Proceedings of The Institute of Acoustics

AN UPDATE ON SPEECH DECODING

Patricia Hollien

Institute for Advanced Study of the Communication Processes,
University of Florida, & Forensic Communication Associates,
Gainesville, Florida

INTRODUCTION

A great increase has been observed in the number of crimes committed during the last decade (BJS, 1983). Indeed, the magnitude of cases law enforcement agencies and the courts currently are having deal with is nothing short of alarming -- and delays/backlogs are straining the procedures, personnel and budgets within these sectors. Worse yet, if the projections of crime rate increases during the next five years are to be believed, there does not appear to be any relief in sight. A recent article in Time (1984) sums up this problem:

> "According to the FBI, a crime is committed every two seconds in the U.S. In the past five years, the number of violent crimes has increased by more than 20%. Nearly one-third of all households were victimized by violence or theft in 1982. Yet while the crime situation has worsened, budgetary pressures have caused most large cities to reduce the numbers of their police officers."

Those of us working in the forensic milieu are acutely aware of these increases in crime and are faced with the demand to find improved, faster and more efficient ways to deal with the ever expanding number of problems associated with this situation. In addition, new and advancing technology is proving to be a mixed blessing. It has helped in some sectors; in others it has but added to our difficulties. The problems being encountered in the use of audio (and more recently video) tape recordings in forensic investigations constitute a prime example of this enigma. For example, as few as 7-10 years ago, the tape recordings made by law enforcement personnel for investigational purposes could be viewed as a "high tech" approach to crime -- and a rather exotic way of doing business. Not so today. Technology has advanced so rapidly that tape recorders (from the small to the very small -- and getting smaller still) are now a permanent part of the equipment associated with any law enforcement agency or court. They are a tool which have been adopted so completely, and used in so many and varied ways, members of the cited organizations find they cannot do without them. Yet, a plethoria of problems are associated with this unpresendented increase in the use of tape recorders. What are these problems and what can be done about them?

THE ISSUE

The primary reason to make a tape recording is to provide a permanent and accurate record of a spoken message. If it is to be useful, the speech it contains must be intelligible. In turn, the intelligibility level of any tape recording is effected in two major ways; by system distortions and by speaker distortions. If no evidence of either of these distortions exists on a recording, there is no problem; if they do exist, precedures to enhance the

speech and decode the message must be applied. However,before any attempt can be made to decode speech on a tape recording, these cited distortions must be identified, examined and understood; the speech signal must be enhanced as much as possible. Application of these operations will allow decoders to work with the best available material and assist them in extracting as much information as possible. However, before proceeding to list the available decoding procedures, it would appear advisable to briefly review the types of distortions which may be encountered and how best to deal with them (for a comprehensive critique of system distortions see Hollien and Fitzgerald, 1977; Rothman, 1977.)

1. System distortions. Very few problems develop when recordings are made in the laboratory (studio); that is, under ideal conditions and/or with high quality equipment. Moreover, sufficient time usually is available to laboratory personnel to check the procedures and equipment employed as they are not in a stressful situation. Obviously, these conditions are not present in police work. Here the equipment often is marginal in quality and usually the recording situation is highly stressful. Worst yet, usually there is but one chance to obtain a recording and the recording conditions very rarely approach the ideal. What are some of these distortions?

a. Equipment quality. It is unfortunate but, due to limited budgets, many agencies which use tape recorders are not able to purchase high quality equipment. Therefore, their efforts are impaired even before any recordings are made. For example, a high quality laboratory tape recorder and microphone (operated under reasonably good conditions) will have a frequency response (bandwidth) of between 40 and 12,000 Hz. with little-to-no internal noise to interfere with the speech signal. In our experience, an inexpensive tape recorder -- such as those often used in forensic work, can have a bandwidth that is significantly reduced -- sometimes to even as little as 200 to 2000 Hz. Thus, the quality of the equipment (or lack of it) often distorts the signal by its limited bandwidth and internally generated noise. Of course, and as most of us are aware, the advances being made in the quality of inexpensive tape recorders are remarkable and improvements in quality continue. Nonetheless, these problems still exist.

