America* 55:1304-1312.

Atkinson, J. E. (1976) Inter- and intraspeaker variability in fundamental voice frequency. *Journal of the Acoustical Society of America* 60:440-445.

Becker, M. H., Gnanadesikan, R., Mathews, M. V., Pinkham, R. S., Pruzansky, S., and Wilk, M. B. (1964) Comparison of some statistical distance measures for talker identification. Abstract. *Journal of the Acoustical Society of America* 36:1988.

Becker, R. W. (1965) Decision procedures for discriminating between speakers. Abstract. *Journal of the Acoustical Society of America* 37:121.

Becker, R. W., Clarke, F. R., Poza, F., and Young, J. R. A Semiautomatic Speaker Recognition System: Final Report. Prepared for the Michigan Department of State Police, 1972.

Black, J. W., Lashbrook, W., Nash, E., Oyer, H. J., Pedrey, C., Tosi, O. I., and Truby, H. (1973) Reply to "Speaker identification by speech spectrograms: some further observations." *Journal of the Acoustical Society of America* 54:535-537.

Bolt, R. H., Cooper, F. S., David, E. E., Jr., Denes, P. B., Pickett, J. M., and Stevens, K. N. (1969) Identification of a speaker by speech spectrograms. *Science* 166:338-343.

Bolt, R. H., Cooper, F. S., David, E. E., Jr., Denes, P. B., Pickett, J. M., and Stevens, K. N. (1970) Speaker identification by speech spectrograms: a scientists' view of its reliability for legal purposes. *Journal of the Acoustical Society of America* 47:597-612.

Bolt, R. H., Cooper, F. S., David, E. E., Jr., Denes, P. B., Pickett, J. M., and Stevens, K. N. (1973) Speaker identification by speech spectrograms: some further observations. *Journal of the Acoustical Society of America* 54:531-534.

Bricker, P. D., and Pruzansky, S. (1966) Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America* 40:1441-1449.

Bricker, P. D., and Pruzansky, S. (1965) Temporal effects in talker identification. Abstract. *Journal of the Acoustical Society of America* 38:924.

Bricker, P. D., Gnanadesikan, R., Mathews, M. V., Pruzansky, S., Tukey, P. A., Wachter, K. W., and Warner, J. L. (1971) Statistical techniques for talker identification. *Bell System Technical Journal* 50:1427-1454.

Broderick, P. K., Paul, J. E., and Rennick, R. J. Semi-Automatic Speaker Identification System.

Bunge, E. (1975) Automatic speaker recognition by computers. *Proceedings*; Carnahan Conference on Crime Countermeasures. Lexington, Kentucky.

Bunge, E. (1976) Automatic speaker recognition by computers. Abstract only. 1976 IEEE International Conference on Acoustics, Speech & Signal Processing. Philadelphia, April 12-14, 1976. *Record:* 738.

Bunge, E., Hofker, U., Jesorsky, P., Driener, B., and Wesseling, D. (1977) Statistical techniques for automatic speaker recognition. 1977 IEEE International Conference on Acoustics, Speech & Signal Processing.

Hartford, May 9-11, 1977. *Record:*772-783.

Calinski, T., Jassem, W., and Kaczmarek, Z. (1970) Investigation of vowel formant frequencies as personal voice characteristics by means of multivariate analysis of variance. Pp. 8-39 in Wiktor Jassem, ed., *Speech Analysis and Synthesis.* Vol. 2. Warsaw: Polish Academy of Sciences.

Carbonell, J. R., Grignetti, M. C., Stevens, K. N., Williams, C. E., and Woods, B. (1965) Speaker Authentication Techniques. Report No. 1296. Cambridge, Mass.: Bolt Beranek and Newman, Inc.

Carbonell, J. R., Stevens, K. N., Williams, C. E, and Woods, B. (1965) Speaker identification by a matching-from-sample technique. Abstract. *Journal of the Acoustical Society of America* 37:1205.

Chapman, W. D., and Li, K-P. (1966) Speaker verification. Abstract. *Journal of the Acoustical Society of America* 40:1282.

Clarke, F. R. (1965) Speaker recognition by humans. Abstract. *Journal of the Acoustical Society of America* 37:1211.

Clarke, F. R. A missing step in 'voiceprint' validation. Unpublished memo, 1973.

Clarke, F. R., and Becker, R. W. (1969) Comparison of techniques for discriminating among talkers. *Journal of Speech and Hearing Research* 12:747-761.

Clarke, F. R., Becker, R. W., and Nixon, J. C. (1966) Characteristics that determine speaker recognition. Technical Report ESD-TR-66-636, Bedford, Mass: Electronic Systems Division, Air Force Systems Command, U.S. Air Force.

Coleman, R. O. (1973) Speaker identification in the absence of intersubject differences in glottal source characteristics. *Journal of the Acoustical Society of America* 53:1741-1743.

Collins, A. M. (1973) Computer Speech Processing for Speaker Identification. Publication EP-RR 37. Department of Engineering Physics Research School of Physical Sciences. Canberra: The Australian National University.

Compton, A. J. (1963) Effects of filtering and vocal duration upon the identification of speakers, aurally. *Journal of the Acoustical Society of America* 35:1748-1752.

Cort, S., and Murry, T. (1972) Aural identification of children's voices. Abstract. *Journal of the Acoustical Society of America* 51:131.

Das, S. K., and Mohn, W. S. (1969) Pattern recognition in

speaker verification. In *Proceedings of the Fall Joint Computer Conference*, Las Vegas, Nev. Montvale, N.J.: American Federation of Information Processing Societies.

Das, S. K., and Mohn, W. S. (1971) A scheme for speech processing in automatic speaker verification. *IEEE Transactions on Audio and Electroacoustics* Au-19:32-43.

Das, S. K., Mohns, W. S., and Saleeby, S. L. (1971) Speaker verification experiments. Abstract. *Journal of the Acoustical Society of America* 49:138.

Das, S. K., Mohn, W. S., Jr., Willett, S. S., and Chapman, W. D. (1972) Two speaker verification experiments. 1972 Conference on Speech Communication and Processing. Newton, Mass., April 24-26, 1972. Sponsored by the Institute of Electrical and Electronic Engineers, Inc. and the Air Force Cambridge Research Laboratories. *Conference Record:*275-278.

Davis, S. B. (1976) Computer Evaluation of Laryngeal Pathology Based on Inverse Filtering of Speech. Monograph No. 13. Santa Barbara, California: Speech Communications Research Laboratory, Inc.

Doddington, G. R. A Method of Speaker Verficiation. PH.D. Thesis, University of Wisconsin, 1970.

Doddington, G. R. (1971) A method of speaker verification. Abstract. *Journal of the Acoustical Society of America* 49:139.

Doddington, G. R. (1974) Speaker Verification. Final Report. RADC-TR-74-179. Griffiss Air Force Base, New York: Rome Air Development Center, Air Force Systems Command, U.S. Air Force.

Doddington, G. R. (1974) Speaker Verification. Technical Report RADC-TR-963700-F. Griffiss Air Force Base, New York: Rome Air Development Center, Air Force Systems Command, U.S. Air Force.

Doddington, G. R., Flanagan, J. L., and Lummis, R. C. (1972) Automatic Speaker Verification by Nonlinear Time Alignment of Acoustic Parameters. U.S. Patent 3700815. October 24, 1972. Washington, D.C.: U.S. Commissioner of Patents.

Doddington, G. R., and Hydrick, B. M. (1976) 'Total voice'--speaker verification goes deluxe. Abstract. *Journal of the Acoustical Society of America* 59 (Supplement 1):S96.

Doherty, E. T. (1975) Evaluation of selected acoustic parameters for use in speaker identification. Abstract. *Journal of the Acoustical Society of America* 58:S107.

Dukiewicz, L. (1970) Frequency-band dependence of speaker identification. Pp. 42-50 in Wiktor Jassem, ed., *Speech Analysis and Synthesis* vol. 2. Warsaw: Polish Academy of Sciences.

