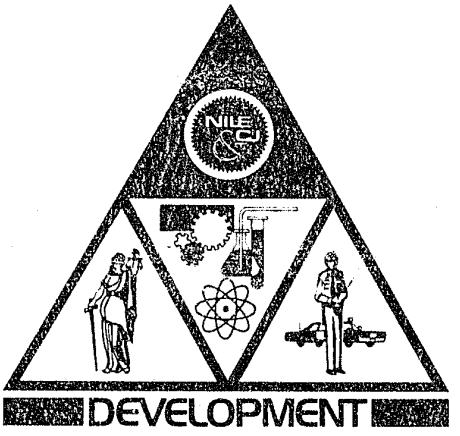
AEROSPACE REPORT NO.
ATR-74(7907)-1

EQUIPMENT SYSTEMS IMPROVEMENT PROGRAM

PRELIMINARY INVESTIGATION OF APPLICATIONS OF THE COMPUTER-AIDED SPEAKER IDENTIFICATION SYSTEM

Law Enforcement Development Group

June 1974



Prepared for

NATIONAL INSTITUTE OF LAW ENFORCEMENT AND CRIMINAL JUSTICE

Law Enforcement Assistance Administration

U.S. Department of Justice

THE AEROSPACE CORPORATION

Report No.
ATR-74(7907)-1

EQUIPMENT SYSTEMS IMPROVEMENT PROGRAM

PRELIMINARY INVESTIGATION OF APPLICATIONS OF THE COMPUTER-AIDED SPEAKER IDENTIFICATION SYSTEM

Law Enforcement Development Group
THE AEROSPACE CORPORATION
El Segundo, California

June 1974

Prepared for

NATIONAL INSTITUTE OF LAW ENFORCEMENT
AND CRIMINAL JUSTICE
Law Enforcement Assistance Administration
U.S. Department of Justice

Contract No. J-LEAA-025-73

This project was supported by Contract Number J-LEAA-025-73 awarded by the Law Enforcement Assistance Administration, U.S. Department of Justice, under the Omnibus Crime Control and Safe Streets Act of 1968, as amended. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

EQUIPMENT SYSTEMS IMPROVEMENT PROGRAM

PRELIMINARY INVESTIGATION OF APPLICATIONS OF THE COMPUTER-AIDED SPEAKER IDENTIFICATION SYSTEM

Approved

John O. Eylar, Jr., Director
Law Enforcement Development Group

ABSTRACT

This report documents an investigation of the present and future uses of voice identification in the law enforcement and criminal justice community. The purpose of the investigation is to estimate the nature and scope of the potential applications of a computer-aided speaker identification system now under development.

While the use of voice spectrograms (or voiceprints) for speaker identification has certain drawbacks, the technique is being used currently in criminal investigations and in courtroom cases. The report provides estimates of the scope of this use and how the use is expected to change as a consequence of supplementing the manual examination of spectrograms with the computer-aided system.

The results of the investigation indicate the principal criticism of manual examinations is that they are subjective and lack general scientific support. The computer-aided approach, being more quantitative in nature and specifying, in a direct way, the points of comparison between speech samples is expected to be more likely to gain acceptance by the scientific and criminal justice community. This acceptance, coupled with increased speed of operation for the computer-aided system, would make recorded speech samples a valuable source of physical evidence.

CONTENTS

CONTENTS (Continued)

TABLES

FIGURES

ACKNOWLEDGMENTS

The preliminary investigation described in this document was conducted by The Aerospace Corporation supported by members of the law enforcement, criminal justice, and speech science communities. It is appropriate to acknowledge the contributions made by the agencies contacted in the course of this investigation. These include: the Washington, D.C., Police Department, the St. Louis Police Department, the Tulsa Police Department, the New Orleans Police Department, the Santa Ana Police Department, the Pasadena Police Department, the Los Angeles Police Department, the Detroit Police Department, the Dade County Sheriff's Department, the Michigan State Police, the Bureau of Alcohol, Tobacco and Firearms of the Treasury Department, and the Los Angeles County Superior Court.

Mr. P. K. Broderick is the principal author of the report; he was assisted in its preparation by Mr. G. Papcun of the University of California at Los Angeles Linguistics Department and by Mr. R. Reider of Rockwell International Corp. Valuable assistance was provided in reviewing the report by Mr. B. L. Adams, Dr. J. F. Carpenter, Mr. T. H. Davies, Dr. D. P. Duclos, Mr. N. A. Mas, and Dr. S. Siegel. Finally, acknowledgment is due to Mr. H. W. Nordyke for his guidance and direction of the program.

SUMMARY

Speech has been demonstrated to be a distinctive feature of an individual's identity.[1] This fact has led to the increasing use of voice recordings as an important source of information in criminal investigations.[2,3] In order to use such recordings in the most effective way, some means of accurately identifying a speaker from recorded samples of his voice is needed. Present methods for comparing speech samples involve examination of the sounds on the recording by listening to them or by displaying them pictorially as "voiceprints" and matching the display patterns. These methods are time consuming, subjective, and often not acceptable as evidence in trials.

In order to improve the effectiveness of police in investigating crimes where voice recordings are involved, the National Institute of Law Enforcement and Criminal Justice has supported development of a computer-aided speaker identification system. The system is intended to provide quantitative, objective measures of the similarity between recorded voice samples. From these measures, and from a knowledge of statistical distributions of speech variations in individuals and populations, it will be possible to compute the likelihood that two independent speech samples were spoken by the same person. Results of analysis and development tests performed to date indicate that a high degree of accuracy is achievable using this method.[4]

The development of such a computer-aided speaker identification system is currently being conducted for the Institute by Rockwell International, Inc., under the direction of The Aerospace Corporation. This system,

termed the Semi-Automatic Speaker Identification System, will be ready for laboratory testing in October 1974. It is planned to follow laboratory testing with a pilot field test in early 1975.

As part of the development effort, The Aerospace Corporation conducted an investigation to estimate the size and scope of the ultimate applications of this equipment, based upon the present and projected use of voice identification by criminal justice organizations.

The investigation into the current use of voice identification revealed that voice samples were analyzed by voiceprint techniques in about 1500 criminal cases per year in the United States, but the results of these analyses were used in court only in about 25 cases. There are several considerations that tend to limit the use of voiceprint evidence in court. The principal one is the question of the admissibility of such evidence. Another is the expense involved since the thorough preparation required for evidentiary use may require a week or more of the voiceprint examiner's time (as opposed to the day or two for a routine investigative case).

In an attempt to estimate the potential applications of voice identification in reducing crime, five police departments were visited and surveyed. In addition, interviews with personnel from a number of other police departments, criminalistics laboratories, and judicial agencies were conducted to sample the attitude of the criminal justice community to voice identification in general and to computer-aided speaker identification in particular.

The survey results indicate that a large city police department would typically generate upwards of 2000 cases per year that could be assisted by voice identification, if a system for providing such identification in an

accurate, effective manner were available, and that the cost savings derived from more efficient use of manpower could approach $300,000 per year. This estimate is based upon the reduced manhours needed to conduct investigations and their greater effectiveness, leading to a higher incidence of confessions and a higher degree of confidence in cases brought to trial. These factors could result in additional savings through the lowering of costs related to judicial procedures for law enforcement agencies. Such overall savings would more than justify the expense of the Semi-Automatic Speaker Identification System. However, an additional important benefit, according to the law enforcement officials interviewed, is that on the basis of demonstrated experience significant deterrent effects could be expected in all crime categories where investigators make use of voice recordings.

It is apparent from the preliminary investigation described in this report that if the voice identification system is to be useful in the environment of the law enforcement community, it must provide results almost independent of the particular examiner using the system. In addition, the system must also be capable of providing voice sample comparisons on the major dialects encountered in police investigations. Furthermore, the effects of emotional stress, disguise and communication link degradation on speech fidelity must be known so that an examiner can take these effects into account when making his decision.

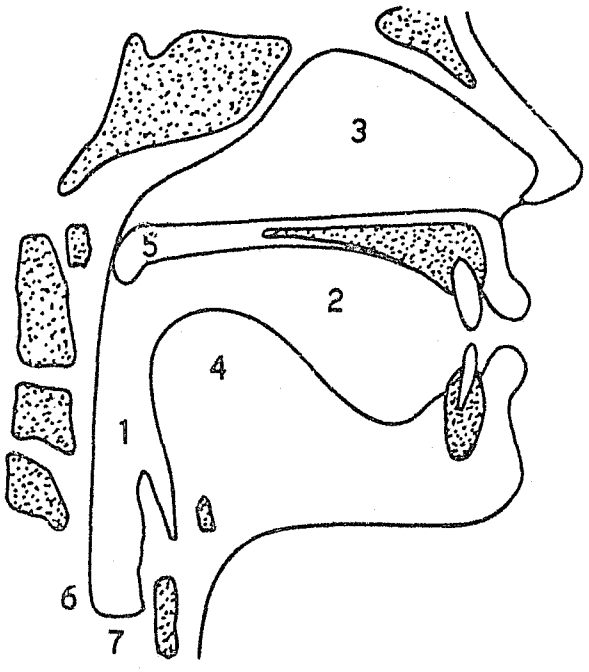
The investigation also shows that one of the most serious drawbacks to the use of voice identification techniques is the lack of general scientific support for its use in court proceedings. A computer-based semi-automatic system for processing voice recording information could conceivably attract

such support since comparisons would be based on objective, quantitative data. It is planned to conduct an intensive series of test programs in order to provide the data and test conditions required to gain acceptance of the computer-aided speaker identification concept by the speech science community.