b. Telephone/transmission lines. As we all know, the frequency band necessary for intelligible speech is within the limits of the bandwidths imposed by telephone lines. However, a passband of 350-3500 Hz does not enhance speech intelligibility. Moreover, other distortions can occur in telephone transmissions; they include fade-outs, "chopped" speech, crosstalk (of all types) and noise.

c. Equipment breakdown. Many problems can be listed under this rubric; they include difficulties such as: 1) battery decay or malfunction, a condition that can cause slowdown in recorder speed (or even variable speed) 2) the "shorting out" of a signal, a condition that can be detremental indeed; its effects depend on the severity and frequency of the interruptions, and 3) power failure and/or current changes, these events can cause severe degradation to the speech message on a recording. Other problems can occur. Occasionally a faulty cassette can result in the recording tape getting stuck, twisted, crimped, broken or stretched or the entire cassette mechanism failing.

d. Noise. Any unwanted sound or sounds that mask and degrade the

speech signal (or message) can be considered to be noise. It can be of a non-speech or aperiodic variety -- such as broad-band, narrow band, steady-state or intermittent noise. Moreover, noise can be periodic in nature; typical noises are: 60 cycle hum, music as well as speech. Friction sources can create noise; they include: wind, automobile motors, fans/blowers, clothing movement. Noise can be intermittment (footsteps, machines of all types), impact (doors closing, gun shots, horns, bells) or just about any sound source that introduces a unwanted signal onto the recordings. It is immaterial if a noise is periodic or aperiodic. If it is a signal which masks the message of interest, it is a forensic noise.

2. Speech Distortions. Speaker distortions have a lesser effect on speech decoding than they do upon speaker identification. However, they are important enough to be reviewed below.

a. Dialects/foreign languages. Sometimes a recording is encountered in which the speaker has a pronounced dialect (including so-called "street talk") or the speech is in a foreign lanugage. Certainly these conditions hamper the normal process of decoding even though the speech can be easily heard. In this case it is necessary to use decoders who are familiar with the particular dialect or language.

b. Print-through/inadequate recording level. A low level recording of the speech signal of interest, often degrades the message severely. The use of quality tape and the rerecording of the material at a high energy level helps reduce this problem. In addition, the immediate dubbing of the recording aids in preserving it and insuring its integrity. Finally, print-through often can be expected if very thin recording tape is used; it is best to use high-quality tape when copies are made.

c. Speech rate. Over the years speakers have been encountered that talk so fast that the difficulties in decoding seem insurmountable (especially if a dialect is also present). This difficulty does not appear to be a severe one -- at least on the surface or until the decoding process is begun. In addition, when several talkers have been recorded and they tend to interrupt each other, or talk in unison, decoding can be an uphill job.

d. Stress/fear/emotions. The speech patterns for a given talker are known to alter dramatically with high stress situations (Hollien 1980). Many of these changes include raised pitch (i.e., elevated fundamental frequency), increased speech intensity, staccatto speech, unintelligibility of phrases or sentences, disconnected speech, stuttering, crying, sobbing, or whispering.

e. Drugs/health states. As is well known, certainly drugs will influence an individual's speech patterns in much the same manner as do stress, fear or other emotional states. The "slurred" speech of the alcoholic is a good example. Moreover, speakers who are under the influence of drugs or experiencing severe health problems often do not talk intelligibly. Good transcripts frequently are difficult to generate from such recordings.

CORRECTION OF SYSTEM DISTORTIONS

In 1977, Hollien and Fitzgerald detailed the various filtering techniques that can be utilized in the enhancement of speech on tape recordings. Accordingly, these techniques will be reviewed only briefly on this paper.