Edie, J., and Sebestyen, G. S. (1962) Voice Identification General Criteria. RADC-TDR-62-278. Griffiss Air Force Base, New York: Rome Air Development Center, Air Force Systems Command, U.S. Air Force.

Egan, J. P. (1948) Articulation testing methods. *Laryngoscope* 58:955-961.

Egan, J. P., Schulman, A. I., and Greenberg, G. Z. (1959) Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America* 31:768-773.

Endres, W. (1970) Changes of human voice caused by age, disguise, and simulation. Abstract. *Journal of the Acoustical Society of America* 48:131.

Endres, W., Bambach, W., and Flosser, G. (1971) Voice spectrograms as a function of age, voice disguise, and voice imitation. *Journal of the Acoustical Society of America* 49:1842-1848.

Fallside, F. (1976) Speaker identification by multivariable linear prediction analysis. Abstract only. 1976 IEEE International Conference on Acoustics, Speech & Signal Processing. Philadelphia, April 12-14, 1976. *Record:* 739.

Fant, G. (1960) *Acoustic Theory of Speech Production.* 's-Gravenhage: Mouton.

Flanagan, J. L. (1958) Some properties of the glottal sound source. *Journal of the Acoustical Society of America* 1:99-116.

Flanagan, J. L. *Speech Analysis Synthesis and Perception.* New York: Academic Press [second ed., Springer].

Floyd, W. (1964) Voice Identification Techniques. No. RADC-TDR-64-312 Griffiss Air Force Base, N.Y.: Rome Air Development Center, Air Force Systems Command, U.S. Air Force.

Fujimura, O. (1962) Analysis of nasal consonants. *Journal of the Acoustical Society of America* 34:1865-1875.

Furui, S. (1974) An analysis of long-term variation of feature parameters of speech and its application to talker recognition. *Electronics and Communications in Japan* 57(A).

Furui, S., and Itakura, F. (1973) Talker recognition by statistical features of speech sounds. *Electronics and Communications in Japan* 56-A(11):62-71.

Furui, S., Itakura, F., and Saito, S. (1972) Talker Recognition by Longtime Averaged Speech Spectrum. Abstract. *Electronics and Communications in Japan* 55: 54-61.

Garvin, P. L., and Ladefoged, P. (1963) Speaker identification and message identification in speech recognition. *Phonetica* 9:193-199.

Glenn, J. W., and Kleiner, N. (1968) Speaker identification based on nasal phonation. *Journal of the Acoustical Society of America* 43:368-372.

Godfrey, J. J. (1974) Voiceprints: fact and science fiction. University of Dayton Research Institute. *NAECON '74 Record:*23-29.

Goldstein, U. G. (1976) Speaker-identifying features based on formant tracks. *Journal of the Acoustical Society of America* 59:176-182.

Gubrynowica, R. (1973) Application of a statistical spectrum analysis to automatic voice identification. Pp. 171-180 in Wiktor Jassem, ed., *Speech Analysis and Synthesis* 3:171-180. Vol. 3. Warsaw: Polish Academy of Sciences.

Haberman, W., and Fejfar, A. Automatic identification of personnel through speaker and signature verification. Carnahan Conference on Crime Countermeasures. Lexington, Kentucky. *Proceedings:*23-30.

Hair, G. D., and Rekieta, T. W. (1972) Automatic speaker verification using phoneme spectra. Abstract. *Journal of the Acoustical Society of America* 51:131.

Hair, G. D., and Rekieta, T. W. (1973) Techniques for objective speaker identification. Abstract. *Journal of the Acoustical Society of America* 53:354.

Hall, M. C. (1974) The current status of speaker identification by use of speech spectrograms. *Journal of the Canadian Society of Forensic Science* 7(3):153-176.

Hall, M. C., and Collins, A. M. Speech Processing in Australia Using a Flexible Computer-Based System. Presented to the Second International Conference on Special Equipment for the Police, Rome, Italy. September 21-24, 1976.

Hall, M. E. Spectrographic Analysis of Interspeaker and Intraspeaker Variabilities of Professional Mimicry. Thesis submitted to Michigan State University, 1975.

Hall, M., and Tosi, O. (1975) Spectrographic and aural examination of professionally mimicked voices. Abstract. *Journal of the Acoustical Society of America.* 58-S107.

Hargreaves, W. A., and Starkweather, J. A. (1963) Recognition of speaker identity. *Language and Speech* 6:63-67.

Hayre, H. S. (1976) Speech--a possible indicator of physical stress. Abstract only. 1976 IEEE International Conference on Acoustics, Speech & Signal Processing. Philadelphia, April 12-14, 1976. *Record:*740.

Hazen, B. M. The Effects of Changing Phonetic Context in the Voiceprint Identification Technique. Thesis submitted to the State University of New York at Buffalo, 1972.

Hazen, B. M. (1973) Effects of differing phonetic contexts on spectrographic speaker identification. *Journal of the Acoustical Society of America* 54:650-660.

Hecker, M. H. L. (1971) Speaker recognition: basic considerations and methodology. Abstract. *Journal of the Acoustical Society of America* 49:138.

Hecker, M. H. L. Three Forensic Techniques Involving Acoustics and the Admissibility of Evidence Obtained From Their Use. Paper presented at Annual Meeting of American Association for the Advancement of Science, 1976.

Hecker, M. H. L., and Guttman, N. (1967) Survey of methods for measuring speech quality. *Journal of Audio Engineering Society* 15:400-403.

Hecker, M. H. L., and Poza, F. (1974) *Voiceprint Field Study*. Task Report, SRI Project 730532-CAN. Menlo Park, Cal.: Stanford Research Institute.

Hecker, M. H. L., and Williams, C. E. (1965) On interrelations among speech quality, intelligibility, and speaker identifiability. Fifth International Congress on Acoustics. Liege. *Proceedings:* Paper A15.

Hecker, M. H., Stevens, K. N., von Bismarck, G., and Williams, C. E. (1968) Manifestations of task-induced stress in the acoustic speech signal. *Journal of the Acoustical Society of America* 44:993-1001.

Helms, R. E., and Doddington, G. R. (1976) Speaker verification using passwords. *Journal of the Acoustical Society of America* 59(Supplement 1):S96(A).

Hemdal, J. F. (1967) Some results from the normalization of speaker differences in a mechanical vowel recognizer. *Journal of the Acoustical Society of America* 41:1594.

Holden, A. D. C., Cheung, J. Y., and Gulut, Y. K. (1976) The role of idiosyncracies in linguistic stressing cues, and accurate formant analysis, in speaker identification. 1976 Carnahan Conference on Crime Countermeasures. Lexington, Kentucky. *Proceedings:*31-37.

Hollien, H. (1974) Peculiar case of "voiceprints." *Journal of the Acoustical Society of America* 56:210-213.

Hollien, H., and McGlone, R. E. (1976) The effect of disguise on "voiceprint" identification. *Journal of Criminal Defense* 2:117-130.

Hollien, H., and McGlone, R. E. An evaluation of the 'voiceprint' techniques of speaker identification. Carnahan Conference on Crime Countermeasures. Lexington, Kentucky. *Proceedings:*39-45.

Hollien, H., Childers, D. G., and Doherty, E. T. (1977) Semi-Automatic system for speaker identification (SAUSI). 1977 IEEE International Conference on Acoustics, Speech & Signal Processing. Hartford, May 9-11, 1977. *Record:*768-771.

Hollien, H., Majewski, W., and Hollien, P. (1974) Speaker Identification by Long-Term Spectra Under Normal, Stress and Disguise Conditions. Eighth International Congress on Acoustics. London. *Proceedings:*1:269.

Holmgren, G. L. (1963) Speaker Recognition. Report no. AFCRL-63-119. Bedford, Mass: Air Force Cambridge Research Laboratories, Office of Aerospace Research.

Holmgren, G. (1966) Speaker recognition, speech characteristics, speech evaluation, and modification of speech signal--a selected bibliography. *IEEE Transactions on Audio and Electroacoustics* AU-14:32.

Holmgren, G. L. (1967) Physical and psychological correlates of speaker recognition. *Journal of Speech and Hearing Research* 10:57-66.

