Penelope O. Pickett,[1] *M.A., M.F.S.*

Transcripts for Law Enforcement: Special Requirements

ABSTRACT: Problems for law enforcement, both in the examination of taped evidence and in the presentation of evidential transcripts in the courtroom, have arisen as a direct result of misconceptions of what constitutes a transcript of a voice recording. This paper introduces a variety of transcript types, pointing out that transcripts for use by the law enforcement community have special requirements that need to be recognized, that are not being met by transcript types presently in use, and that would distinguish this type of transcript from other types. To assist the law enforcement community in meeting these special requirements, a unique format for transcribing voice recordings is presented, with accompanying guidelines.

KEYWORDS: jurisprudence, evidence, law enforcement, transcripts, voice recordings, forensic linguistics, voiceprints

While it may be true that "a rose is a rose," it is not true that "a transcript is a transcript." Increased judicial system use of voice-recording transcripts is highlighting the growing need to address the special requirements of transcripts for law enforcement. Just as other user groups have developed their own types of voice-recording transcripts, based on their special needs, so, too, must the law enforcement community.

Law enforcement use involves what we might call evidential transcripts, that is, those that are offered as evidence in court, as a guide to listening to taped evidence in court, as evidence to be examined in the forensic science laboratory, or as a guide to the examination of evidence in the forensic science laboratory. The view that the law enforcement type of transcript is "just another transcript" gives rise to problems, to controversy, and to opposition in court.

For example, a typical opposition statement occurred in a trial in Texas, when an expert witness for the defense commented on the accuracy of a transcript introduced into evidence by the prosecution: "[There were] one hundred and twenty-five differences between what I heard on the tape and the transcript" [1]. Unfortunately, similar statements occur far too frequently, indicating not that the transcribers are inept, but rather that the task of producing an evidential transcript is misunderstood, as is the concept of what constitutes a transcript.

This paper attempts to expand the common concept of "transcript," to show that the special requirements of law enforcement transcripts are not being met, and to entice the law enforcement community to use a uniform transcript format, such as the one presented here.

The World of Transcripts

It is commonly thought that the basic requirement of all voice transcripts is the same: *verbatim* representation on paper of what was said on the recording. Strangely, not only is this not the case, but also, most transcript users do not wish it to be the case. A quick look into the world of transcripts will provide some insight.

In a broad sense, the most familiar type of "transcript" is the play script studied in school literature classes. Years after their schooling, transcribers naturally key into this memory base to retrieve general guidelines for formatting. Play scripts, however, and other written representations of spoken language, such as in novels, are not true to natural spoken language. They are formatted, rather, for ease in reading and so omit much of the duplication and other "extraneous" material found in natural language.

Another familiar type of transcript is that used by the business community and other office environments. The boss uses a dictaphone to record a letter, or other document, and the secretary transcribes the recorded message onto paper. They both know that the expected transcription format is a typed letter, for example, with correct grammar and complete thoughts. This most often does not reflect exactly what was dictated, nor does the boss really want it to. The secretary must eliminate false starts, hesitation sounds, instructions, and other extraneous material from the dictation and must correct any grammatical errors occurring in the dictation. This type of transcript captures the content of the message, not the exact dictation.

Persons interested in the study of effective communication, and thus in analyzing interviews or conversations, have developed a third type of transcript. From recorded conversations, they produce written transcripts, using a play script format, but in a column style down the page, with another column for adding commentary alongside. For ease in handling, there is a tendency to join split sentences and to represent questions and answers each as complete, uninterrupted entities, with no interference of simultaneous utterances. The focus of this type of transcript, again, is on content [2].

A fourth type of transcript is that used by historians to document "oral history interviews" of important persons [3]. The historian's goal is to capture not only the content of the speaker's message, but also the flavor of speech and accent of the particular person interviewed, while at the same time not necessarily including all false starts, hesitation marks, or grammatical errors. To achieve the ideal transcript, from the historian's viewpoint, one "must learn to retain the essential and desirable qualities of spoken language and refine or delete less essential ones while shaping the interview into a different form, the written transcript" [4].