CHAPTER I. INTRODUCTION

One of the most effective deterrents to crime is the knowledge on the part of the potential criminal that some trace or clue left during the crime may serve to identify him. This identification, leading to subsequent apprehension, prosecution and conviction, may be made in a variety of ways: fingerprints, bloodstains, hair, handwriting and body fluids are some of the individual specific features used by criminalists to identify criminals. In a communication-oriented society such as ours, however, crimes can be and often are committed remotely by telephone. Often, the call itself is the crime, as in a bomb threat or extortion case. Recorded voice communication is frequently the chief source of information to investigators of narcotics or gambling transactions. Increasingly, police and prosecutors are relying on the identification of criminals through the matching of their voices.

Speech is used for identification because it is a product of the speaker's individual anatomy and linguistic background. When air is expelled from the lungs, it passes through the glottis, which is the opening bounded on either side by the vocal folds (see Figure 1).[1] When the vocal folds are drawn together and air from the lungs is forced through them, they vibrate, making a buzzing sound. This sound is modified as it passes through the vocal tract, which is the tube formed principally by the pharyngeal cavity and the oral cavity. The shape of the vocal tract serves to concentrate sound energy at certain frequencies and reduce it in others. The formants are the areas in which vocal energy is concentrated by the effect of the vocal tract shape.



LEGEND

1. PHARYNGEAL CAVITY
2. ORAL CAVITY
3. NASAL CAVITIES
4. TONGUE MASS
5. SOFT PALATE
6. VOCAL FOLDS
7. GLOTTIS

Figure 1. Saggital Section of the Vocal Tract

During speech the shape of the vocal tract is continuously modified by movements of the tongue, lips, and other vocal organs. Thus, the quality of the speech sounds a speaker produces represents the sizes and shapes of his vocal organs and the way he uses them in speaking.

From a relatively short conversation, a listener can often detect substantial information about the speaker's characteristics, without regard to the content of the speech. Characteristics may include the speaker's sex, emotional state, ethnic or geographical origins, approximate age and, perhaps, state of health.

By hearing longer samples of an individual's speech, a listener would eventually learn to distinguish the speaker from a large number of different speakers. The listener has, in effect, derived a set of features which are highly specific to a given individual. The listener, using his knowledge of this feature set, can make decisions on whether or not a given speech sample was spoken by the original speaker. The feature set thus provides an index as to the speaker's identity. The feature set, in this case, is entirely subjective and is of use only to the specific listener. It is also very hard to quantify the degree of certainty when decisions are made based upon this recognition technique. The existence of some form of feature set, however, forms the basis for speaker identification.

While the term "speaker identification" is used throughout this report, the techniques discussed herein are often described as belonging to a "speaker discrimination" task. A discrimination task always involves two speech samples. In this task, a decision is rendered as to whether the speech samples were produced by the same speaker or by different speakers. Since this is the scenario of most forensic speaker identification cases, the term speaker identification can be considered in this report to be synonymous with speaker discrimination.

There are three general methods for speaker identification.[1] They are: by listening, by visual comparison of voice spectrograms, and by machine. Speaker identification by listening is most common. While under certain circumstances it is more accurate and reliable than the other methods, it is limited in that it is entirely subjective. The use of spectrograms is considered to

be a more objective method, because they exhibit graphic features of a speech signal which can be discussed in a fairly objective manner. This objectivity is compromised, however, since the features are interpreted subjectively and often are supplemented by listening when a decision is being reached. To overcome these limitations, considerable effort has been expended to develop machines capable of speaker identification. Decisions arrived at using such machines will be inherently objective and repeatable. But such machines will require high accuracy, which is equivalent to a high degree of discrimination capability, if they are to be used in law enforcement and criminal justice applications. For example, in the forensic situation, false identification could erroneously single out a particular individual as a suspect and result in the possible conviction of an innocent person. In investigative work, on the other hand, false rejections are equally important because they may lead to the elimination of a guilty person from consideration as a suspect.

CHAPTER II. BACKGROUND

A. The Use of Voiceprints[2]

A voiceprint (also called a sonogram) is a three-dimensional graph representing time, frequency, and intensity of speech sounds and is technically known as an acoustic spectrogram. These characteristics, as illustrated in Figure 2, are depicted as follows:

- Time is represented from left to right along the dimension of the horizontal axis.
- Frequency is represented along the dimension of the vertical axis with height along the axis proportional to frequency.
- Intensity is represented along the dimension of a gray scale in which darkness is proportional to intensity. Thus, dark areas represent regions of relatively intense sound energy.

In 1944 Gray and Kopp coined the word "voiceprint" in a report discussing identification of speakers by visual inspection of spectrograms and concluded that this method seemed to offer good possibilities. They aimed their efforts at helping the military. After World War II, voiceprint was almost forgotten. In 1962 Kersta reexamined voiceprint and claimed that spectrograms of several utterances of the same words by a given speaker always contain more similar spectral features than those produced by different speakers. Kersta concluded, therefore, that speaker identification by visual examination of spectrograms was reliable. According to Kersta, speaker recognition by visual inspection of spectrograms consists of

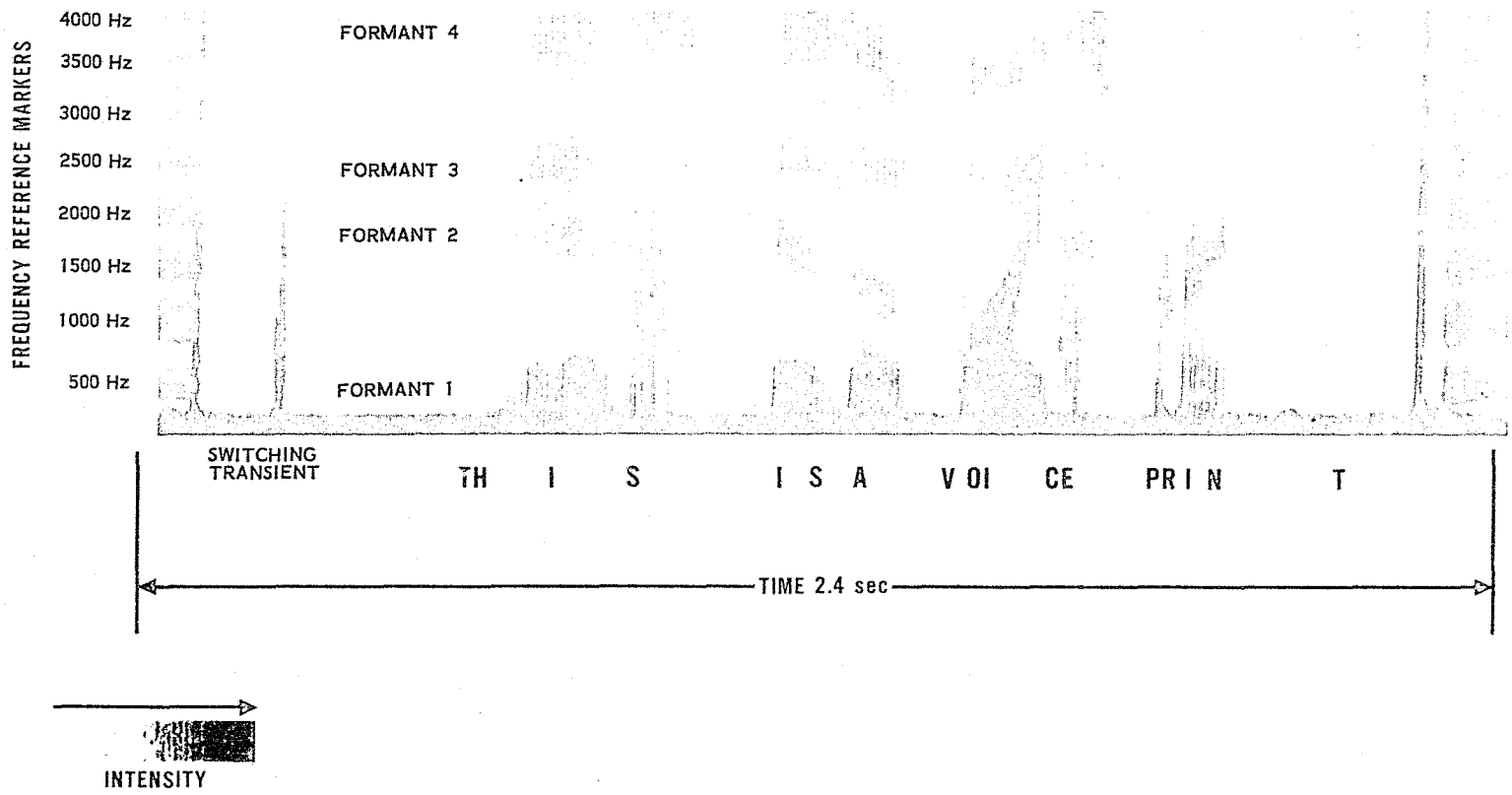


Figure 2. Labeled Spectrogram of "THIS IS A VOICEPRINT"

subjectively matching similarities found in pairs of spectrograms from different persons. The dissimilarities presented by the matched spectrograms are disregarded; they are assumed to be a result of intraspeaker variability. To back his claim, Kersta published the results of experiments he performed at Bell Laboratories. In these experiments, he observed fewer than one percent wrong identifications. Since 1962, Kersta has been producing legal testimony on speaker identification by using the voiceprint methods, and has also been offering training to law enforcement officers. Several speech scientists have questioned the results of Kersta's experiments and have expressed their concern over the legal application of the voiceprint technique prior to proving its accuracy and reliability through comprehensive and properly controlled experimentation.

To overcome these objections, the National Institute of Law Enforcement and Criminal Justice has supported efforts to evaluate the effectiveness of present voiceprint identification techniques. One such effort was conducted by Dr. O. Tosi of Michigan State University under an Institute grant.[3] After one month of training, 29 examiners were tested with spectrograms from 250 different speakers in a large variety of tests (34,992 trials).

The results of these experiments and their interpretation have been the subject of considerable controversy in the speech science community and have not resulted in the widespread acceptance of the technique. The spectrographic identification of a voice by a trained observer appears to rely on a broad assessment of loosely defined points of similarity rather than a carefully specified set of objectively defined spectrographic attributes. This makes

replication of experimental results by independent investigators highly unlikely and therefore prevents the acceptance of the technique as a recognized scientific procedure.