Hunt, M. J., Yates, J. W., and Bridle, J. S. (1977) Automatic speaker recognition for use over communication channels. Abstract. 1977 IEEE International Conference on Acoustics, Speech and Signal Processing. Hartford, May 9-11, 1977. *Record:*464-467.

Hydrick, B. M., and Doddington, G. R. (1976) Speaker verification--an operational example. *Journal of the Acoustical Society of America* 59(Supplement 1):S96(A).

Ingemann, F. (1968) Identification of the speaker's sex from voiceless fricatives. *Journal of the Acoustical Society of America* 44:1142-1144.

Jakobson, R., Fant, C. G. M., and Halle, M. (1963) *Preliminaries to Speech Analysis.* Cambridge: MIT Press.

Jansen, L. P. C. (1974) Objective speaker identification by means of some relations between vowel parameters. Eighth International Congress on Acoustics. London. *Proceedings:*1:270.

Jassem, W. (1968) Vowel formant frequencies as cues to speaker discrimination. Pp. 10-41 in Wiktor Jassem,ed., *Speech Analysis and Synthesis*, vol. 1, Warsaw: Polish Academy of Sciences.

Jassem, W., Steffen-Batog, M., and Czajka, S. (1973) Statistical characteristics of short-term average F_o distributions as personal voice features. Pp. 209-225 in Wiktor Jassem, ed., *Speech Analysis and Synthesis,* vol. 3. Warsaw: Polish Academy of Sciences.

Jayant, N. S. (1972) Decision-theoretic approach to speaker verification. Abstract. *Journal of the Acoustical Society of America* 51:132.

Kaczmarek, Z., and Krzysko, M. (1973) An attempt to use Anderson and Bahadur's separating hyperplane to identify a population among many normal populations. Pp. 160-169 in Wiktor Jassem, ed., *Speech Analysis and Synthesis,* vol. 3. Warsaw: Polish Academy of Sciences.

Kersta, L. G. (1962) Voiceprint identification. *Nature* 196:1253-1257.

Kersta, L. G. (1962) Voiceprint-identification infallibility. Abstract. *Journal of the Acoustical Society of America* 34:1978.

Kersta, L. G. (1965) Environmental influence on the speech of family members shown by spectrographic speech matching. Abstract. *Journal of the Acoustical Society of America* 38:935.

Kersta, L. G. (1965) Voiceprint classification. Abstract. *Journal of the Acoustical Society of America* 37:1217.

Kersta, L. G. (1966) Voiceprint classification for an extended population. Abstract. *Journal of the Acoustical Society of America* 39:1239.

Kersta, L. G. Automated Talker Identification by Quantized Spectrography. Conference on Speech Communication and Processing, 1967.

Kersta, L. G. (1968) Instruction and application of voiceprint identification to law enforcement. Abstract. *Journal of the Acoustical Society of America* 44:390.

Kersta, L. G. Speaker Identification by Spectrographic Voiceprints of Voiceprint Laboratories. Sixth International Congress on Acoustics. Tokyo. B147-B150: *Proceedings:* Paper B-4-10.

Kersta, L. G. (1971) Progress report on automated speaker-recognition system. Abstract. *Journal of the Acoustical Society of America* 49:139.

Klatt, D. H., and Stevens, K. N. (1973) On the automatic recognition of continuous speech: implications from a spectrogram-reading experiment. *IEEE Transactions on Audio and Electroacoustics* AU-21:210-216.

Kosiel, U. (1973) Statistical analysis of speaker-dependent differences in the long-term average spectrum

of Polish speech. Pp. 181-208 in Wiktor Jassem, ed., *Speech Analysis and Synthesis,* vol. 3. Warsaw: Polish Academy of Sciences.

Krzysko, M., Jassem, W., and Frackowiak-Richter, L. (1973) Statistical discrimination functions and their application to the problem of voice identification. Pp. 144-157 in Wiktor Jassem, ed., *Speech Analysis and Synthesis,* vol. 3. Warsaw: Polish Academy of Sciences.

Kurtzberg, R. L., Alpert, M., and Friedhoff, A. J. (1963) Identification from voice: techniques for the reduction of trial-retrial variability. Abstract. *Journal of the Acoustical Society of America* 35:1877.

Ladefoged, P. An Opinion on "Voiceprints." Working Papers in Phonetics No. 19. UCLA, 1971.

Ladefoged, P., and Papcun, G. Two "Voiceprint" Cases. Working Papers in *Phonetics* No. 31. UCLA, 1976.

Ladefoged, P., and Vanderslice, R. (1967) The voiceprint mystique. Abstract. *Journal of the Acoustical Society of America* 42:1164.

Lass, N. J., and Davis, M. (1976) An investigation of speaker height and weight identification. *Journal of the Acoustical Society of America* 60:700-703.

Lass, N. J., and Harvey, L. A. (1976) An investigation of speaker photograph identification. *Journal of the Acoustical Society of America* 59:1232-1236.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (1976) Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America* 59:675-678.

Li, K-P., and Hughes, G. W. (1974) Talker differences as they appear in correlation matrices of continuous speech spectra. *Journal of the Acoustical Society of America* 55:833-837.

Li, K-P., Dammann, J. E., and Chapman, W. D. (1966) Experimental studies in speaker verification, using an adaptive system. *Journal of the Acoustical Society of America* 40:966-978.

Li, K-P., Hughes, G. W., and House, A. S. (1969) Correlation characteristics and dimensionality of speech spectra. *Journal of the Acoustical Society of America* 46:1019-1025.

Li, K-P., Hughes, G. W., and House, A. S. (1970) Approaches to the characterization of talker differences by statistical operations on speech spectra. Abstract. *Journal of the Acoustical Society of America.* 47:66.

Lin, W. C., and Pillay, S. K. (1976) Feature evaluation and sèlection for an on-line adaptive speaker verification system. 1976 IEEE International Conference on Acoustics, Speech, and Signal Processing. Philadelphia, April 12-14, 1976. *Record:* 734-737.

Luck, J. E. (1969) Automatic speaker verification using cepstral measurements. *Journal of the Acoustical Society of America* 46:1026-1032.

Luck, J. E. (1969) Discription of a completely automatic speaker-verification system. Abstract. *Journal of the Acoustical Society of America* 46:90.

Lummis, R. C. (1971) Real-time technique for speaker verification by computer. Abstract. *Journal of the Acoustical Society of America* 50:106.

Lummis, R. C. (1972) Implementation of an on-line speaker verification scheme. Abstract. *Journal of the Acoustical Society of America* 52:181.

Lummis, R. C. (1972) Speaker verification: a step toward the 'checkless' society. *Bell Laboratories Record* 50:254-259.

Lummis, R. C. (1973) Speaker verification by computer using speech intensity for temporal registration. *IEEE Transactions on Audio and Electroacoustics AU-21:* 80-89.

Lummis, R. C., and Rosenberg, A. E. (1972) Test of an automatic speaker verification method with intensively trained professional mimics. Abstract. *Journal of the Acoustical Society of America* 51:131.

Majewski, W., and Hollien, H. (1974) Euclidean distance between long-term spectra as a criterion for speaker identification. Speech Communications Seminar. Stockholm. *Proceedings:* 3:303-310.

Majewski, W., and Hollien, H. (1974) Speaker Identification by Means of Long-Term Speech Spectra. Eighth International Congress on Acoustics. London. *Proceedings:* 1:274.

Makhoul, J. (1972) Computer-assisted reading of spectrograms. Abstract. *Journal of the Acoustical Society of America* 51:131.

Markel, J. D., Oshika, B. T., and Gray, A. H., Jr. (1977) Long-term feature averaging for speaker identification. *IEEE Transactions on Acoustics, Speech and Signal Processing* ASSP-25(4):330-337.

Matsumoto, H., Hiki, S., Sone, T., and Nimura, T. Multidimensional representation of personal quality of vowels

and its acoustical correlates. 1978 IEEE International Conference on Acoustics, Speech & Signal Processing. Tulsa, April 10-12, 1978. *Record:*384-386.