For these users of transcripts of spoken language, an exact rendering of the language would be cumbersome and even detrimental to the purpose of the transcript. In the above examples, the transcriber is required to do some editing. This becomes an important factor when the secretary in a law enforcement organization, who is used to the boss-secretary type of transcribing, or has only a play- and novel-reading background to draw upon, is tasked to produce an evidential transcript from a voice recording.

Evidential Transcripts

The law enforcement evidential transcript has a different focus from the four types of transcripts described above, with different formatting requirements. The focus is on the surface structure of the underlying content, that is, on the words, sounds, and pauses, and how they are combined. The forensic linguistic examiner and the spectrographic (voiceprint) examiner, as well as others dealing with evidential transcripts in courts of law, require a *precise* written rendering of the words in question, as well as a close indication of the timing of intersecting sounds. Transcripts need to accommodate these requirements.

For the forensic linguistic (including psycholinguistic) examiner of taped evidence, the

focus is on language usage. Hesitation sounds, for example, can represent an individual's speech habit and thus aid in matching speakers from two different recordings, or in providing a clue to demographic information about a speaker, or a clue as to the speaker's determination, the validity of a threat or statement, and the psychological state of mind of the speaker. The forensic linguist focuses on the minutiae of the language usage—exactly which words or sounds were uttered, exactly how they sounded, and exactly how they fit together in the structure of the conversation or monologue. The forensic linguist looks for patterning and other clues in the language. Editing and errors in a transcript tend to contaminate the examination results when the transcript only, and not the tape, is submitted for examination. Imprecise transcripts also tend to misguide the examiner and cause time delays when both the tape and transcript are submitted.

For the spectrographic (voiceprint) examiner of taped evidence, the focus is on comparing identical words. Before the spectrographic machine itself is called into play, an aural examination is conducted on the tapes to determine whether the recordings of identical words are adequate for spectrographic comparison [5]. For ease in locating identical words on the tapes, transcripts are used. However, when a transcript is inaccurate, such as in the omission of words, it could affect the projected number of comparable words and lead the examiner to a preliminary conclusion that an adequate number of comparable words is not present to make spectrographic examination results meaningful. In addition, when two persons are speaking simultaneously, or when an intrusive background noise occurs, an indication in the transcript of the timing of this is needed, since such occurrences alter the "voiceprint," causing some word matches to be unusable. This, again, affects the meaningfulness of the spectrographic examination results.

For those involved in legal proceedings, the focus is on who said what, and when, as portrayed in the transcript offered in evidence. The transcript is generally offered as an aid in following what is spoken on tape or as a way of capturing a pertinent portion of the taped speech in a visible, nontransient, form, so that it can be further scrutinized, or both. As such, the transcript needs to be accurate. The timing of overlapping speech, or the ringing of a phone, for example, may play an important role in the case, and such things need to be marked accurately on the transcript.

The focus on a true *verbatim* rendering, with a timing element included, sets the evidential transcript apart from others. That this is the case must be made known not only to the law-enforcement transcribers themselves, but also to those who train the transcribers and, perhaps most importantly, to those who task the transcribers with specific assignments.

Suggested Format

The ephemeral, flowing quality of speech is similar to that of music. There is a similarity, too, in the problems involved in attempting to capture both on paper. Although a musical score represents, on paper, the continuous flow of music, it must do so in a noncontinuous manner. The musical notation must break when the side of the page is reached and then continue on the next "line." So, too, must the representation of the continuous flow of sounds in a conversation be constrained by the ready accessibility of standard paper sizes, as well as by standard machines such as typewriting keyboards.[2] To clarify the transcriptual representation of sequencing of phrases, words, and turn-taking of recorded conversation and to facilitate the examination of evidential transcripts, I have designed a format for law enforcement use and have developed some guidelines for transcribing.