AN UPDATE ON SPEECH DECODING

1. Filtering Techniques.

a. Frequency biasing. Any home music system will include bass and treble controls. Manipulation of these circuits often can serve to mildly enhance speech intelligibility on a tape recording. The process is one where the operator moderately reduces the level of the unwanted signals (noise) which lie above or below the speech band.

b. Notch filters. Spectrum analysis can provide information about a noise source which is producing a relatively narrow band of high energy at or around a specific frequency. Application of a notch filter (at that frequency) reduces the effect of the noise. These filters can be fabricated/purchased for most relevant frequencies and are particularly good at reducing the speech masking effects of 60 Hz or hum (it must be remembered that low frequencies mask high ones, as well as distract).

c. Band-Pass filters. Highly intelligible speech needs only a frequency pass-band of about 200-5000 Hz and good quality speech (for decoding purposes anyway) can be obtained even if the pass-band is reduced to (roughly) about 300-3500 Hz. Filters that pass these speech frequencies but materially reduce higher and lower frequencies are especially useful in the enhancement of speech on tape recordings. They operate to eliminate masking signals without seriously degrading the speech itself.

d. Comb filters. A comb filter consists of a fairly large series of sharply tuned and separately controllable notch or band-pass filters. Troublesome sections on a tape recording can be played through a comb filter a number of times with the operator adjusting the various settings by very small increments each time.

e. Digital filters. The previously cited procedures involved only analog filtering. It is now possible also to apply especially designed (and rather expensive) digital filtering procedures to a tape recording. However, even though they are based on different principles and are effective, the results are similar to those obtained by procedures a-d.

2. Isolation of filters. It is very important to isolate filters from each other and from any other electronic device (such as a tape recorder) as they have a tendency to interact with each other (and with other units). They do so even though the manufacturer's specifications will assure the operator that they do not. As a result, interacting filters produce a variety of unwanted side effects (such as rebound) and serve to degrade, rather than enhance speech intelligibility.

PROCEDURES FOR DECODING

Any attempt to obtain a "quick and dirty" transcript i.e., one as quickly as possible without utilizing a systematic approach to the decoding process will lead only to major errors in the text, or to additional hours spent in attempts to find and correct errors. Indeed, the decoding process itself often can be among the easier operations within the overall task. It is the thorough and systematic preparation for decoding that can make the resulting transcript both accurate and unchallengable. Please note again that this paper is an up date on the decoding process; hence, while appropriate steps are outlined, specific techniques are not detailed.

AN UPDATE ON SPEECH DECODING

1. Prepartion of the enhanced, working copy tape.

a. The log. It is important that an accurate log be kept of all operations. Such a log is invaluable as the project advances or the individual who carries out the processing is required to testify.

b. Dubbing/filtering. First, a high quality copy of the original recording should be made -- and by means of a hard-line link between two laboratory quality, dual-channel, tape recorders. It is important that the original tape is never used in the processing (except for dubbing or tape authentication purposes of course) as accidents will happen even to those who "should know better". The orginial recording should be stored away safely and only the copy used. Any necessary filtering is applied to the copy.

c. Binaural tapes. Basically, we have found that speech intelligibility can be improved when the original signal is split and the filtered product fed to the decoders dominant ear at a relatively high intensity and the unprocessed signal simultaneously fed to the other ear at a level just above the threshold of hearing for speech (Carhart, 1967; and Hollien and Fitzgerald; 1977). This method has become one of our standard procedures and is used for all processed recordings regardless of the problems encountered.

d. Filtering via computer processing. Modern technology now permits the digitizing of speech and the reconstruction of a speaker's mode for purposes of enhanced speech intelligibility. These procedures are expensive, found only a a few laboratories (world wide) and are but marginally effective. These limitations are such that procedures of this type can be considered only in that rare instance where all other approaches have failed and where the case is very important.

e. Tape speed variations. Decoders appear to do a much better job when the basic recording speed is 7.5 ips. It also is more convenient due to the constant rewinding that is necessary (it is possible to find a given place on the recording quite accurately). A reel-to-reel tape recorder (rather than a cassette) also provides ease of operation. Finally, slight variations in tape speed (either with or without vowel formant normalization) often assist the decoder.

f. Earphones. Once the filtered, binaural recording has been made, the decoder is almost ready to begin. However, the one remaining piece of equipment that is necessary for the decoder to obtain/use is a high quality set of earphones with adjustable volume control for each ear/channel. This procedure allows the binaural tape to be played with the maximum efficiency. Earphones such as these are widely available at modest prices.