McGee, V. E. (1965) Invariance of personal characteristics of voice over two vowel sounds. *Perceptual Motor Skills* 21:519-529.

McGehee, F. (1937) The reliability of the identification of the human voice. *Journal of General Psychology* 17:249-271.

McGehee, F. (1944) An experimental study in voice recognition. *Journal of General Psychology* 31:53-65.

McGlone, R. E., and Hollien, H. Partial analysis of the acoustical signal of stressed and unstressed speech. Carnahan Conference on Crime Countermeasures. Lexington, Kentucky, *Proceedings:*19-21.

Mead, K. O. (1974) Identification of Speakers from Fundamental-Frequency Contours in Conversational Speech. Joint Speech Research Unit, Report No. 1002.

Meeker, W. F., Martin, T. B., and Herscher, M. B. (1967) Automatic speaker authentication using speech recognition techniques. Abstract. *Journal of the Acoustical Society of America* 42:1182.

Meeker, W. F. (1967) Speaker authentication techniques. Technical Report Ecom-02526-F. Ft. Monmouth, N.J.: U.S. Army Electronics Command.

Meltzer, D., and Lehiste, I. (1972) Vowel and speaker identification in natural and synthetic speech. Abstract. *Journal of the Acoustical Society of America* 51:131.

Miller, J. E. (1964) Decapitation and recapitation, a study of voice quality. *Journal of the Acoustical Society of America* 36:2002.

Miller, J. E., and Mathews, M. V. (1963) Investigation of the glottal waveshape by automatic inverse filtering. *Journal of the Acoustical Society of America* 35:1876.

Mohn, W. S., Jr. (1970) Two statistical feature evaluation techniques applied to speaker identification. IEEE Symposium on Feature Extraction and Selection in Pattern Recognition. Chicago, Illinois. *Proceedings:* 246-255.

Mysak, E. D. (1959) Pitch and duration characteristics of older males. *Journal of Speech and Hearing Research* 2:46-54.

National Research Council (1971) *Research on Speaker Verification.* Working Group 53, Committee on Hearing, Bioacoustics and Biomechanics. Washington, D.C.: National Academy of Sciences.

Nelson, A. L., Herscher, M. B., Martin, T. B., Zadell, H. J., and Falter, J. W. (1967) Acoustic recognition by analog feature-abstraction techniques. In W. Wathen-Dunn, ed., *Models for the Perception of Speech and Visual Form.* Cambridge: M.I.T. Press.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. (1957) *The Measurement of Meaning.* Urbana: University of Illinois Press.

Paul, J. E. Semi-Automatic Speaker Identification System: Final Report. Prepared for The Aerospace Corporation. 1976.

Paul, J. E., Ellington, J. W., and Rennick, R. J. Semi-Automatic Speaker Identification System: Laboratory Test Report. Prepared for the Aerospace Corporation. 1975.

Paul, J. E., Rabinowitz, A. S., Riganati, J. P., and Richardson, J. M. C74-1184/501 Semi-Automatic Speaker Identification System (SASIS) Analytical Studies. Summary. Prepared for the Aerospace Corporation. December 1974.

Paul, J. E., Rabinowitz, A. S. Riganati, J. P., and Richardson, J. M. C74-1184/501 Semi-Automatic Speaker Identification System (SASIS) Analytical Studies Final Report. Prepared for the Aerospace Corporation. December 1974.

Paul, J. E., Jr., Rabinowitz, A. S., Riganati, J. P., Richardson, J. M. Development of Analytical Methods for a Semi-Automatic Speaker Identification System.

Peters, R. W. (1954) Studies in extra messages: listener identification of speakers' voices under conditions of certain restrictions imposed upon the voice signal. Project Report No. MM 001 064-01-30. Pensacola, Fla.: U.S. Naval School of Aviation Medicine.

Peters, R. W. (1954) Studies in extra messages: the effects of various modifications of the voice signal upon the ability of listeners to identify speakers' voices. NM 001-104-500. Joint Report 61. Pensacola, Fla.: USNSAM

Pfeifer, L. L. (1974) Inverse Filter for Speaker Identification. Speech Communications Research Laboratory RADC-TR-74-214. Final Report. Griffiss Air Force Base, New York: Rome Air Development Center, Air Force Systems Command, U.S. Air Force.

Pollack, I., Pickett, J. M., and Sumby, W. H. (1954) On the identification of speakers by voice. *Journal of the Acoustical Society of America* 26:403-406.

Potter, R. K. (1945) Visible patterns of sound. *Bell System Monograph #1368.*

Potter, R. K., Kopp, G. A., and Green, H. C. (1947) *Visible Speech.* New York: D. van Nostrand (reprinted by Dover).

Poza, F. (1974) Voiceprint identification: its forensic applications. Carnahan Conference on Crime Countermeasures. Lexington, Kentucky. *Proceedings.*

Poza, F. Comments on Voiceprint Identification. Presented at the California Association of Criminalists Forty-Second Semiannual Seminar at Stanford Research Institute, Menlo Park, California, 1973.

Prestigiacomo, A. J. (1962) Amplitude contour display of sound spectrograms. *Journal of the Acoustical Society of America* 34:1684-1688.

Presti, A. J. (1966) High-speed sound spectrograph. *Journal of the Acoustical Society of America* 40:628-634.

Pruzansky, S. (1963) Pattern-matching procedure for automatic talker recognition. *Journal of the Acoustical Society of America* 35:354-358.

Pruzansky, S., and Mathews, M. V. (1964) Talker-recognition procedure based on analysis of variance. *Journal of the Acoustical Society of America* 36:2041-2047.

Pruzansky, S., and Stevens, B. A. (1974) Speaker Recognition 1937-1973. Bibliography 266. Murray Hill, N.J.: Bell Laboratories.

Ptacek, P. H., Sander, E. K., Maloney, W. H., and Jackson, C. C. R. (1966) Phonatory and related changes with advanced age. *Journal of Speech and Hearing Research* 9:353-360.

Ramishvili, G. S. (1965) Automatic Recognition of Speaking Persons. Report FTD-TT-65-1079. Wright-Patterson Air Force Base, Foreign Technology Division: Air Force Systems Command, U.S. Air Force.

Ramishvili, G. S. (1966) Automatic voice recognition. *Engineering Cybernetics* 5:84-90.

Ramishvili, G. S. (1972) On the problem of automatic identification of voices. 1972 Conference on Speech Communication and Processing. Newton, Mass., April 24-26, 1972. Sponsored by the Institute of Electrical and Electronic Engineers, Inc. and the Air Force Cambridge Research Laboratories. *Conference Record:* 279-282.

Ramishvili, G. S. (1974) Experiments on automatic verification of speakers. Second International Joint

Conference on Pattern Recognition. Copenhagen. *Proceedings:*389-393.

Ramishvili, G. S. and Tushishvili, M. A. (1976) On the connection of some time characteristics of speech signal with the individuality of voice. 1976 IEEE International Conference on Acoustics, Speech, and Signal Processing. Philadelphia, April 12-14, 1976. *Record:* 730-733.

Reich, A. R., Moll, K. L., and Curtis, J. F. (1976) Effects of selected vocal disguises upon spectrographic speaker identification. *Journal of the Acoustical Society of America* 60:919-925.

Reitboeck, H. J. Speaker identification over telephone transmission channels. Carnahan Conference on Crime Countermeasures. Lexington, Kentucky. *Proceedings.*

Reitboeck, H. J., Brody, T. P., and Thomas, D. T., Jr. Speaker-identification with real time formant extraction. 1977 IEEE Conference on Acoustics, Speech, and Signal Processing. Hartford, May 9-11, 1977. *Record:* 761-763.

Rekieta, T. W. and Hair, G. D. (1972) Mimic resistance of speaker verification using phoneme spectra. Abstract. *Journal of the Acoustical Society of America* 51:131.

Rennick, R. J. Semi-Automatic Speaker Identification System Final Report. Prepared for the Aerospace Corporation. 1974.

Richards, J. R. (1969) Speaker authentication using a formant-tracking vocoder. Abstract. *Journal of the Acoustical Society of America* 46:90.