Despite these shortcomings, the use of voiceprints for speaker identification is expected to increase over the next few years as police and criminalistics laboratory personnel are being trained in this technique. Recent judicial decisions regarding the admissibility of evidence and the manner in which suspects are interrogated have forced police agencies to rely more on new techniques and equipment for investigative purposes.

B. The Use of Machines for Speaker Identification

The basic function of an automatic speaker recognition system is to extract speaker-dependent parameters from the speech signal and subject them to a statistical analysis. The function therefore involves two distinct processes: (a) parameters thought to be useful for differentiating among speakers are extracted from the speech signal, and (b) decision rules are applied to combinations of parameter values that represent particular speech samples. The choice of parameters (or features) influences the instrumentation requirements of the technique, the complexity of the decision rules, and the level of recognition that can be achieved.

A program to investigate the capabilities of machine-assisted speaker identification was initiated by the Michigan State Police. Supported by an Institute grant, this program involved a subcontract with Stanford Research Institute and Texas Instruments to develop a computer-assisted speaker identification system. Both organizations developed systems for extracting

features from speech in the form of a predefined utterance and processed these features using digital computers, recorders, and other general-purpose equipment. Their results were comparable and indicated that (a) the technique supplements information available from the voice spectrogram, and (b) the achieved accuracy encouraged further development.[4, 5]

Current efforts are aimed toward extending the machine-assisted technique into more general usage. A subcontract is presently in effect with Rockwell International to develop a computer-aided speaker identification system based on the concept developed by Stanford Research Institute and Texas Instruments. This system, termed the Semi-Automatic Speaker Identification System (SASIS), will be ready for laboratory and pilot field testing in FY 75. It will enable an operator to rapidly obtain measurements regarding the likelihood that a given speech sample and any one of several suspect speech samples were spoken by the same person.

The system configuration is shown in Figure 3. It utilizes a general-purpose computer coupled with high-speed data processing and pattern recognition algorithms specially designed for the speaker identification task. By statistical methods, the parts of the speech samples which best contribute to speaker discrimination are selected and compared with other samples. On the basis of these comparisons, the computer can measure the degree of similarity between a sample (e.g., from a bomb threat recording) and that from a suspect.

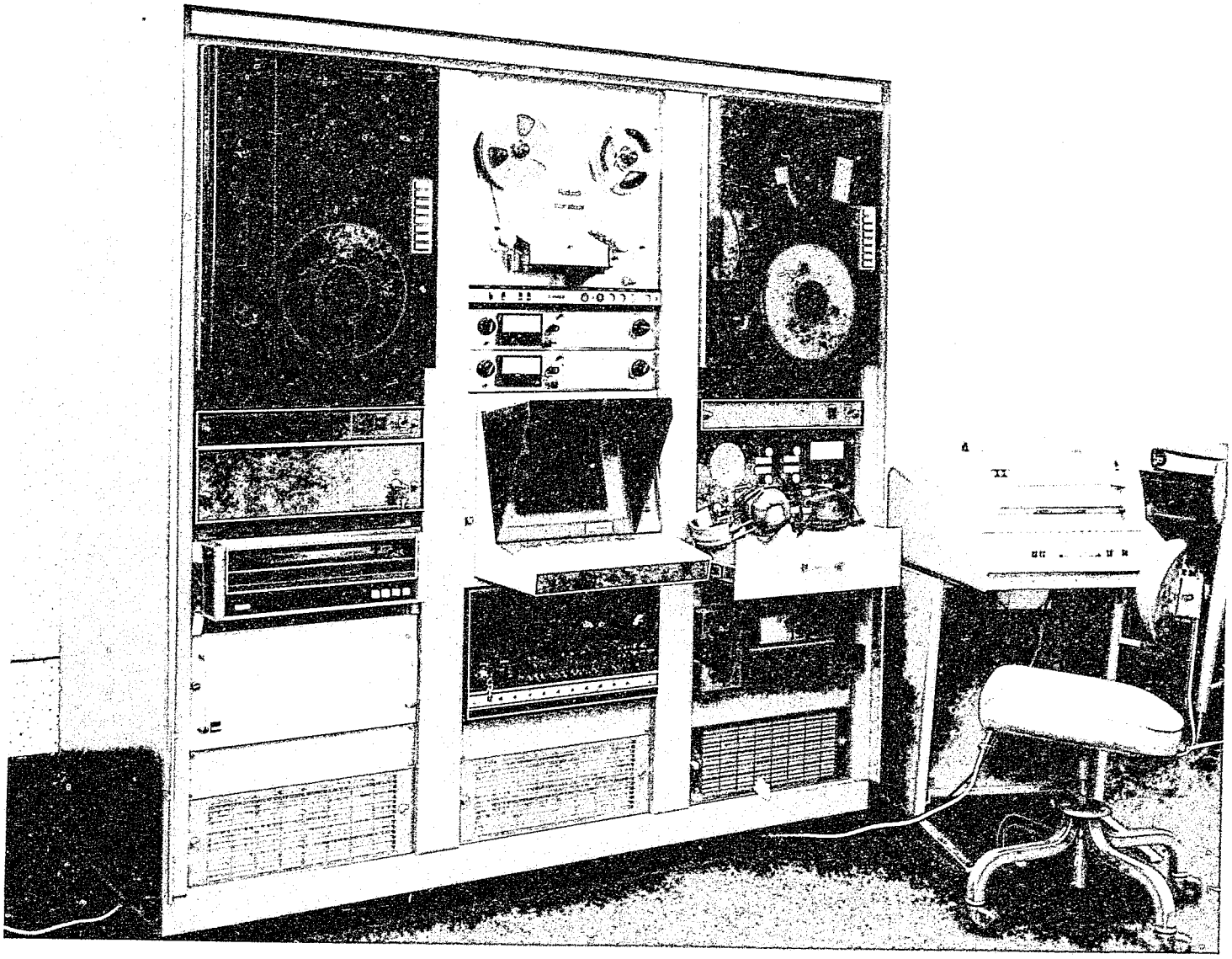


Figure 3. Semi-Automatic Speaker Identification System (SASIS)

This system offers a greater degree of flexibility than any previous approach. It performs comparisons on diverse, frequently occurring phonetic events and will analyze these phonetic events in any of a large number of combinations. This flexibility is expected to provide capabilities for using the system in actual investigative situations.

CHAPTER III. SCOPE

The intent of this applications investigation is to identify problems and deficiencies in law enforcement and criminal justice operations which can be solved or assisted by the science of voice analysis and identification. The impact on these problems of current and projected efforts in speaker identification systems and techniques will be analyzed and quantitative cost-benefit evaluations will be made.

Initial efforts regarding this investigation have been completed. The preliminary portion of the investigation involved surveys of police and criminalistics agencies to identify problem areas and to assess the magnitude of the problems. The surveys and interviews have provided a preliminary set of data regarding the uses and limitations of current identification techniques.

Early estimates have also been made as to the potential uses of improved voice analysis and identification equipment in investigative operations and in providing courtroom evidence. Finally, alternative approaches to utilizing the computer-aided voice identification equipment have been defined and preliminary evaluations have been made.

The results of these initial efforts are presented in this report. It is planned to continue this survey and analysis task into the next phase of the program, at the completion of which a final applications assessment report will be prepared.

The next phase of the applications investigation will incorporate the technical results of The Aerospace Corporation and subcontractor efforts conducted in FY 74 to extend and refine the findings of the preliminary investigation. These results will be evaluated to assess their application to the general problem of identification, in addition to the specific investigative and forensic areas. This general area may include voice identification to improve: computer access security, area access security, identity verification for check and credit card usage, and remote identity verification by police in the field.

The cost-benefit relationships of each concept to each application will be quantified and the most cost-effective approaches to system development will be identified. Finally, alternative methods for concept implementation will be postulated and recommendations made as to the most advantageous strategy to pursue.

CHAPTER IV. CURRENT APPLICATIONS

A. Usage

Although the number of reported criminal cases in which voiceprints have been used to make an identification for a trial is still limited, the voiceprint technique has been used extensively in the past six years in the investigative process to identify or eliminate criminal suspects.[6]

A preliminary survey of practicing voiceprint examiners was performed to obtain estimates regarding the current use of this technique. The organizations contacted were the following:

- Identification Services of the Bureau of Alcohol, Tobacco and Firearms
- Dade County, Florida, Sheriff's Identification Bureau
- Michigan Department of State Police
- Detroit Police Department
- Voice Identification Inc., Summerville, New Jersey

These organizations process approximately one-third of the voice identification cases currently occurring in this country.

Table 1 shows the average current caseload for each organization.

Table 1. Voiceprint Examination Caseloads

Organization	Caseload
Bureau of Alcohol, Tobacco & Firearms	140/yr
Dade County Sheriff	30/yr
Michigan State Police	200/yr
Detroit Police Dept.	40/yr
Voice Identification, Inc.	35/yr
Total	445/yr

From their total caseload, an overall nationwide caseload of approximately 1500 cases per year may be estimated. Typically, these cases break down as follows:

Lewd or harassing calls	35%	Murder	7%
False alarms	20%	Narcotics	3%
Bomb threats	10%	Miscellaneous	15%
Extortion	10%		

The miscellaneous category includes: kidnapping, rape, burglary, arson, robbery, gambling and revoked confessions.

Relatively few of the cases result in courtroom testimony whereby voiceprint evidence is presented during a trial. In the past three years, about 25 trial cases each year have involved voiceprint evidence. Of those cases that have been brought to court, approximately 50 percent have resulted in convictions, and 50 percent in acquittals, dismissals, suppressions and withdrawals. In almost all of the convictions, there were no expert witnesses called by the defense to challenge the voiceprint evidence. These results imply the inherent weakness involved in the use of voiceprints for forensic applications and their vulnerability to challenge from knowledgeable witnesses.