---

[2]The ideal situation would be to have a continuous-feed paper supply moving right-to-left and a typing machine to accommodate it, so that the continuous conversation on a recording could be represented in a continuous manner on paper.

Musical Score, Rather Than Play Script Format

The reference to music is no accident. Music and speech are closely parallel and interrelated systems. It is appropriate, then, to format them in a similar manner. But *similar,* not identical, since it would be extremely difficult for the person who transcribes a tape recording of speech to indicate the actual musical tones, synchrony, cadence, and vowel/consonant occurrences that would accurately represent the spoken word at each point of the speech (at our current level of technology). We settle, then, for similarity.

We can begin the formatting process by thinking of a conversation as an opera, where more than one person can "sing" at the same time, while saying different things. We simplify this by eliminating the tune and staff lines, but assign each speaker a part, allowing for both solo and combined speech. Such representation may look like the partial conversation depicted in Fig. 1.

Features of the Suggested Format

  Identification Block:
     Situation (circumstances of tape)
     Date and time (of tape)
     Speakers/sounds

  Transcript:
     Speakers labelled on separate lines
     Each speaker assigned one, consistent line (example: Speaker 1 always on top line)
     Horizontal (left to right) reading of block of lines represents time passing on tape, with continuation blocks below (number of lines in block dependent on how many lines designated for speakers, background noise, and so forth)
       Adequate vertical spacing between blocks (at least triple spaced)
       Optional marking of approximate time lapse along horizontal line
       Vertical only, or box enclosure of each block (drawn in, or typewriter symbol—see Fig. 2)
       Each utterance/noise indicated at appropriate location along time segment
       Open-ended, allowing for additional lines for other speakers, background noises

Where there is a marked difference in the rate of speech between the speakers, this can be represented by splitting the syllables or spacing the words, or both, of the slower speaker:

$$\begin{matrix} \text{Joe:} & \left[\quad \text{Ne- ver} \quad\quad \text{mind.} \right. \\ \text{John:} & \left[\text{But you always said that} \right] \end{matrix}$$

The basic format presented here eliminates much of the transcribing ambiguity generally found in representing simultaneous speech, overlapping speech, and the areas in which there is clearly only one voice. It also allows the addition of a line for indicating the timing of interfering noises, such as telephone rings, a radio or television broadcast in the background, traffic noises, train whistles, and so on. An added line may also be used for vocal sounds or words for which the specific speaker cannot be identified, for example, a cough or a laugh.