Rosenberg, A. E. (1972) Listener performance in a speaker-verification task with deliberate imposters. Abstract. *Journal of the Acoustical Society of America* 51:132.

Rosenberg, A. E. (1973) Listener performance in speaker verification tasks. *IEEE Transaction on Audio and Electroacoustics* AU-21:221-225.

Rosenberg, A. E. (1974) A practical implementation of an automatic speaker verification system. Eighth International Congress on Acoustics. London. *Proceedings* 1:268.

Rosenberg, A. E. (1976) Evaluation of an automatic speaker-verification system over telephone lines. *Bell System Technical Journal* 55:723-744.

Rosenberg, A. E. and Sambur, M. R. (1975) New Techniques for Automatic Speaker Verification. *IEEE Transactions on Acoustics, Speech and Signal Processing* ASSP-23.

Rothman, H. B. (1975) Perceptual (aural) and spectrographic investigation of speaker homogeneity. Abstract. *Journal of the Acoustical Society of America* 58:S107.

Sambur, M. R. Speaker recognition and verification using linear prediction analysis. PH.D. thesis, Massachusetts Institute of Technology, 1972.

Sambur, M. R. (1975) Selection of acoustic features for speaker identification. *IEEE Transactions on Acoustics, Speech, and Signal Processing* ASSP-23:176-182.

Sambur, M. R. (1976) Text independent speaker recognition using orthogonal linear prediction. 1976 IEEE International Conference on Acoustics, Speech, & Signal Processing. Philadelphia, April 12-14, 1976. *Record:*727-729.

Sarma, V. V. S. and Yegnanarayana, B. (1976) Cascade realization of digital inverse filter for extracting speaker dependent features. 1976 IEEE International Conference on Acoustics, Speech, & Signal Processing. Philadelphia, April 12-14, 1976. *Record:*723-726.

Schroeder, M. R. (1968) Similarity measure for automatic speech and speaker recognition. *Journal of the Acoustical Society of America* 43:375-377.

Schwartz, M. F. (1968) Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America* 43:1178-1179.

Sebestyen, G. S. (1962) *Decision Making Processes in Pattern Recognition.* New York: Macmillan.

Shearme, J. N., and Holmes, J. N. (1959) An experiment concerning the recognition of voices. *Language and Speech* 2:123-131.

Shridhar, M., and Vidalon, M. (1976) An algorithm for speaker verification. Abstract. *Journal of the Acoustical Society of America* 60:S13.

Silbiger, H. R. (1966) Voice classification by hierarchical clustering. Abstract. *Journal of the Acoustical Society of America* 40:1282.

Smith, J. E. (1962) Decision-theoretic speaker recognizer. Abstract. *Journal of the Acoustical Society of America* 34:1988.

Smrkovski, L. L. (1975) Collaborative study of speaker identification by the voiceprint method. *Journal of the Association of Official Analytical Chemists* 58: 453-456.

Smrkovski, L. L. Voice Identification: Michigan Department of State Police, Voice Identification Unit. 1976.

Smrkovski, L. L. (1976) Study of speaker identification by aural and visual examination of non-contemporary

speech samples. *Journal of the Association of Official Analytical Chemists* 59:927-931.

Snider, D. L. (1970) Frequency and time-domain differentiation of speech contour data for voice recognition. Abstract. *Journal of the Acoustical Society of America* 48:129.

Speaker Verification. (1974) RADC-TR-74-179 Final Report. Griffiss Air Force Base, New York: Rome Air Development Center, Air Force Systems Command, U.S. Air Force.

Speaker Verification II. RADC-TR-75-274 Final Technical Report. Griffiss Air Force Base, New York: Rome Air Development Center, Air Force Systems Command, U.S. Air Force.

Starkweather, J. A. (1956) Content-free speech as a source of information about the speaker. *Journal of Abnormal Social Psychology* 52:394-402.

Steffen-Batog, M., Jassem, W. and Gruszka-Koscielak, H. (1970) Statistical distribution of short-term F_o values as a personal voice characteristics. Pp. 196-206 in Wiktor Jassem, ed., *Speech Analysis and Synthesis*. Vol. 2. Warsaw: Polish Academy of Sciences.

Stevens, K. N., Williams, C. E., Carbonell, J. R., and Woods, B. (1968) Speaker authentication and identification: a comparison of spectrographic and auditory presentations of speech material. *Journal of the Acoustical Society of America* 44:1596-1607.

Stewart, L. C., Houde, R. A., and Larkin, W. D. (1973) The VSTA: An Approach to the Speech Training Instrumentation Problem. Presented at the 1973 Carnahan Conference on Electronic Prosthetics, Lexington, Kentucky.

Stuntz, S. E. (1963) Speech intelligibility and talker recognition tests of Air Force Communication Systems. Technical Document Report ESD-TDR-63-224. L. G. Hanscom Field, Bedford, Mass.: Electronic Systems Division, Air Force Systems Command, U.S. Air Force.

Su, L-S., Li, K-P. and Fu, K. S. (1974) Identification of speakers by use of nasal coarticulation. *Journal of the Acoustical Society of America* 56:1876-1882.

Tosi, O. (1968) Speaker Identification Through Acoustic Spectrography. 14th International Congress on Logopedics and Phoniatrics. Paris. *Proceedings des séances du 14e congrès international, Paris, 1968*.

Tosi, O. (1975) The problem of speaker identification and elimination. Pp. 399-430 in S. Singh, ed., *Measurement Procedures in Speech, Hearing, and Language*. Baltimore: University Park Press.

Tosi, O., Greenwald, M., and Supal, C. Voice Identification Study of Minority Group. Paper presented at the May 1976 Meeting of the International Association of Voice Identification. Lexington, Kentucky. 1976.

Tosi, O., Oyer, H., and Nash, E. (1972) Latest developments in voice identification. Abstract. *Journal of the Acoustical Society of America* 51:132.

Tosi, O., Oyer, H., Pedrey, C., Lashbrook, B., and Nicol, J. (1971) An experiment on voice identification by visual inspection of spectrograms. Abstract. *Journal of the Acoustical Society of America* 49:138.

Tosi, O., Oyer, H. J., Lashbrook, W. B., Pedrey, C., and Nicol, J. Voice Identification Through Acoustic Spectrography. Department of Audiology and Speech Sciences, Michigan State University, East Lansing, Michigan. Prepared for the Law Enforcement Assistance Administration, U.S. Department of Justice.

Tosi, O., Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J., and Nash, E. (1972) Experiment on voice identification. *Journal of the Acoustical Society of America* 51:2030-2043.

Tosi, O., Oyer, H., Lashbrook, W., Pedrey, C., Nicol, J., and Nash, E. (1972) Experiment on voice identification. Abstract. *Journal of the Acoustical Society of America* 51:132.

Truby, H. M. (1972) Voice recognition by man, animal, and machine. 7th International Congress of Phonetic Sciences. Montreal. *Proceedings* 233-257.

Ungeheuer, G. (1965) A simple procedure for the acoustical classification of speakers. (in German). Fifth International Congress on Acoustics. Liege. *Proceedings*.

Vidalon, M., Shridhar, M., and Canas, M. (1977) Speaker verification using composite reference. 1977 IEEE International Conference on Acoustics, Speech, & Signal Processing. Hartford, May 9-11, 1977. *Record:*758-763.

Voiers, W. D. (1964) Perceptual bases of speaker identity. *Journal of the Acoustical Society of America* 36:1065-1073.

Voiers, W. D. (1965) Performance Evaluation of Speech Processing Devices II. The role of individual differences. Report AFCRL-66-24. Bedford, Mass.: Air Force Cambridge Research Laboratories, Office of Aerospace Research.

Voiers, E. D., Cohen, M. F., and Mickunas, J. (1965) Evaluation of Speech Processing Devices. I: Intelligibility, Quality, Speaker Recognizability. Final

Report. Bedford, Mass.: Air Force Cambridge Research Laboratories.

Wachter, K. W. (1970) Talker recognition on large populations. Abstract. *Journal of the Acoustical Society of America* 47:66.