The use of voiceprint evidence in a particular case in court is affected by a number of considerations. Among these are: the relative seriousness of the case, the quality of the evidence (recorded voice samples), the amount and quality of other evidence, and the availability of a qualified examiner to testify. Other factors are the problems related to the admissibility of such evidence and the expense involved in its preparation.

B. Reliability

A study of court records cannot be used to estimate reliability since it is not known whether a person convicted on voiceprint evidence was actually guilty or whether a person acquitted was actually innocent. These factors make it impossible to infer "error rates" for the technique on the basis of judicial records. Lack of data regarding the reliability of the technique has been one of the chief reasons for the criticism of voiceprint identification by the scientific community.

The experiments of Dr. O. Tosi mentioned in Chapter II were aimed at developing quantitative measures of the reliability of voiceprint identification. His experiments examined the effects of five variables:

- The number of speakers in the set
- Open vs. closed tests*
- Context of speech in sentences or isolated words
- Effects of noise and distortion
- Duration of time between samples

Errors of less than one percent were obtained for closed tests with words spoken in isolation and with contemporary voice samples.† Test conditions more relevant to forensic applications involve noncontemporary voice

*In a closed test, the examiner knows, a priori, that the unknown speaker is contained in the test set. In an open test, this constraint is not imposed.

†Contemporary voice samples are made very close together in time. Noncontemporary samples are made several days or weeks apart.

samples, open tests, and words spoken in various sentence contexts. The experiments showed that the reliability of speaker identification varied according to the particular conditions included in these tests with a range of errors from 0.9 to 29.1 percent. For tests in which all unknown spectrograms were contemporary, the rate of false identification ranged from 2 to 4.5 percent depending upon the number of voices in the known set. When the unknown spectrograms were noncontemporary, this error rate increased to between 4.9 and 9.8 percent.

The experiments also showed increased error rates when the context of the test words changes from words in isolation (7 percent error) to words imbedded in random sentence contexts (16 percent error). Not examined in the study were other factors which tend to be present in forensic situations. These include changes in the emotional state of the speaker, intentional mimicking or disguise, background noise and distortion caused by telephone channels and recording devices. These factors can be expected to cause significant increases in intraspeaker variability -- and in turn, the probability of error.

One of the major problems in assessing the reliability of voice identification is that there is scant literature available to confirm or reject the fundamental premise upon which the technique is founded, namely: voice uniqueness. Nor is there available at this time any authoritative data on the effect on voice spectrograms of nasal or oral surgical operation, muffling of the voice, mimicking, use of dentures, tooth extractions, as well as, for example, the effects of illness, colds, puberty, external influences, background noise,

emotional state, or indeed the effect on spectrograms of isolated cue words of the preceding and following sounds in a sentence. Such data would appear to be a necessary prerequiste to scientific acceptance based on proven reliability.[7]

C. Admissibility of Voiceprint Evidence

The identification of an individual by his voice, when made aurally and not by voiceprint, has long been held admissible in criminal trials. Aural voice identifications have been held admissible by analogy to identifying techniques which involve bodily or physical examinations. Compelled aural voice identification is not protected by the self-incrimination privilege even where a suspect is required to repeat the same words which a witness has indicated were used by the perpetrator of the crime.[6]

The response of the courts to the use of spectrographic identification has been uneven and contradictory. The New Jersey Supreme Court and the California Court of Appeal have ruled that voiceprint evidence is not admissible due to lack of scientific acceptance while the Minnesota Supreme Court has allowed the technique to be used. One of the most recent rulings has involved the U.S. Court of Appeals in Washington, D.C. In this case, Judge Carl McGowan has ruled that the technique is not sufficiently accepted by the scientific community to form a basis for a jury's determination of guilt or innocence.

The principal objections to the use of voice identification in court proceedings are: the lack of data regarding the invariance of speech, the inadequacy of identification experiments conducted thus far, the contradictions

between the results of experiments conducted by different investigators, and the poor quality of the recordings used in the field for identification purposes.

As far as the criminal recordings are concerned, poor quality can be expected. Although some improvements are possible, the factors affecting the recorded signal quality are, for the most part, uncontrollable. An equally significant problem associated with the use of voiceprints results from poor quality of the exemplar material (i.e., the suspect recordings). These recordings are usually made by police interrogators with no consideration given to the quality of the recorded speech or the presence of background noise, reverberations, etc. To minimize this problem at least one agency has begun to obtain exemplars over the telephone. The situation is often crucial to an identification process, since the law requires a suspect to provide only one exemplar. If the recording process is faulty, this item of evidence could be irretrievably lost.

Spectrographic voice identification offers great hopes as a reliable means of establishing identity, provided the claims by its proponents can be substantiated by reliable, unbiased research of the type called for by its critics. At the present time, admissibility appears to hinge on whether the test meets the "general acceptance" standard for novel scientific methods. This acceptance has not yet occurred either for the principle of voice uniqueness or for the reliability of the art of comparing speech spectrograms.[7]

CHAPTER V. POTENTIAL APPLICATIONS

A. Trends in Voice Identification

The problems associated with voiceprint identification and the objections raised regarding its application make it unlikely that it will be a widely used, universally accepted technique for obtaining courtroom evidence in the foreseeable future.[8] While its popularity may grow, particularly as an investigative tool, it will eventually be replaced or supplemented by better, more objective methods. These methods will probably include a computer-based automatic or semiautomatic system for processing and comparing voice information.

The computer system approach to speaker identification has two distinct advantages over the voiceprint examination technique. First, in programming the computer to make the identification, the expert is forced to clearly define what he considers to be significant similarities between the voices. This definition and analysis will ultimately lead to an examination of the basic hypothesis of voice uniqueness. Second, once the computer is programmed the actual identification will be made on a consistent objective basis. Table 2 summarizes the basic functional differences between the two approaches.

The computer-assisted speaker identification technique now under development resolves the voice samples to be compared into a set of well-defined features or parameters. A given feature may be a power spectral density component at a specific frequency, the frequency of a particular formant, or an algebraic combination of two or more physical attributes of the speech sample.

Table 2. Functional Comparison of Voiceprint Examination and Computer-Assisted Speaker Identification Techniques

Voiceprint Examination	Computer-Assisted Technique
1. Decisions are subjective and dependent upon the individual examiner's expertness.	1. Decisions are objective - repeatable results may be obtained with different examiners.
2. Decisions are based upon loosely defined points of similarity.	2. Measures are made of differences between well-defined features of each voice sample.
3. Comparisons are qualitative in nature.	3. Quantitative measurements form the basis of comparisons.
4. Effects of distortion, disguise, etc., are difficult to assess.	4. Effects can be measured and the degree of confidence in the decision can be adjusted quantitatively.

The features are derived by first dividing each of the two speech samples into its constituent phonetic events and then extracting feature sets for selected pairs of equivalent phonetic events from the samples.

In the speech comparison process, the operational sequence is reversed. For each pair of phonetic events, the equivalent features are compared and a measure is computed for the separation between pairs of features. The particular magnitude of this separation is termed the "distance" between the features. The set of distance magnitudes for a pair of phonetic events forms a measure of the distance between the two phonetic events. There will be such a set for each of the pairs of events used in the samples. Finally the measures

for the pairs of events are combined to provide a measure of the distance between samples. This distance between samples is used to estimate the similarity between the voices on the two speech samples.

A laboratory prototype system was developed and system tests were run on a single utterance in a fixed context.[4] The test results are shown graphically in Figure 4. The figure shows for any given degree of distance (i.e., any given point on the X-axis) the proportion of pairs of utterances from different speakers which were at least that close and the proportion of pairs of utterances from the same speaker which were farther apart than the given distance. For example, if a distance of 20 is selected as a threshold to decide if two speakers are the same or different, then two different speakers will be judged the same only two percent of the time (false accusation), while a speaker match will be incorrectly rejected approximately 20 percent of the time (false dismissal). If two thresholds were drawn such that no decision would be made approximately 30 percent of the time, the error rate would be less than one percent when a decision was made.

The prototype Semi-Automatic Speaker Identification System now under development extends the concepts utilized in the laboratory prototype system so that the system can be used in a wide variety of investigative and forensic situations. Figure 5 illustrates the overall operation of the system. Criminal speech samples from police station monitors, covert recordings or authorized wire taps are processed and stored on digital magnetic tape. In the processing operation, specific phonetic events which are known to have a high degree of discriminating power are identified and labeled. When a suspect sample is