Use of this format attenuates many of the problems referred to above; however, little will really be accomplished for law enforcement purposes without a primary focus on changing our normal listening pattern. Accurate transcription will only be derived from accurate listening. Normally, our natural, subconsciously triggered editing tendencies are activated when we listen to someone speak. We focus on the message (content), placing extraneous sounds in other categories to be discarded or used as concurrent information. For example, when we ask a question, we listen for the answer and tend to wade through other comments and communication signals until we either get the answer, choose to ask again, or cease trying. For a "well, uh, yes" answer, we may record this in our memory as a simple "yes,"
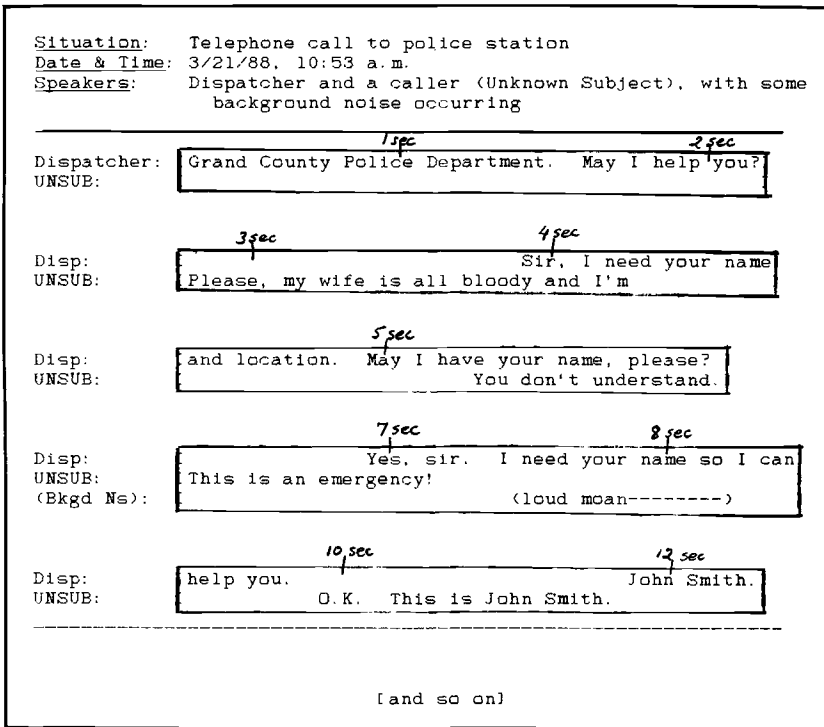
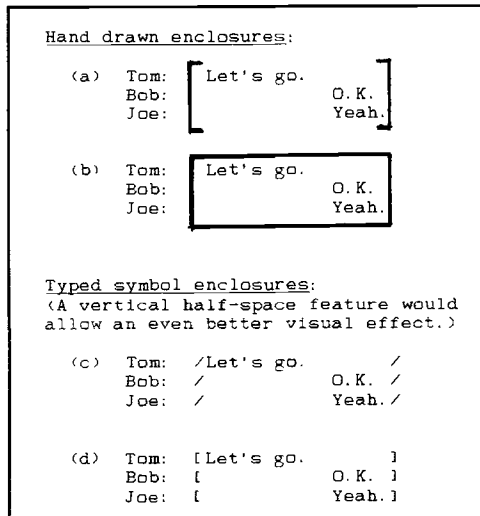FIG. 1—*Sample format for conversation.*

FIG. 2—*Various methods for enclosing blocks of speech.*

while perhaps making a mental note that there was an accompanying hesitation. The "yes" is the message, in this case, and the "well, uh" are extraneous sounds which we may discard, or we may use for our mental note. We would normally consider that we heard a simple "yes" answer. This mode of listening will not result in accurate transcribing, however. The transcriber must work at suppressing the editor within.

Mechanical problems are also encountered when one attempts to transcribe a tape. Often, problems of the mechanisms of recording equipment, compounded with the nature of speech itself, render portions of the recording inaudible, unintelligible, or extremely difficult to hear. Even when one listens to these portions on good equipment a multitude of times, and even when "enhanced" tapes are used, attempting to discern what was said on the tapes can be a frustrating and very time-consuming experience. Those persons dealing with transcripts realize, of course, that not all problems can be solved and that transcribing is deceptively difficult.

The suggested format is aimed at facilitating the task of the transcriber and, most assuredly, will more closely fill the needs of the law enforcement examiner. At a quick glance, one can identify the conversational turn-taking points, the areas of simultaneous speech, and the points at which the speech signals would be altered by additional voices or background noises. A great deal of time will potentially be saved.

Since this formatting is nontraditional (although more functional), it will require some readjustment in both the producing and the reading. It must be kept in mind that, rather than being like a literary play script, it is like a musical score. It is to be kept in mind, too, that this format is offered as a rudimentary construction, with the expectation that refinements and improvements will be made by the user group. The applications can also extend beyond the law enforcement community.

Suggested Technique and Style

Techniques and intricacies of formatting transcripts are not taught in typing classes or secretarial courses. Secretarial handbooks and procedures manuals do not include adequate, if any, coverage or guidelines concerning this topic. Extending this observation, we see that even novels and other written materials which include representations of conversation do not demonstrate consistency in this task. Each office or organizational grouping may develop its own technique, for internal use, such as in the case of court reporters, who may be centrally trained within a jurisdictional system. For the most part, however, each secretary (or transcriber) is left to wrestle with transcripts as well as he or she is able. To remedy this situation, guidelines are needed concerning technique and style. Following are suggestions to provide uniformity of transcribing within the law enforcement community.