Wakita, H. An approach to vowel normalization. Paper presented to the 89th Meeting of the Acoustical Society of America, 1975.

Welch, P. D. and Wimpress, R. S. (1961) Two multivariate statistical computer programs and their application to the vowel recognition problem. *Journal of the Acoustical Society of America* 33:426-434.

Williams, C. E. (1964) The Effects of Selected Factors on the Aural Identification of Speakers. Section III: Methods for Psychoacoustic Evaluation of Speech-Communication Systems. Technical Document Report ESD-TDR-65-163. L. G. Hanscom Field, Bedford, Mass.: Electronic Systems Division, Air Force Systems Command, U.S. Air Force.

Williams, C. E. and Stevens, K. N. (1972) Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America* 52:1238-1250.

Williams, C. E., Stevens, K. N., and Hecker, M. H. L. (1970) Acoustical manifestations of emotional speech. Abstract. *Journal of the Acoustical Society of America* 47:66.

Williamson, J. A. An Investigation of Several Factors which Affect the Ability to Identify Voices as Same or Different. Unpublished dissertation in phonetics, University of Edinburgh, 1961.

Wolf, J. J. (1969) Acoustic measurements for speaker recognition. Abstract. *Journal of the Acoustical Society of America* 46:89.

Wolf, J. J. (1970) Simulation of the measurement phase of an automatic speaker recognition system. Abstract. *Journal of the Acoustical Society of America* 47:83.

Wolf, J. J. (1972) Efficient acoustic parameters for speaker recognition. *Journal of the Acoustical Society of America* 51:2044-2056.

Young, M. A. and Campbell, R. A. (1967) Effects of context on talker identification. *Journal of the Acoustical Society of America* 42:1250-1254.

Zalewski, J., Majewski, W., and Hollien, H. (1974) Cross-correlation between long-term speech spectra as a criterion for speaker identification. Eighth International Congress on Acoustics. London. *Proceedings* 1:275.

APPENDIX

C

Legal Issues in Voice Identification: A Bibliographic Review

CHRISTOPHER SMEALL

GENERAL LITERATURE

Treatises

C. McCormick, EVIDENCE (2d ed. 1972)
Criticizes *Frye* test; brief discussion of voicegrams in chapter on experimental and scientific evidence.

A. Moenssens, R. Moses, & F. Inbau, SCIENTIFIC EVIDENCE IN CRIMINAL CASES (1973)
Separate chapter on spectrographic voice identification; discussion of technique and early case law.

J. R. Richardson, MODERN SCIENTIFIC EVIDENCE (2d ed. 1974)
Separate chapter on spectrographic voice identification; discussion of the history of the technique.

J. Weinstein & M. Berger, WEINSTEIN'S EVIDENCE (1977)
Multivolume treatise on the federal rules of evidence; treatment of case law under each rule.

Periodicals

Beuscher, *The Use of Experts by the Courts,* 54 HARV. L. REV. 1105 (1941)
Proposes a greater role for reference of complex factual issues to experts.

Boyce, *Judicial Recognition of Scientific Evidence in Criminal Cases,* 8 UTAH L. REV. 313 (1963)

Discusses problems with *Frye* rule and compares radar, breath intoxication, lie detector, and psychiatric evidence; argues for rejection of *Frye*.

Korn, *Law, Fact, and Science in the Courts,* 66 COLUM. L. REV. 1080 (1966)
Proposes a general framework for the analysis of the use of expert evidence by the courts; discusses doctrine of judicial notice, relative competence of court and jury.

Martin, *The Proposed Science Court,* 75 MICH. L. REV. 1058 (1977)

Strong, *Questions Affecting the Admissibility of Scientific Evidence,* 1970 U. ILL. L. FOR. 1
Discussion of *Frye* standard and of the characteristics of those types of scientific evidence that have been excluded under *Frye*.

Note, *Did Your Eyes Deceive You? Expert Psychological Testimony on the Unreliability of Eyewitness Identification,* 29 STAN. L. REV. 969 (1977)
Proposes admission of expert psychological testimony on the unreliability of eyewitness identification; discusses criteria for admissibility of such evidence, including conformity to generally accepted explanatory theory and balancing of probative value and prejudice.

VOICEGRAM IDENTIFICATION

Decker & Handler, *Voiceprint Identification Evidence--Out of the Frye Pan and into Admissibility,* 26 AM. U. L. REV. 314 (1977)
Favoring admission.

Greene, *Voiceprint Identification: The Case in Favor of Admissibility,* 13 AM. CRIM. L. REV. 171 (1975)
Favoring admission.

Jones, *Danger--Voiceprints Ahead,* 11 AM. CRIM. L. REV. 549 (1973)
Opposing admission.

Kamine, *The Voiceprint Technique: Its Structure and Reliability,* 6 SAN DIEGO L. REV. 213 (1969)

Siegel, *Cross-examination of a "Voiceprint" Expert: A Blueprint for Trial Lawyers*, 12 CRIM. L. BULL. 509 (1976)

Comment, *The Evidentiary Value of Spectrographic Voice Identification*, 63 J. Crim. L., Criminology & Police Science 343 (1972)

Note, *The Voiceprint Dilemma: Should Voices be Seen and Not Heard?*, 35 MD. L. REV. 267 (1975)
Opposing admission.

Note, *Voiceprints: The End of the Yellow Brick Road*, 8 U.S.F.L. REV. 702 (1974)

APPENDIX

D

Committee on Evaluation of Sound Spectrograms

COMMITTEE FORMATION

The Committee on Evaluation of Sound Spectrograms was established by the National Research Council in 1976 at the request of the Federal Bureau of Investigation (FBI) of the United States Department of Justice.* The personnel in the FBI laboratory noted that there is a continuing controversy regarding the reliability of speaker identification technology and the admissibility of testimony based on that technology and that courts at various levels have been ruling both for and against the admissibility of evidence based on application of that technology. Accordingly, the FBI asked the National Academy of Sciences to form a committee to review the status of speaker identification techniques and practices and provide an assessment of that status and probable future developments.

The National Academy of Sciences accepted this task through the Assembly of Behavioral and Social Sciences of the National Research Council. That Assembly sought the advice of its Committee on Hearing, Bioacoustics, and Biomechanics (CHABA); CHABA convened an ad hoc panel that considered the task, developed an initial work statement, and suggested persons for committee membership.

*The name given to the Committee is somewhat misleading on two counts: (1) the Committee concerned itself with all forms of voice identification and (2) at present, voicegram examiners use a combination of aural and visual methods; they do not examine voicegrams alone.

Initially, a committee of seven members was formed, with specialties representing the multiplicity of disciplines involved in voice identification and its uses. By the time of the Committee's third meeting, the final membership of eight was established. Brief biographical sketches of the Committee members and staff appear in Appendix E.

COMMITTEE TASK

Scope

During its first two meetings, the Committee developed and refined a set of tasks that it believed were responsive to the sponsor's interests and the general need and that could be addressed within a reasonable time span from first meeting to published report. These tasks are described below.

1. To examine the characteristics of speech relevant to talker identification, their representation by voicegrams, and the use of voicegrams in the task of identifying speakers.
2. To consider error rates in identifying speakers by the use of information contained in speech sounds and the factors that influence these rates.
3. To consider new methods of identifying speakers by the use of information contained in speech sounds.
4. To describe the training of voice identification examiners and to consider optimal training and testing of the examiners.
5. To search the relevant scientific and legal literature. To examine and describe existing data bases for evaluating techniques of talker identification. To prepare a review paper and a bibliography.

In addition to addressing the primary questions of the accuracy of voice identification by the present combination of aural and visual techniques, the Committee gave consideration to potential new methods and to areas of research, development, and practice that bear on all methods of voice identification.

Excluded Issues

The Committee considered several issues that are related to voice identification, and decided not to address them.

These issues, and the Committee's reasons for not addressing them, are detailed below.

Admissibility Chapters 3 and 4 of this report present the issue of admissibility of voice identification evidence and make explicit the Committee's decision to take no position on the issue of admissibility. The Committee has taken care throughout the report to separate issues of scientific facts from issues of values, and the admission of voice identification evidence is not a question of scientific fact. For a discussion of this issue, see the body of the report.