2. The Decoder. Personnel who are utilized to decode tape recordings should be trained to the task, exhibit good hearing, be familiar with police work and enjoy this type activity. Trained personnel such as phoneticians, linquists or individuals with speech training in their background should be used; however, we realize that this is not always possible or feasible. As an alternative, we suggest anyone that has had training with degraded speech or

difficult transcriptions. As stated, individuals who have experience with law enforcement and/or the courts often make good decoders; however, they need at least some formal training in phonetics as very often information about the manner or place of phoneme or word production -- or the acoustics of speech -- is required If reasonable transcripts are to be developed. Moreover, technical training is desirable if the decoder finds it necessary to carry out electroacoustic analysis of a speech segment in order to discover its nature. Finally, individuals who are blind or partially sighted should be considered as potential decoders (P.A. Hollien, 1983). As is well known, the blind often exhibit other sensory modalities (such as hearing) that are highly developed. All that is necessary is to provide the blind decoder with some basic training in phonetics, experience and a method of recording the transcripts (either a braille typewriter or a second tape recorder.

THE BASIC PROCESS

a. Listening. Once the enhanced/filtered copy of the tape is available, the necessary equipment has been assembled and the decoder (or decoders) is(are) identified, it is possible to begin the task. The first procedure is to listen to the entire tape recording. This technique constitues more than a simple familiarization process. It also permits particularly difficult places to be identified, proper names to be learned and any idiosyncratic features to be noted. The listening process, then, is repeated (any number of times) and the transcript developed.

b. Panels and Specialists. To have one person decode and a second person review and refine this (first) attempt is the procedure most often used. Indeed, a panel of listeners or decoders, assembled for the purpose is even more effective. With a panel of decoders, each phrase of sentence can be finalized by the consensus of the group -- a procedure which is an effective one relative to airplane crash recordings. Finally specialized personnel must be brought in to assist with the decoding task if the recording includes a foreign language (or dialect) -- or if the material is highly technical or contains language not familiar to ordinary decoders. However, specialists need only be used for such extraordinary purposes.

c. Codes. Once the listening procedure is complete and a reasonable transcript made, it is possible to draw up a tentative list of the codes to be used for identifying the talkers, events, etc. For example, it is necessary to identify talker number 1 (or male/female number 1), talker number 2 and so forth. For the sake of consistency, a systematic way of numbering inaudible words also must be structured. One of our systems is to use the following: (2-3 words ?) or (10-14 words ?) rather than simply inserting the word "inaudible." This method can be used for each talker if there are sections where two or three talkers are inaudible. Another way in which the flow of events can be documented is simply by describing the occurance, as follows: */footsteps, door closing, two gun shots, loud thump/. Indeed, any systematic set of codes is acceptable providing the are consistant and that they aid the reader to develop a clearer picture of the events. Occasionally a "Summary of the Events" must be developed and included. This (transcript) cover page, should, contain the following:

AN UPDATE ON SPEECH DECODING

Case name:	=	
Case number:	=	
Case Summary:	=	
Tape I.D.	=	Identification of tape recording.
Male 1:	=	First voice heard on tape or use the name.
Male 2:	=	Second voice heard on tape or use the name.
*/ /	=	Explaination of events.
(2-3 words ?):	=	Approximate number of words inaudible.
(word):	=	Words in parenthesis, not sure.
000 to 782:	=	Approximate tape recorder meter number (for ease in finding particular places on the recording).
Misc. Info.:	=	As needed for a particular recording.
Time:	=	Minutes and seconds, as needed.
Proper names, places:	=	"Sounds" like the spelling given.