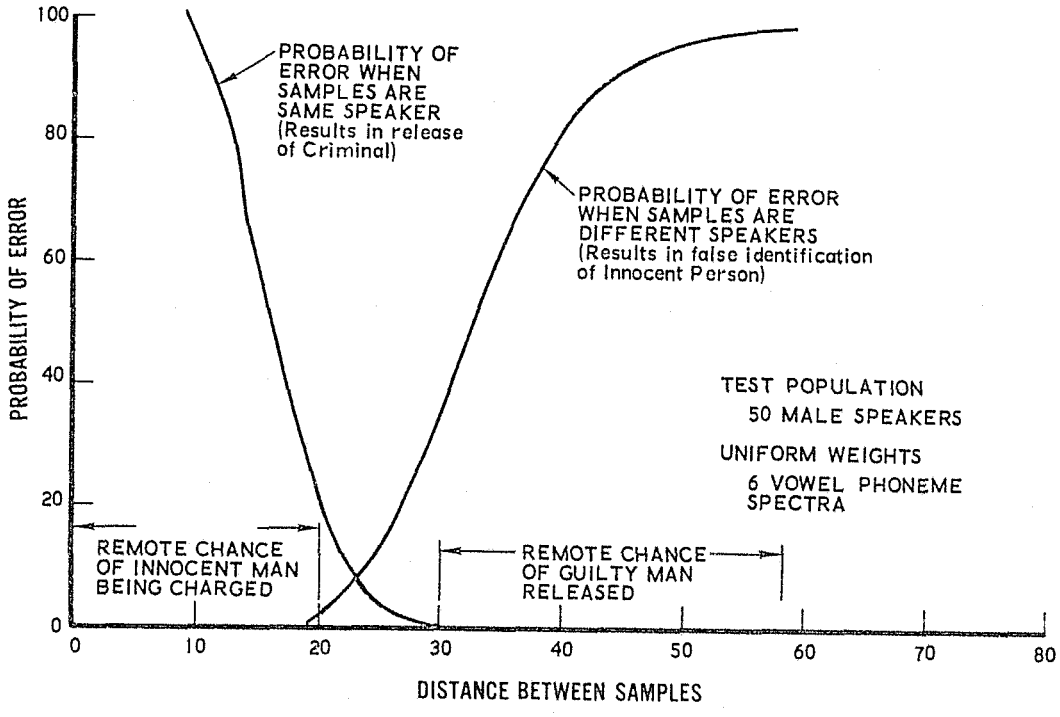


Figure 4. Error Rates as a Function of Distance Between Speech Samples

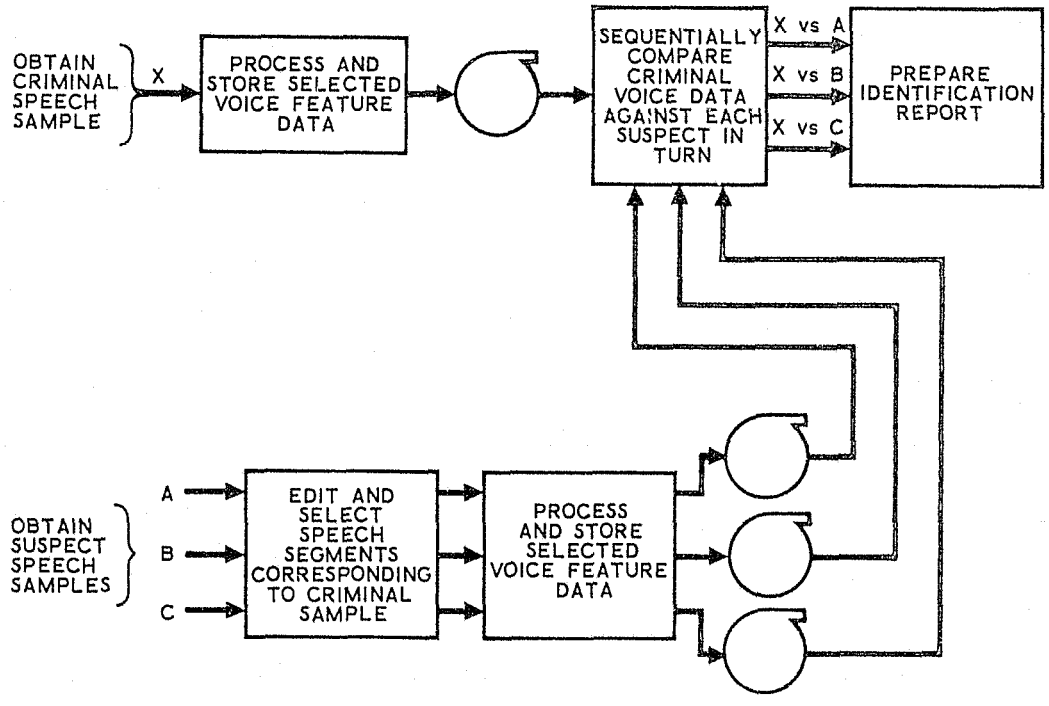


Figure 5. Semi-Automatic Speaker Identification System Operation

obtained, the same phonetic events are selected for processing. In the comparison phase each selected event from the criminal sample is compared with a like event from a suspect sample. The points of comparison are well defined and yield quantitative results. The system is thus able to generate accurate and objective results on a repeatable basis.

The detailed comparison process is shown diagramatically in Figure 6. Each speech sample goes through the same series of steps whereby the sample is digitized, sonograms and other displays are generated, and phonetic events are selected. For speech sample of suspect A the selected events could be designated 1A, 2A, 3A... Each event is further broken down into a set of parameters or features. Each event will produce a set of approximately 30 features. The features used will be those which exhibit the best discriminating power for that event. Event 1A will thus produce features 1A1, 1A2, 1A3..... 1AK, as shown. Event 1B from speech sample B will likewise produce feature set 1B1, 1B2, 1B3..... 1BK. The two feature sets are combined in a manner which produces a distance measure set. The manner in which the distance is derived is such that the widest separation between different speakers is achieved while maintaining the smallest distance between different utterances by the same speaker.

A distance measure is obtained for each phonetic event in the sample. Only like events are compared. Finally the various distance measures are combined to arrive at an overall similarity measure for the two samples. As before, the method of combination is selected to maximize the system's

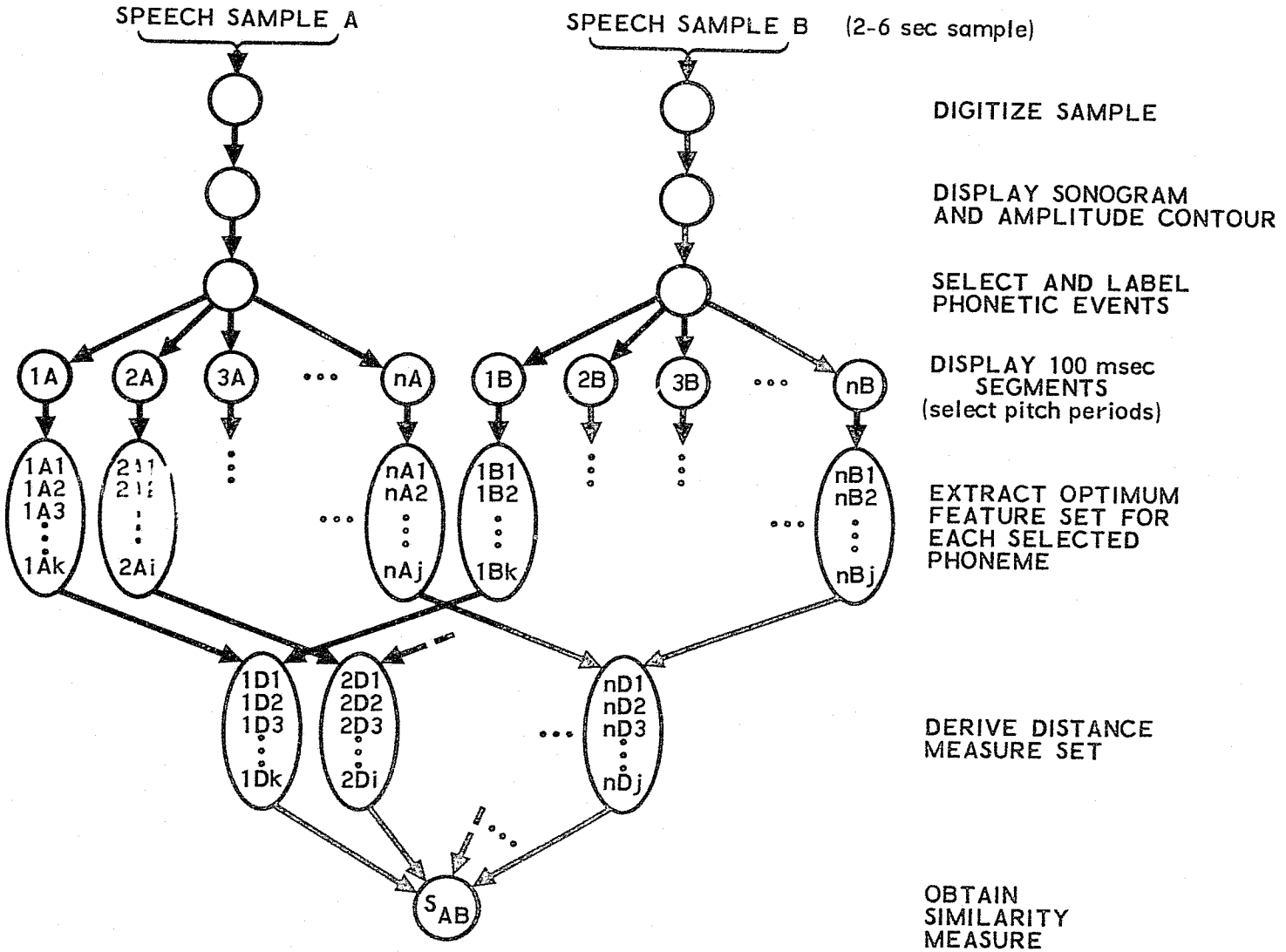


Figure 6. Semi-Automatic Speaker Identification System Identification Process

speaker discrimination capabilities. It is expected that, when fully implemented, this technique will overcome the objections which currently hamper the use of voice identification.

B. Applications Survey

As previously mentioned, there are about 1500 criminal cases occurring each year in which voice identification via voiceprints plays a part. This low figure appears due more to the limited availability of qualified examiners and facilities than to a lack of case material. In an effort to obtain an estimate of the potential usage of voice identification technology, a survey was conducted of police investigators in five cities. The cities were:

- Washington, D.C.
- New Orleans, Louisiana
- St. Louis, Missouri
- Tulsa, Oklahoma
- Santa Ana, California

The aggregate crime rate for each of these cities, as listed in the 1972 Uniform Crime Reports, is shown in Table 3. Also listed is the overall crime rate for the five cities taken as a whole. It is seen that this rate is very close to the overall crime rate for all U.S. cities over 50,000 in population. These 5085 cities comprise the bulk of the U.S. urban population.