Privacy Issues of privacy can arise in connection with voice identification of the kind with which the Committee is concerned. First, the recording of the voice of the unknown criminal might have been made in a way that violates privacy. Second, the effort to obtain a voice sample from a suspect also might violate privacy. The Committee decided not to deal with those issues. Because it is possible to obtain both original recordings and suspect samples without violating privacy, the Committee takes the position that the issue of the accuracy of reaching conclusions from comparisons of speech samples can be considered separately from the question of the source of the samples.

Tape Tampering The authenticity of voice recordings is an important issue raised during the Committee's open meetings and during its information-gathering phase. Since this issue does not involve the technical question of voice identification, the Committee did not address it, except to note that, in the voice identification situations with which it was concerned, authenticity can be controlled to some extent by administrative efforts to safeguard the tapes.

Other Issues Two other issues were not considered. One issue is the observation of psychophysiological stress in speech. Continuing publicity concerning voice stress analyzers and their implications for lie detection in speech led the Committee to consider whether they would address this point. The other issue is the determination of the linguistic content of an utterance that is difficult to understand (for reasons of either noise or speaker-related difficulties). Although both topics bear on voice identification, the Committee felt they could not address

them. Neither of these topics was in the original request; both would have required extensive additional time and effort.

To recapitulate, the Committee operated under the assumption that a speech signal has been recorded, that it it pertinent to a legal or investigative action, that samples of speech have been obtained from other persons, and that all of the samples have been acquired legally and have been protected from tape tampering. Then the task is determining whether the unknown speaker is or is not one of the known speakers. In short, the Committee confined its work to the question of voice identification and the degree of accuracy with which it could be accomplished.

COMMITTEE OPERATION

In addition to holding closed deliberative sessions, the Committee interacted with relevant parties through correspondence, invited discussions, open meetings, and reviews of literature. Over 200 individuals and organizations were informed of the Committee's existence and purpose. These covered the scientific, legal, and law enforcement communities, including all those whose interest in the subject was known to the Committee.

Five major meetings were held, the first of two days', the last four of three days' duration. At the first four of these meetings, selected individuals talked to the Committee as invited discussants. At the second and third meetings, open sessions were held to which all interested persons were invited by public notice.

One open session was held at a meeting in San Diego and one in Washington, D.C. At the San Diego meeting, the Committee reached a large number of scientists who were attending relevant professional meetings at that time as well as some West Coast members of the legal community. In Washington, D.C., the open session of the meeting drew mainly law enforcement and investigative personnel and members of the press. As a result of these sessions, the Committee was able to hear the views of and have discussions with: (1) voice identification practitioners, (2) scientists who were both "pro" and "con," (3) lawyers who were both "pro" and "con," (4) one judge, and (5) identification specialists in other technologies. In addition, small subgroups of the Committee met from time to time, mainly for writing sessions.

In doing its work, the Committee benefited from the assistance of a large number of people from the various constituencies concerned with voice identification. Some of these people met with the Committee as a whole, some participated in the Committee's open meetings, others met with members of the Committee or the staff, and some communicated with the Committee by mail or telephone. The Committee is grateful to all of them for providing valuable information; their names are listed below. Finally, the Committee would like to thank the project coordinators at the Federal Bureau of Investigation, who helped us in many ways: William E. Harward, Steven A. Killion, and Bruce E. Koenig.

Speech Science Community

Joseph G. Agnello, Professor, Speech Pathology/Audiology, University of Connecticut
James E. Atkinson, Naval Underwater Systems Center, New London, Connecticut
John Boehm, Department of Defense
Peter D. Bricker, Bell Laboratories, Murray Hill, New Jersey
David Broad, Speech Communications Research Laboratory, Inc., Los Angeles
Frank R. Clarke, Consultant, San Diego
Anthony M. Collins, Department of Engineering Physics, Australian National University
Edith L. R. Corliss, Physicist, National Bureau of Standards, and Vice-President-Elect, Acoustical Society of America
Joseph de Clerk, Electronic Engineer, Technical Management Division, SINGARS
Peter B. Denes, Head, Speech and Communications Research Department, Bell Laboratories
George R. Doddington, Texas Instruments, Inc.
Michael Earle, Sound Research Service, Carpinteria, California
Benjamin N. Early, Program Manager, Aerospace Corporation
Victoria A. Fromkin, Linguistics Department, University of California at Los Angeles
John J. Godfrey, University of Texas at Dallas
Barry Hazen, University of North Carolina
Michael H. L. Hecker, Senior Research Engineer, Stanford Research Institute International
John M. Heinz, Johns Hopkins University School of Medicine
Harry Hollien, Institute for Advanced Study of Communication Processes, University of Florida

John N. Holmes, Head, Joint Speech Research Unit, Cheltenham, Gloustershire, England
Michael Hoshiko, Southern Illinois University
Robert A. Houde, Director, Center for Communications Research, Rochester
Kathleen Houlihan, Assistant Professor, Department of Linguistics, University of Minnesota
Arthur S. House, Institute for Defense Analysis, Princeton, New Jersey
Wiktor Jassem, Head, Acoustic Phonetics Laboratory, Polish Academy of Sciences
Robert Kennel, Aerospace Corporation
Dennis H. Klatt, Research Laboratory of Electronics, Massachusetts Institute of Technology
Peter N. Ladefoged, Linguistics Department, University of California at Los Angeles
Norman J. Lass, Department of Speech Pathology and Audiology, West Virginia University
Wayne A. Lea, Speech Communications Research Laboratory, Inc., Los Angeles
Wojciech Majewski, Institute of Telecommunication and Acoustics, Technical University of Wroclaw
John D. Markel, Signal Technology, Inc., Santa Barbara
Robert E. McGlone, University of Nebraska
W. F. Meeker, Radio Corporation of America
Kenneth H. Mill, Associate Dean of Faculties, University of Iowa
Edward P. Neuberg, Department of Defense
A. Michael Noll, American Telephone and Telegraph
John Ohala, Department of Linguistics, University of California at Berkeley
George Papcun, Phonetics Laboratory, University of California at Los Angeles
Larry L. Pfeifer, Signal Technology, Inc., Santa Barbara
L. C. W. Pols, Institute for Perception TNO, Netherlands
Sandra Pruzansky, Bell Laboratories, Murray Hill, New Jersey
Herbert Reitboeck, Westinghouse Research Laboratory, Pittsburgh
Aaron Rosenberg, Acoustics Research Department, Bell Laboratories
Howard B. Rothman, Institute for Advanced Study of Communication Processes, University of Florida
Edgar A. G. Shaw, Head, Acoustics Section, Division of Physics, National Research Council, Ottawa, Canada
Lowell K. Smith, U.S. Government
Connie A. Supal, Michigan State University

Henry M. Truby, Board of Directors, International Association of Voice Identification
Visvaldis A. Vitols, Rockwell International
Hisaski Wakita, Research Engineer, Speech Communications Research Laboratory, Inc., Los Angeles
Clifford A. Wood, Department of Defense

Legal and Judicial Community

John E. Ackerman, Dean, National College of Criminal Defense Lawyers and Public Defenders, Bates College of Law
John Carney, Deputy Attorney General, State of California
Raymond P. Flynn, Attorney, Washington, D.C.
Carroll L. Gilliam, Attorney, Grove, Jaskiewicz, Gilliam and Cobert, Washington, D.C.
Henry F. Greene, Assistant United States Attorney, Washington, D.C.
Stephen Heiser, Attorney, Singer and Osterhoudt, San Francisco
William R. Jones, Center for the Study of Law, Nova University
Gene C. Lange, Attorney, Ragan and Mason, Washington, D.C.
Thomas D. Luchetti, Attorney, Miller & Hickey, Rockford, Illinois
W. Brown Morton, Jr., Attorney, Irons and Sears, Washington, D.C.
Tim Murphy, Judge, Superior Court of the District of Columbia
James M. Russ, Attorney, Law Offices of James M. Russ, Orlando, Florida
Kenneth R. Thomas, Attorney, Los Angeles
William A. Thomas, Research Attorney, American Bar Foundation
William F. Urich, Attorney, Law Offices of Robert L. Moran, San Francisco
Edwin Williams, Attorney, Kellog, Williams and Lyons, Washington, D.C.