Technique

1. Use the best equipment available to you, including good quality earphones.

2. If you know it is a short recording, listen to it all the way through before starting to transcribe. This will help you plan. If it is a long recording, listen to a good portion of it first.

3. With a poor quality recording, have audio-"enhanced" copies prepared by a well-equipped forensic science laboratory, if possible, to attenuate extraneous noises and acoustical room distortions. Use the enhanced tape.

4. Begin the transcript with an identification block (see Fig. 1).

5. Listen to a phrase at a time, type the phrase, back the tape up to listen to a repeat of that phrase, and continue into the next phrase. This allows for checking your work as well as correctly interpreting transition sounds.

6. When you cannot clearly hear a segment of speech, play that portion over several times (20 to 30 times is not an unusual necessity).

7. When two or more persons are speaking simultaneously, play that portion, listening to only one speaker, then play it back, listening this time to the next speaker, repeating this process until all speakers and background noises are accounted for.

8. When finished, play back the entire recording while following along in your transcript.

9. Remember, transcribing taped speech is a slow process.

Style

Rule of Thumb—The suggested "rule of thumb" is to use conventional spelling, with only slight modifications, so that the words are easily readable.

Agreement/Disagreement (Alternate Forms)—In informal speech, an actual "yes" or "no" is not nearly so often heard as is some alternate form. There is such a variety possible in the spelling of the alternate forms, however, that confusion often results. The following spellings that appear in quotations are suggested:

	Mouth Open	Mouth Closed
Agreement		
by grunt	"uh huh"	"mm hmm"
by word	"yeah"	. . .
Disagreement		
by grunt	"huh uh"	"hmm mm"
by word	"nah"	. . .

Participation Markers—During conversations, while one person "has the floor" for an extended period of time, the other participant(s) is expected to give occasional gestures, or vocal markers or both to indicate that he or she is listening and is still participating in the conversation. These markers are either solicited or unsolicited by the speaker. They often take the form of the agreement/disagreement indicators (see above), or slightly more simplified versions, such as "mmm" or "hmm." The transcript should, of course, note the gesture (when videotape is being transcribed) or vocal marker, or both without attempting to interpret whether it is a participation marker or an actual agreement.

Hesitations—Hesitations frequently occur when one feels uncomfortable conversing (as when feeling "put-on-the-spot"), or when one is trying to interrupt another. In either case, the hesitation usually manifests itself by the following sequence: beginning sound of a word, pause, optional repeat(s) of these two, then the complete word and those (if any) following. To represent this on paper, you must first listen through to hear what the initial word is, then write the letters that represent the portion of the word that begins the sequence, with a hyphen after, and then continue on, as appropriate.

Example: Th- Th—That guy is O.K.

Time Fillers—Time fillers usually occur when one is searching for a word or, for some other reason, needs a bit of time to think about just how to say something. A short multiple of the lengthened sound is sufficient to convey the idea.

Examples: Welll;
Uhhh.
Uhmm.
Mmmm.

Emphasis—When you hear words spoken with emphasis, usually indicated by the voice being louder and higher pitched, this can be transcribed by using all capital letters for those words.

<div align="center">

Example: That is HIS way of doing.

</div>

Tone of Voice—Although there is sometimes a temptation to indicate in a transcript that a speaker's tone of voice was angry, menacing, jovial, and so forth, it is recommended that such notations be left out, since they are of an interpretative nature.

Pauses—Pauses can be of varying length. Using the format of Fig. 1, the slight pauses are easily indicated by the placement of the words along the horizontal time line; the longer pauses of one speaker when another is speaking are also represented in this way. When there is a silence on the part of all speakers at once, a simple notation, such as "(long pause)," can be made.