C. Study Method

Several law enforcement agencies were contacted by telephone to develop a gross list of potential speaker identification system applications

Table 3. Crime Statistics for Selected Cities

City	Population	Total Crime	Rate/1000 Pop.
Washington	750,000	37,446	50
New Orleans	650,000	30,000	46
St. Louis	600,000	42,580	71
Tulsa	400,000	12,611	31
Santa Ana	172,000	7,928	42
Total	2,572,000	130,565	50.5
5085 cities over 50,000 population	124,592,000	6,324,982	50.8

(shown in Table 4). This list was reviewed in detail with selected members of the law enforcement community to arrive at six specific crime types for the survey:

- Disturbing telephone calls
- Bomb threats
- Kidnapping
- Narcotics
- Gambling
- Prostitution

Data collection forms were designed to facilitate accumulation of information from which to determine, for each crime type for the year 1973:

- Number of cases handled
- Number of cases where speaker identification techniques could have been applied

Table 4. Gross List of Potential Speaker Identification Applications in Local Law Enforcement Agencies

MAJOR USE: Personal Identification Verification Through Voice Comparison

- Narcotics Stake Out
- Narcotics Buy
- Narcotics Soliciting Buy
- Prostitution Stake Out
- Prostitution Buy
- Prostitution Soliciting Buy
- Burglary Stake Out
- Armed Robbery Stake Out
- Assault Stake Out
- Kidnap Threat Monitor
- Kidnap Payoff
- Murder Contract Buy
- Murder Contract Monitor
- Vandalism Stake Out
- Vandalism Buy
- Secure Area Access
  - General Area
  - Police Computer Room (Records, etc.)
  - Communications Room (Command, Control, etc.)
  - Confession
  - Interview Interrogation
  - Bomb Threats
  - Bomb Extortion Payoff
  - Lewd/Obscene Telephone Calls
  - Suspicious Telephone Calls
  - Harassment Telephone Calls
  - Merchandise Fence Monitor
  - Merchandise Fence Payment
  - Extortion By Wire
  - Extortion Payoff
  - False Fire Alarm
  - Arson Threats
  - Arson Payoff Monitor
  - False Police Alarm
  - Riot/Insurrection Monitor
  - General Investigative Tool
  - Court Authorized Wire Taps
  - Court Authorized Eavesdropping
  - Attorney of Record Verification
  - Insurance Investigator/Other Verification
  - Control of Inmate Egress/Ingress - Jail/Other

- Number of cases where speaker identification could have been a crime deterrent
- Number of investigative manhours which could have been saved by employing speaker identification techniques
- Number of confessions which speaker identification techniques could have encouraged
- Number of court convictions which speaker identification techniques could have encouraged.

Five different local law enforcement agencies were selected for the survey. Eight functional areas in each police department were visited: Management (Chief of Police or representative), Records Division, and the heads of each of the six crime-type investigation units.

D. Summary of Relevant Comments

1. Washington, D.C. Police Department

a. Bomb threats. There were almost 600 bomb threat cases reported in 1973. In some of the cases (about 20), the person who placed the bomb also called the police department with the threat. In these instances the criminal sample was readily available since all calls into the department are recorded. Many of the other cases involved threats to banks. A capability to record these calls is either already there or would be easy to provide.

b. Kidnapping. All cases encountered in 1973 were disputes regarding custody of children between divorced or separated parents. While there were no kidnap for ransom cases, this type of crime as well as

political kidnapping is potentially very likely for Washington. Speaker identification techniques would have extensive usage in such investigations.

c. Narcotics. Criminals in the narcotics trade make great use of code words and names. Recordings of conversations are readily available since undercover investigators use concealed voice recorders and transmitters in present operations. The ability to subject these recordings to analysis to identify speakers would aid investigators in matching users and dealers with previously collected voice samples. A library of voice samples (similar to a latent print file) would enable investigators to break codes and gain information regarding the organizational structure and communication channels of the narcotics traffic.

d. Disturbing telephone calls. There is no way to accurately estimate the number of complaints regarding disturbing calls, since no records are kept. Despite the distress endured by the victims of these calls, they are a low priority item because of the extreme difficulty of investigation. The use of an effective means of identifying speakers from voice samples would allow the department to establish a program to pursue these crimes.

e. Gambling. The major gambling activities, such as the numbers lottery, are highly communications dependent. Currently, convictions are difficult to obtain due to lack of evidence. A library of voice samples would be useful in investigations of gambling rings.

f. Prostitution. Prostitution and other vice operations are often linked to narcotics activities. Vice arrests frequently lead to breaks in narcotics investigations. Voice recordings are readily available from covert recorders carried by undercover investigators.

2. New Orleans, Louisiana Police Department

a. Bomb threats. Currently, there is a very small chance of solving crimes of this type. While speaker identification techniques would improve investigative opportunities for solving these cases, the greatest effect would be to reduce the incidence of these crimes.

b. Kidnapping. Comments generally similar to Washington Police Department.

c. Narcotics. The use of speaker identification techniques would greatly depend upon the ability of the equipment to overcome intentional voice disguise by criminals. If this problem were solved, the system would be invaluable in tracing the organizational structure and logistics of narcotics rings.

d. Disturbing telephone calls. Obscene and threatening calls often involve repeated calls by the offender. The ability of investigators to identify offenders from voice samples would increase chances of apprehension dramatically. Even more significant would be the deterrent effect.

e. Gambling. A major problem in gambling investigations is the difficulty in connecting recorded voice samples with apprehended suspects. The use of speaker identification techniques would greatly reduce this problem. It would also be possible to relate the various gambling operations through coordinated use of speech identification equipment.

f. Prostitution. Comments generally similar to Washington Police Department.

3. St. Louis, Missouri Police Department

a. Bomb threats. Bomb threat investigations require very positive identification. The use of speaker identification techniques with other identification methods would increase the accuracy and credibility of investigations.

b. Kidnapping. Comments generally similar to Washington Police Department.

c. Narcotics. The use of speaker identification would increase the usefulness of covert recordings by an order of magnitude. This would be especially useful in cases where dealers or pushers are involved. Since 80 to 90 percent of narcotics addicts are also involved in burglaries, robberies and thefts, application of better identification techniques to reduce or deter narcotics activities would also reduce or deter these other crimes.

d. Disturbing telephone calls. Comments generally similar to Washington Police Department.

e. Gambling. Bookmaking is almost entirely dependent upon telephone communications. Speaker identification techniques could be applied when bookmaker's telephones are tapped or when an undercover agent places a bet and uses a concealed recorder. A fully implemented speaker identification capability would virtually eliminate bookmaking in this city.

f. Prostitution. Comments generally similar to Washington Police Department.

4. Tulsa, Oklahoma Police Department

a. Bomb threats. Most bomb threats are either to High Schools or Junior High Schools. If an effective speaker identification system were available, the police department would equip all schools with voice-actuated recorders to obtain criminal voice samples. Such recorders would also aid in vandalism investigations by recording voices in schoolrooms after hours.

b. Kidnapping. Comments generally similar to Washington Police Department.

c. Narcotics. A speaker identification system would be used both in undercover buys and telephone-arranged buys. Currently many undercover buys are from persons unknown to the police. Speech analysis would aid in later investigation and would assist in connecting seemingly unrelated cases through common parties. A noticeable reduction in the number of cases would be expected as well as a significant reduction in the total narcotics traffic volume.

d. Disturbing telephone calls. If a speaker identification system were available for use, the procedure would be for the Police Department to loan and connect a voice recorder to the telephone of the person receiving the repeat disturbing telephone calls. This would probably include all lewd calls, all threatening calls, and some nuisance calls. Since the greatest number of calls are of the lewd type, and since these callers tend to be calling several people over a short period of time, speaker identification from voice samples would aid in solving multiple cases involving the same caller.

e. Gambling. The vast majority of gambling cases are for bookmaking. This is a telephone-communications oriented crime; therefore,

speaker identification could be used with virtually every case. In addition, due to the acquisition of additional information, it would be possible to uncover 50 percent more new cases.

f. Prostitution. Comments generally similar to Washington Police Department.

5. Santa Ana, California Police Department

a. Bomb threats. Ninety percent of bomb threats are to schools. Comments are similar to Tulsa Police Department. Other threats are from hard core criminals who are bomb threat repeaters. A library of voices would assist in these cases if banks, stores, etc., had a capability for recording voices during the call.

b. Kidnapping. Comments generally similar to Washington Police Department.

c. Narcotics. Most suspects are especially difficult to identify and compare to existing known offenders. Speaker identification would provide a much faster method for identification of suspects. It would also provide an additional investigative tool. This would be particularly helpful in narcotics cases because it is best to have several independent methods for investigations.

d. Disturbing telephone calls. Comments generally similar to Washington Police Department.

e. Gambling. Current policy is to enforce gambling statutes on a complaint basis only. Speaker identification techniques would enable investigation on a more systematic basis.

f. Prostitution. Comments similar to Washington Police Department.

E. Quantitative Results

Table 5 presents a summary of the estimated caseload for a potential speaker identification system obtained from representatives of the law enforcement agencies in the five cities in the survey. Members of Management (Chief of Police or staff), Records Division, and heads of each of the six crime-type investigative units were asked to estimate the number of cases in which use of voice identification could have a significant effect.

Table 5. Summary of Applicable Cases for 1973

City / Crime Type	Washington	New Orleans	St. Louis	Tulsa	Santa Ana
Bomb Threats	75	210	25	60	7
Kidnapping	64	1	-	-	-
Narcotics	350	1300	700	400	750
Disturbing Tel. Calls	-	171	-	432	240
Gambling	1000	270	350	63	10
Prostitution	550	100	350	35	100
Total	2039	2052	1425	990	1107

From the standpoint of a speaker identification system, each case would consist of an examiner receiving two or more speech samples, one of which is made by a criminal. The examiner would be asked to render an opinion on

the likelihood that the criminal sample and a given suspect sample were made by the same person. It was estimated -- based on the time required to initialize the system, to input data, to perform data processing and comparison operations, and to output the required data and summary reports -- that an examiner using the computer-aided system will require between one and two hours for each case. Since part of this time will consist of the examiner waiting for the computer to run through its computations, an efficient examiner will be able to perform other work for a portion of the time. As an estimated average, a figure of 1.5 hours per case appears reasonable at this time.