Forensic Community

George J. Bonebrake, Supervisory Fingerprint Specialist, Federal Bureau of Investigation
Linda Chiari, Voice Identification, Inc.
Owen Greenspan, Detective, Police Department, City of New York
Malcolm C. Hall, Principal Research Officer, Commonwealth Police, Canberra City, Australia

Malcolm E. Hall, Detective Sargeant, Michigan State Police
Fred Lundgren, Bureau of Alcohol, Tobacco and Firearms, U.S. Department of the Treasury
David Paige, Trooper, Forensic Laboratory, Connecticut State Police
Joseph L. Peterson, Executive Director, Forensic Sciences Foundation
Eyal Shy, Criminal Identification Division, Israel Police H.Q.
Lonnie L. Smrkovski, Detective-Lieutenant, Michigan State Police
Charles T. Spitzer, Special Agent, Laboratory Division, Document Section, Federal Bureau of Investigation

APPENDIX

E

Biographical Sketches of Committee Members and Staff

Richard H. Bolt is chairman of the board emeritus of Bolt Beranek and Newman, Inc. With a BA in architecture and a PhD in physics from the University of California, he went in 1939 to the Massachusetts Institute of Technology as a National Research Council fellow in physics. During the next 30 years at the Massachusetts Institute of Technology, he held several positions including professor of acoustics, director of the acoustics laboratory, and lecturer in political science. He has been associate director of the National Science Foundation, principal consultant in biophysics for the National Institutes of Health, a fellow of the Center for Advanced Study in the Behavioral Sciences, and president of the International Commission on Acoustics and of the Acoustical Society of America, from which he received its biennial award for research contributions. He served as first chairman of the Committee on Hearing, Acoustics, and Biomechanics of the National Research Council, chairman of the U.S. District Court's Advisory Panel on White House Tapes, and visiting scientist of the Massachusetts legislature. He is a fellow of the American Academy of Arts and Sciences and the Institute of Electrical and Electronic Engineers and a member of the Institute of Noise Control Engineering, the National Academy of Engineering, and the American Bar Association-American Association for the Advancement of Science joint committee on science and law.

Franklin S. Cooper, an associate director of research for Haskins Laboratories, was one of its founders and president

from 1955 to 1975. He did undergraduate and graduate work at the University of Illinois and received a PhD in physics from the Massachusetts Institute of Technology. He began work at Haskins Laboratories after three years at the General Electric Research Laboratories. His other professional experience includes wartime service as head of the Liaison Office of the Office of Scientific Research and Development; postwar work with the United Nation's Atomic Energy Commission Group; advisory services to the National Research Council, the National Institutes of Health, and various agencies in the Department of Defense; and adjunct affiliations with Columbia University, Yale University, and the University of Connecticut. He was a fellow of the Center for Advanced Study in the Behavioral Sciences in 1964-1965; he was a member of the U.S. District Court's Advisory Panel on White House Tapes in 1973-1974; he is a fellow of the Acoustical Society of America and of the Institute of Electrical and Electronic Engineers, and a member of the Council on Foreign Relations and the Cosmos Club in Washington.

David M. Green is professor of psychophysics at Harvard University and chairman of the Department of Psychology and Social Relations. He received degrees from the University of Chicago and the University of Michigan and has served on the faculties of the Massachusetts Institute of Technology, the University of Pennsylvania, and the University of California at San Diego, before coming to Harvard in 1974. A recipient of a Guggenheim fellowship in 1973, he spent that year at St. John's College in Cambridge, England as an overseas fellow. Green is a former editor of the *Journal of the Acoustical Society of America* and the author of *Signal Detection Theory and Psychophysics* with John A. Swets (1966, 1974) and *An Introduction to Hearing* (1976) as well as many journal articles. He was elected to the National Academy of Sciences in 1978.

Sandra L. Hamlet is associate professor in the Department of Hearing and Speech Sciences at the University of Maryland. She received a BA from the University of Wisconsin in linguistics and the MA and PhD degrees from the University of Washington, in speech pathology and speech science, respectively. She is an associate member of the Acoustical Society of America, and a member of the American Association of Phonetic Sciences.

John G. (Jay) McKnight is president of Magnetic Reference Laboratory Inc. in Mountain View, California. He is president-elect of the Audio Engineering Society and president for the year 1978-1979. He received a BS in electrical engineering from Stanford University in 1952. In 1973-1974 he was a member of the U.S. District Court's Advisory Panel on White House Tapes. From 1972 to 1974 he was a consultant on audio systems and magnetic recording and engineering vice president of Magnetic Reference Laboratory Inc. From 1952 through 1972 he worked for Ampex Corporation, serving in the magnetic recording research group, the stereo tape division, and the professional audio division. He has published more than 40 technical papers on magnetic recording and audio engineering. He is a fellow of the Audio Engineering Society, and recipient of the Audio Engineering Society Award; a member of its review board; and has been a governor three times and chairman of the Standards Committee and of the Publications Policy Committee. He is a senior member of the Institute of Electrical and Electronic Engineers.

James M. Pickett is professor of speech communication research and, since 1964, director of the Sensory Communication Research Laboratory at Gallaudet College. He received a BA from Oberlin College and a PhD in experimental psychology from Brown University. He has served the National Research Council through his work on the Committee on Hearing, Bioacoustics, and Biomechanics. He is a fellow of the Acoustical Society of America and a member of the American Association for the Advancement of Science and the Psychonomic Society. In 1961 he was awarded a National Institutes of Health Special Research fellowship for research and study at the Speech Transmission Laboratory of the Royal Institute of Technology, in Stockholm. He has been chairman of the Technical Committee on Speech Communication of the Acoustical Society of America; of the Subcommittee on Speech Processing Aids for the Handicapped of the Insitute of Electrical and Electronic Engineering; of a writing group on a standard for calculating a speech interference index for noises of the American National Standards Institute and has served on other committees concerned with speech communication. He is currently a member of the Communicative Sciences Study Section of the National Institutes of Health and the Research Advisory Committee of the Boys Town Institute for Communicative Disorders in Children. Since 1952 he has published numerous research works in speech science and its applications.

Oscar Tosi is professor of audiology and speech sciences, and director of the Speech and Hearing Research Laboratory and the Institute of Voice Identification at Michigan State University. He received the ScD degree in engineering and applied physics from the University of Buenos Aires and the PhD in audiology, speech sciences, and electronics from Ohio State University. He directed the major study to date on voice identification at Michigan State University in 1968-1970. Since 1968 Tosi has testified as an expert on voice identification in many state and federal courts in the United States and Canada. He is a member of the American Speech and Hearing Association; an appointed member of the technical committee on speech communication from 1974-1977 of the Acoustical Society of America; a founder and director of the International Association of Voice Identification; and a member of the International Association of Phonetics. His publications include *Voice Identification: Theory and Legal Applications* (1978).

Barbara D. Underwood is professor of law at Yale Law School, where she has served on the faculty since 1972. She received a BA from Radcliffe College in history and science and a JD from the Georgetown University Law Center. She served as a law clerk to Chief Judge David L. Bazelon of the United States Court of Appeals for the District of Columbia Circuit and to Justice Thurgood Marshall of the United States Supreme Court. She is a member of the Bar of the District of Columbia and of the Bar of the United States Supreme Court. Her teaching and research interests include criminal law, evidence, and constitutional law.

Douglas L. Hogan, who was study director for the Committee on Evaluation of Sound Spectrograms, is a project manager at the System Development Corporation in McLean, Virginia. He received a BS from George Washington University and the SM and Electrical Engineer degrees from the Massachusetts Institute of Technology. He worked in the Department of Defense for 28 years in computer and communications engineering and in speech research. From 1970 to 1973 he was chairman of the Speech Communication Technical Committee of the Acoustical Society of America.