At this juncture a "first order transcript" should become available. Of course, it should be typed with plenty of room left for corrections/additions/deletions. It is important to develop such a working copy as soon as possible; it certainly is the most difficult part of the task. However, once it has been completed, a "second level" attempt can be initiated. That is a second decoder now can review the recording using this transcript, and continue the process.

d. Settlement of Disputes. Whenever more than one person is involved in decoding a difficult recording, disputes naturally will arise. Here again, there are options to be considered in dealing most effectively with these differences of opinion. Either the decoding supervisor, or (ideally) a panel of listeners should make the judgement. If, as it happens, the segment simply is unclear, it is necessary to qualify that portion, i.e., put it in parenthesis so as to indicate the uncertainty. A very difficult sentence, then might appear as follows:

145 Male 3: (If we) are (2 words ?) kill (Bor'-chen-ko), */door closes/, (he) can get (Lake-land) in (September)...

e. Proper names. It should be stressed that proper names often are quite difficult to decode. While speech and language are quite redundent -- and markedly affected by coarticulation -- there is little within a proper name to assist the decoder in discovering its characteristics. Of course, the problem is mitigated if the name is repeated throughout the text. However, if often is necessary to obtain a machine processing "assist" if the decoder is to accurately identify severely distorted proper names and other such material.

f. The final transcript. Once all the decoding is complete and all confusions resolved, the entire transcript should be retyped in final form and proofed for errors. The power of the written word is not to be underestimated as a conversation recorded on paper can turn out to be the deciding factor in a case. If on the other hand, many errors (even simple typing errors) can be detected, the transcript is impeached and contributes little. In any case, it is the obligation of a decoding team to provide as accurate a transcript as possible.

AN UPDATE ON SPEECH DECODING

SUMMARY

As many law enforcement groups have discovered -- and to their distress -- only those transcripts of tape recordings that are accurate and reliably developed should be utilized. Decoding is not a simple task, it involves a long and rigorous process. As has been pointed out, efficient decoding demands a good understanding of the law enforcement milieu, the sources of distortion, the acoustics of speech and language, phonetics and the actual decoding process. Further, the decoder must be able to identify problems, enhance the speech on the recording and be adept with the procedures of dubbing, filtering, development of binaural recordings and the decoding technique itself. While the approach allows very little room for error, handled well, a good transcript will stand as a solid piece of evidence. In any case, it can greatly help law enforcement personnel do their job well.

ACKNOWLEDGEMENTS

The author wishes to thank Brian Klepper for his cheerful support and encouragement and and Dr. Howard Rothman, whose pioneering efforts in the field made this article possible. In addition, assistance was provided by the Institute for the Advanced Study of the Communication Processes at the University of Florida, Dr.W.S. Brown, Jr.,Director. Finally I wish to thank my husband, Harry Hollien, for his constant support and editorial assistance and for being such a good friend.

REFERENCES

1. Anonyomous (1984) Time ,February 27, pg.26. B.A. in Blue.
2. Anonyomous (1983) U.S. Bureau of Justice Statistics, NCJ-87068, October. Report to the Nation on Crime and Justice: The Data.
3. Carhart, Raymond (1967) in Sensorineural Hearing Processes and Disorders, (A. B. Graham, Ed.), Little Brown and Co., Boston, 153-168. Binaural Reception of Meaningful Material.
4. Hollien, Harry (1980) Chapter in Forensic Psychology and Psychiatry (F. Wright, C. Bahn and R. Richer, Eds.) New York Academy of Sciences, New York, 47-72. Vocal Indicators of Psychological Stress.
5. Hollien, Harry and Fitzgerald, James T. (1977) Proceedings: 1977 International Conference on Crime Countermeasures, Science and Engineering, Oxford, England. Speech Enhancement Techniques for Crime Lab Use.
6. Hollien, Patricia (1983) Abstracts of the Tenth International Congress of Phonetic Sciences, (A. Cohen and M.P.R. v.d. Broecke Eds.), Foris Publications Dordrecht, Holland/Cinnaminson, U.S.A. pg. 532. Utilization of Blind Decoders in Phonetics.
7. Rothman, Howard B. (1977) Proceedings, 1977 Carnahan Conference on Crime Countermeasures, Lexington, KY, 63-67. Decoding Speech from Tape Recordings.