A one-shift operation on the computer-aided system as presently configured would handle approximately 1330 cases per year. This would increase to 1600, if a six-day week were scheduled, and to 2660 for a two-shift operation. Requirements for the examiner to appear in court to testify could reduce this caseload capability. In any event, the speed of the prototype system appears adequate for any of the cities included in the survey.

1. System costs. The cost of the equipment for the current configuration is approximately $75,000. It is expected that quantity purchases and their associated original equipment manufacturer (OEM) discounts would enable a production system to reduce this cost somewhat. Since the development and software are supported by the Institute, it is expected that a production system, including installation and checkout, could be supplied in the $80,000 to $100,000 range.

If one assumes a ten-year amortization for the equipment and a $1000 to $6000 yearly cost for operational supplies and maintenance service

(depending on the intensity of usage), the effective annual equipment cost will be between $9,000 and $16,000.

The salary expense for speech examiners will vary greatly with the caseload. Table 6 presents one estimate of cost components for various caseload levels.

Table 6. Speaker Identification System Costs per Year

Cases per Year	Hardware Costs	Maintenance and Supplies	Examiner Costs	Total
0 - 100	$8K	$1K	$2K	$11K
100 - 500	$8K	$2K	$10K	$20K
500 - 1000	$9K	$4K	$25K	$38K
1000 - 1500	$10K	$5K	$35K	$50K
1500 - 2500	$10K	$6K	$50K	$66K

The figures are based on an overall examiner cost (including overhead) of $25,000 per year. Five usage categories are rather loosely defined. When the caseload is under 100 per year, the maintenance requirements are minimal. An examiner whose principal duties involve other work, is only called upon when an operation is required. For this level of usage, the cost per case is high, as shown in Figure 7. At this usage level, a department may find it economically advantageous to offer the facility to other agencies on a cost sharing basis.

When the usage increases to between 100 and 500 cases per year, it would be worthwhile to assign a staff member to work half time on the speaker

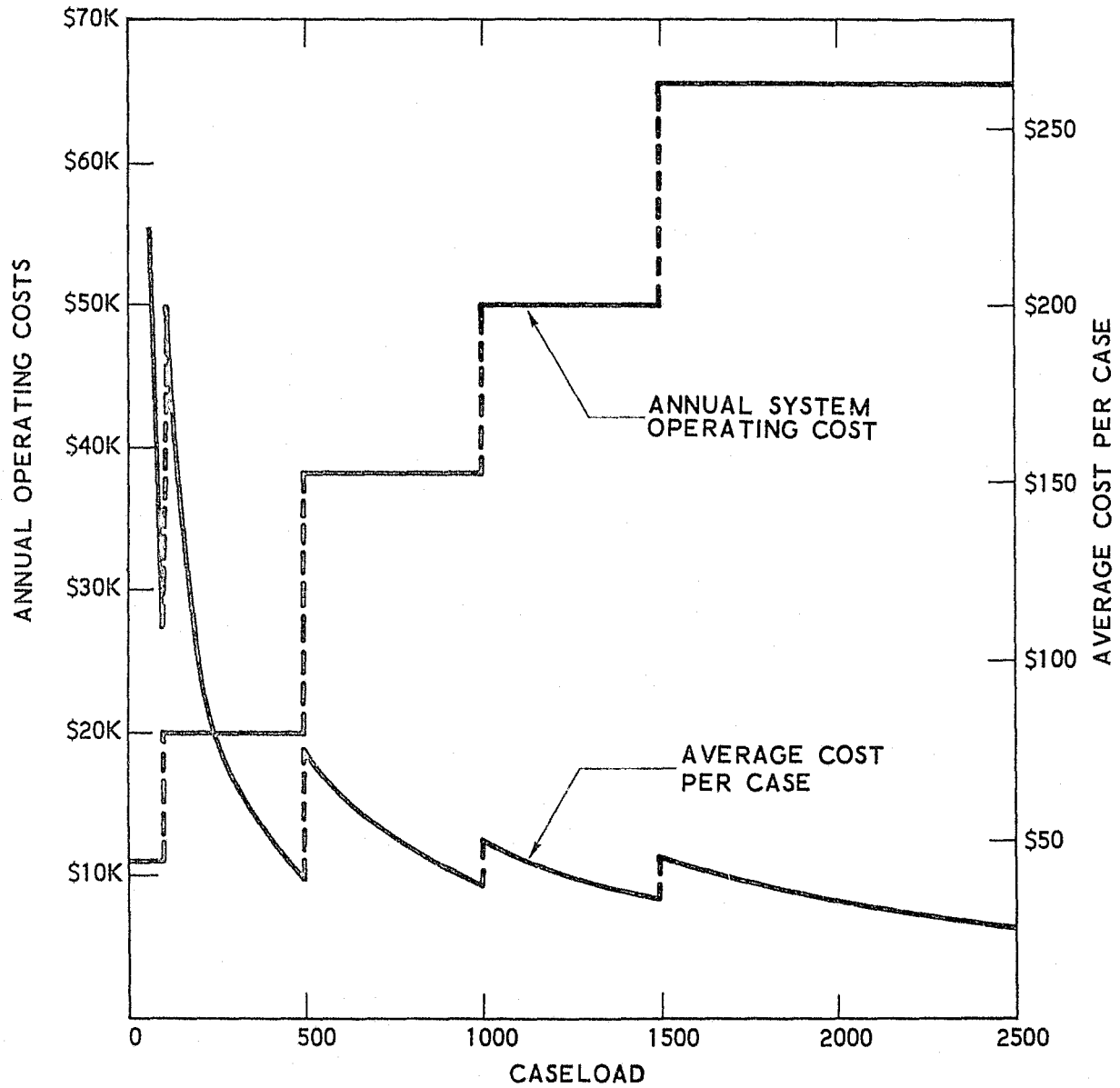


Figure 7. Annual Costs and Costs per Case as a Function of Caseload for Speaker Identification

identification system on a continuing basis. As the caseload increased further (to between 500 and 1000 cases), the examiner would be assigned full time to speaker identification. At this level of operation, the operating expense (maintenance and supplies, etc.) could be expected to increase somewhat. Also, it might be useful to employ a more elaborate hardware configuration in order to increase speed and efficiency (e.g., use of a high-speed printer). The average cost per case could now be expected to drop below $50.

As the caseload increased beyond 1000 per year, the examiner would be required to work overtime, and might employ part-time assistance. A still more elaborate hardware configuration might be desirable. Finally, for caseloads over 1500 cases per year, two examiners working on a two-shift basis would be needed. The average cost per case could be expected to reach the $30 to $40 range. Beyond this point, a more sophisticated system configuration would be considered, using a fast, powerful central computer and a number of time-shared terminals.

2. Estimated effectiveness. The effectiveness of an improved identification investigative tool such as the speaker identification system lies in three principal categories: deterrence, investigation, and judicial.

a. Deterrence. The deterrent effect of speaker identification in the area of illegal telephone calls has been demonstrated in a Michigan program whereby the intent of the police to trace calls and match speech samples was publicized. This program resulted in a 50-percent decline in the incidence of these calls. The police departments surveyed reinforced

this estimated deterrent effect for telephone calls and also emphasized a significant potential for deterrence in the area of narcotic traffic. The widespread use of covert recording by undercover narcotics agents is known to criminals engaged in the drug traffic. If voices taped on these recorders could be used for identification, the detectives interviewed estimated that it would deter approximately ten percent of all narcotics activities.

b. Investigation. The principal impact of using the speaker identification system will be in improving the effectiveness of police investigation techniques in several crime categories. These categories include: bomb threats, false alarms, extortion, receiving stolen goods, gambling and narcotics investigations. In conjunction with other technology, such as voice-actuated recorders, the system could also be used in armed robbery, vandalism and burglary cases. The effects in investigative applications would be due to the improved ability to focus investigations on the most promising suspects. Suspects whose voice samples are dissimilar to the criminal sample can be eliminated from further consideration.

c. Judicial. The applications of the speaker identification system would include: providing investigators with greater weight of evidence (this could be used in court or might induce a confession from a guilty suspect), making it less likely that a suspect would later disavow a recorded confession, and providing a better basis for decisions to prosecute.

The effectiveness of these several applications was quantified when the police investigators estimated the number of manhours that would be saved

by using speaker identification techniques. The manhours were then translated into equivalent costs. No attempt was made to assign cost saving estimates in the deterrence category. Table 7 summarizes the savings in investigation and judicial areas and indicates that a typical large city police department such as Washington, D.C., New Orleans or Tulsa could experience approximate cost savings of $300,000 per year. These savings would be based on the processing of about 2000 cases each year. The cost of such processing, as shown in Figure 7, would be between $60,000 and $70,000 per year. On a simple cost comparison, therefore, the use of such a system would be worthwhile. However, the real benefit would result from increased effectiveness of law enforcement activities and deterrence of crimes which cause fear and anguish to the general public.

Table 7. Summary of Speaker Identification System Savings Benefits by City Police Department

City Police	Investigation		Judicial		Total	
	Man Hours	Dollars[a]	Man Hours	Dollars[a]	Man Hours	Dollars[a]
Washington, D.C.	15,800	237,000	9,300	139,500	25,100	376,500
New Orleans, La.	13,392	200,880	6,696	100,440	20,088	301,320
St. Louis, Mo.	6,000	90,000	2,300	34,500	8,300	124,500
Tulsa, Okla.	12,360	185,400	6,180	92,700	18,540	278,100
Santa Ana, Ca.	3,776	56,640	1,698	25,470	5,474	82,110
Total for 5 Cities	51,328	769,920	26,174	392,610	77,502	1,162,530

[a] Fifteen dollars per hours, estimated average cost.