Spelling/Number Listing—When the speaker spells a word out, the recommended written representation is to use underlined, capital letters, with commas and spaces separating the letters of the series.

<div align="center">

Example: That's Miller. $\underline{M}$, $\underline{I}$, $\underline{L}$, $\underline{L}$, $\underline{E}$, $\underline{R}$.

</div>

When the speaker lists a series of numbers, such as in an address or phone number, the recommended written representation is to use words, rather than numerals, again with commas and spaces separating the members of the series. For the phone numbers, a dash in the conventional place makes it easier to read. Conventional hyphens should also be used, such as "twenty-two."

Example: My address is one, seven, nine, My Home Lane.

Example: The phone number is one, two, three-four, five, six, seven.

Dialect/Accent/Idiosyncracies—When you identify the speaker as having an accent or a dialect, a notation of your observation can be made in the identification block; however, it is not recommended that an attempt be made in the transcript itself to imitate the dialect/accent pronunciations of words, such as altering the vowels. Attempted imitation could lead to difficulty in reading the transcript. Of course, you transcribe what is said, so that if no verb is used in a sentence, you do not add one to make it grammatically correct. The general "rule of thumb" applies: use conventional spelling, with only slight modifications that would not deter from easy reading. With idiosyncratic speech habits, slight, conventional modifications of spelling can be used, since these are in common usage and are, therefore, easily read:

Example for dialect/accent: (Use conventional spelling.)

Examples for idiosyncracies: I s'pose you're right.
Yes, he's comin' home.

Whispers—There is a difference between a whisper ("voice box" not used at all) and whisper-like, breathy, speech. An appropriate notation, such as "whispered,"or "almost whispered," should appear in the transcript. This notation may have to be placed between the lines to maintain the proper timing when more than one speaker is involved. When the entire taped speech is whispered or almost whispered, the notation can appear at the beginning to cover the whole tape.

Unintelligible/Inaudible—Frequently, taped speech has portions that the transcriber finds impossible to transcribe. Care should be taken to make notations in the transcript

accurately; that is, if the sound of the voices were actually not heard, then that portion is "inaudible," and if voices were heard, but the words could not be distinguished, that portion is "unintelligible."

Conclusion

The law enforcement community has long been struggling to "force fit" traditional types of transcripts of tape-recorded spoken language to its own use. The traditional types of transcripts do not, however, adequately address the special requirements of law enforcement. A more productive approach is to develop a unique type of transcript—one suited to the needs of the law enforcement community and other user groups having the same special requirements of exactness and time indicators. The transcript format and guidelines presented in this paper are designed as a base for development of this new type of transcript. Manuals and secretarial training programs need to reflect this new mode of thinking about transcripts for change to be implemented. Perhaps the most important starting point for implementing change in the law enforcement community, though, is the community's recognition that the evidential transcript has special requirements that can best be addressed by a unique format such as that presented in this paper.

References

[1] *State of Texas v. Thomas Cullen Davis*, Criminal District Court, Tarrant County, TX, Case 16838, 1979.
[2] Molyneaux, D. and Lane, V. W., *Effective Interviewing: Techniques and Analysis*, Allyn and Bacon, Boston, 1982.
[3] Deering, M. J. and Pomeroy, B., *Transcribing Without Tears: A Guide To Transcribing and Editing Oral History Interviews*, Oral History Program, George Washington University, Washington, DC, 1976.
[4] Davis, C., Back, K., and MacLean, K., *Oral History: From Tape To Type*, American Library Association, Chicago, 1977, pp. 35–36.
[5] Koenig, B. E., "Spectrographic Voice Identification," *Crime Laboratory Digest*, Vol. 13, No. 4, Oct. 1986, p. 109.

Address requests for reprints or additional information to
Penelope O. Pickett
Document Section, Laboratory
Federal Bureau of Investigation
10th and Pennsylvania Ave, N.W.
Washington, DC 20535