Note: All manhour figures are based on estimates from the five city police departments for the six crimes of Disturbing Telephone Calls, Bomb Threats, Kidnapping, Narcotics, Gambling, and Prostitution. All figures are for the year 1973.

CHAPTER VI. TECHNICAL REQUIREMENTS

A. Accuracy

Under laboratory conditions, experimental machine identification has achieved a one-percent error rate.[4] To achieve this accuracy decisions were made only when the machine measurements were in the high confidence region; as a consequence decisions were made in only 70 percent of the cases tested. Accuracy of an operational system will be a function of many factors. Among these are: length of samples, quality of samples, effects of natural variation in speech, effects of emotion, effects of disguise, and effects of ethnic and geographical origin. Extensive testing will be required to evaluate the effects of the various factors upon decision accuracy.

B. Repeatability

The effects on system performance of using different examiners must not be appreciable. Since the system is semi-automatic, certain operations and decisions are made by the examiner. These include the editing and selection of phonetic events for comparison, the labeling of these events, and the selection of a steady-state portion of the event waveform for detailed analysis. This last operation requires positioning of the terminal markers (or cursors) so as to isolate three glottal periods.

Different operators may choose different portions of a sample or may select different phonetic events. A given phonetic event may be labeled differently, and a different portion of the waveform could be selected for detailed analysis. Each of these decisions could introduce variation in the

final result. The guidelines and training materials developed for the system must be evaluated to ensure that such variations are minimized.

C. Versatility

The system should accommodate all major dialects and regional accents encountered in normal law enforcement operations. These should include as a minimum: general American English, Urban Black and Chicano, both male and female.

The system should also be able to operate under the recording conditions normally encountered in police work. These include:

- The 24-hour, slow speed tape recording usually made at the police dispatching center. In this case, the call is made from a public or private telephone. Acoustic background noise is present at the caller end of the line as well as at the police end.
- An inexpensive cassette tape recorder supplied to the victim of repeated disturbing telephone calls. The victim's telephone is fitted with an induction pickup device. The audio output of the pickup is recorded. Again, background noise is present at both ends of the line.
- A standard tape recorder of the type used on an automatic telephone answering device. Several kinds of devices are commercially available with different acoustic specifications.
- A tape recorder, used in a police headquarters for suspect interrogation. The acoustic properties of the room, as well as background noise, have been found to have a significant effect on such recordings.

- A concealed tape recorder, used by undercover agents. These units must be very limited in size and, as a consequence, the signal quality is generally poor.
- A concealed transmitter, used by an undercover agent. In this case the signal is received and recorded at a remote location. While the recording equipment may be superior, the problems associated with transmission over a radio channel are included in the final sample. In addition, covert recordings of necessity use small, inefficient microphones in locations and orientation which make reception poor. Considerable signal fading is often encountered.

D. Judicial Acceptance

Acceptance of a scientific method to provide evidence for use in judicial proceedings generally requires prior acceptance by the relevant scientific community. The usual basis for scientific acceptance of any new procedure is an explicit description of experimental methods and results of relevant tests. In the case of a computer-aided speaker identification system, such acceptance becomes a matter of replicable experiments on the machine/examiner combination with suitable definitions regarding the background and training of the examiner. The test procedures should be based upon specifications of the voice features used for identification. It will be important to know how these features are distributed in the population since such distributions will permit an estimate of the size of the population of discriminable voices, and so give an indication of the reliability that would be theoretically attainable in specific situations.

The experiments with trained examiners should be statistically valid models of the practical task. The tests should involve judgments of whether two speakers are identical when one sample is available from each speaker, and when more than one sample is available. It may also be appropriate to perform tests in which the unknown speaker, whose identity is to be determined from a speech sample, may be drawn from a set of known speakers or may not be a member of the set (open vs. closed tests). Test formats should yield information about the probabilities of missed identification, as well as false identification, and the tradeoff between them. Further, information about the effects of the size of population, the nature of the spoken context in both known and unknown samples and the sensitivity to noise, distortion or deliberate attempts to disguise the unknown voice should be considered.

CHAPTER VII. CONCLUSIONS AND RECOMMENDATIONS

In reviewing the results of this preliminary investigation a number of conclusions seem indicated. At this time, these must be regarded as tentative conclusions, since further investigations may indicate that a reassessment is necessary. These tentative conclusions are of interest, however, since they define the environment of the speaker identification field and also provide direction regarding the allocations of resources for future efforts. The major conclusions are:

- Speaker identification using spectrogram examination has become an important tool in criminal investigations. Because of the subjective nature of the examination, however, and because little is known about the effects of noise, distortion and other factors upon the reliability of examiner decisions, voiceprint identification as a forensic tool does not perform satisfactorily. Major limitations on the use of this technique are the lack of skilled, qualified examiners; the high cost and time delays associated with the examination technique; and the lack of objectivity in arriving at decisions.
- Despite these limitations, the use of speech identification will increase as more examiners are trained and as knowledge of the technique becomes widespread. Police investigators are increasingly reliant on more sophisticated investigative techniques, since they feel limited by recent court decisions regarding use of the traditional methods. Members of the Federal

Strike Force in Los Angeles, who are concerned principally with investigations of organized crime, currently utilize some means of voice identification in 50 to 75 percent of their cases.

- The use of voiceprint testimony in judicial proceedings will come under increasing attack as more defense counsels become aware of its lack of general scientific acceptance. Ironically, the existence of more trained experts in voiceprint examination, while increasing the use of the technique, will also provide more defense witnesses to contest its use in court. Most representatives of the judicial system interviewed in this survey echoed the view of U.S. Court of Appeals' Judge McGowan that the technique "is not sufficiently accepted by the scientific community" to be accepted as evidence. They added, however, that they would welcome the use of voice identification when the general scientific acceptance was achieved.

- The management and investigators of the police departments surveyed were highly receptive to the concept of a computer-aided voice identification system. Those familiar with voiceprint techniques readily saw the advantages in being able to precisely define the points of comparison and to obtain quantitative measures of the degree of similarity between samples. On several occasions, the system was compared to the polygraph which, although severely circumscribed in courtroom use, is still an invaluable investigative tool.

- The preliminary estimates made of the potential caseload for a speaker identification system, when compared to the estimated costs associated with its use, indicate that the system is feasible on a simple cost-effectiveness basis. Other intangible benefits, such as crime deterrence and an improved capability for quickly freeing innocent suspects, must also be considered.

In making recommendations regarding the scope and directions for future efforts in the field of speaker identification, one must weigh the large number of areas wherein work is needed against the limited resources available to support this work. The value of further development in each area both in the short and long term must be compared and relative priorities set to ensure that resources are allocated in the best manner.

Among the items where support could be expected to produce useful results are: development and evaluation of the computer-aided speaker identification system; studies to evaluate the performance of voiceprint examiners under field conditions; research to develop techniques for classifying voices for investigative purposes; and research to define the optimum voice features for identification. A serious drawback to the use of voiceprint techniques is the lack of feedback regarding the accuracy of the decisions made by examiners working in the field. The tasks involved in these examinations are highly subjective. A method for conducting periodic evaluations of an examiner's performance may be useful for maintaining a high level of performance. On the basis of priorities, the status of development, the momentum of current efforts, and the potential for maximum long range effect

indicate that emphasis should be placed on continued support of development of the computer-aided speaker identification system. Specific recommendations regarding this development are:

- Complete the development of the prototype Semi-Automatic Speaker Identification System including laboratory testing to evaluate the effectiveness with female speakers, with speakers under stress conditions and with speakers using disguise. The system should also undergo testing to evaluate the effects of channel and recording equipment normally encountered in law enforcement investigations. These tests should be followed by a design optimization effort to maximize system accuracy under the stated conditions of speakers and recordings.
- Conduct a pilot field test of the system in conjunction with user agencies (public and private criminalistics laboratories, police departments, etc.) to determine its accuracy and its usefulness in concert with existing voiceprint examination techniques. This test should cover the general American dialect, as well as other dialects.
- Prepare a detailed plan for a Formal Field Test and Evaluation of the Semi-Automatic Speaker Identification System. This test will be designed to provide the data and experimental evidence necessary to gain the acceptance of the scientific (and subsequently the legal) community for the concept of computer-aided

speaker identification. The test plan will be reviewed by recognized experts in the speech science community to ensure that the formal field testing provides the data and test conditions required.

- Continue analytical activities in support of the test effort and to optimize the system performance. These activities should include: extension of the data bases, evaluation of additional features, development of new or modified distance and similarity measures and improvement of equalization and sound discrimination techniques.
- Complete the survey and analysis of the uses of voice identification techniques in law enforcement. The problems and limitations encountered with current practices should be evaluated and methods for using the computer-aided speaker identification system to alleviate these problems should be developed.

NOTES

1. M. H. L. Hecker, Speaker Recognition - An Interpretive Survey of the Literature, ASHA Monographs Number 16, Washington, D.C. (January 1971).
2. G. Papcun, G. Ducoff, R. D. Smith, Voiceprint Applications Manual, TOR-0073(3654-06)-1, The Aerospace Corporation, El Segundo, Calif. (July 1973).
3. Tosi, O., et al., An Experiment on Voice Identification, Report SHSLR 171, Department of Audiology and Speech Sciences, Michigan State University, East Lansing, Michigan (July 1971).
4. R. W. Becker, et al., A Semiautomatic Speaker Recognition System, Stanford Research Institute Project 1363, Final Report (August 1972).
5. G. D. Hair, T. W. Rekieta, Speaker Identification Research, Texas Instruments, Inc., Final Report (August 1972).
6. Voiceprint Identification, Georgetown Law Journal, Vol. 61, Issue 3 (February 1973).
7. A. A. Moenssens, et al., Scientific Evidence in Criminal Cases, The Foundation Press Inc., Mineola, New York (1973).
8. E. J. Welch, Jr., Voiceprint Identification, TRIAL Magazine (January/February 1973).

END