THE ROLE OF TRANSCRIPTIONS IN THE COURTROOM: A SCIENTIFIC EVALUATION

FAUSTO POZA[1] AND DURAND R. BEGAULT[2]

[1] *Poza Consulting Services, Menlo Park, CA, USA*
tito@poza.net
[2]*Audio Forensic Center, Charles M. Salter Associates, San Francisco, CA, USA*
Durand.Begault@cmsalter.com

The past thirty-five years have seen a heated debate in both scientific and legal venues as to the proven accuracy of Forensic Voice Identification, which is now admissible in some jurisdictions but not in others. It is notable that there has been little, if any, scientific reporting on the common practice of allowing the use of transcriptions of difficult to understand recordings as aids to the trier of fact in court proceedings. The use of such transcripts, whether produced by "experts" or laypersons is discussed in the context of their potential for anonymously biasing the trier of fact. Signal Detection Theory shows that when subjective judgments are made in the presence of uncertainty, as is the case when trying to understand marginally intelligible recordings, the criterion of the decision maker can be significantly affected by external factors. When ruling on whether to allow transcripts of marginally intelligible recordings to be used as "aids" to the trier of fact, the Court should consider whether such "aids," rather that the recordings themselves will effectively become the evidence.

INTRODUCTION

In this paper we hope to illuminate certain scientific issues that we believe should be considered by Courts in the course of reviewing the admissibility of transcriptions of recordings that contain marginally intelligible speech utterances. Our comments will be especially directed at such transcriptions that are produced for use in court by "so called" experts, as distinguished from transcripts that are prepared to assist attorneys in the preparation of their cases.

When audio recordings are used as evidence in civil and criminal trials, it is common for the side introducing the recording to submit a transcript of the recording as an aid to the trier of fact. In most such cases the recorded speech is totally intelligible, or at least neither side objects to the submission of a specific transcript as an aid to the trier of fact. In such cases the transcripts were usually prepared by administrative assistants or by other staff members of one side or the other.

There are cases, however, in which the speech content of an audio recording may include crucial portions that are very difficult, if not impossible, to make out. Understandably, one side in a legal dispute may find the words transcribed by the other side in such portions to be objectionable. In most disputes of this kind, the Court has been inclined to rule that the trier of fact may be given the disputed transcript as an aid and, in order to even the playing field, the other side may submit its own version of the transcript to the trier of fact as well. The fairness of this decision, from a scientific standpoint, is the central issue we will discuss in this paper.

In these kinds of cases, the side introducing the recording may have retained an expert to attempt to enhance the recording in hopes of improving its intelligibility. In a majority of cases, even the most advanced enhancement software does not significantly improve the intelligibility of the difficult portions of such recordings. When enhancement does not produce a new version of the recording that is understandable to the attorneys offering the recording, or to their staff, they may ask the expert who enhanced the tape to also produce a transcript, or they may retain a second expert to produce a transcript using either the original or the enhanced recording, or both.

1 AN ILLUSTRATIVE EXAMPLE

In order to provide a narrative to which we can refer during our discussion, we present the following hypothetical situation that could occur within a criminal trial. A 911 recording is placed into evidence by the prosecution in a murder case. The 911 call had been made by one of two parties involved in a domestic altercation. The recording shows that after the 911 operator has responded, the two quarreling parties continue the argument, paying little, if any, attention to the operator's queries. At some point in the recording

the person who did not initiate the call (who will be referred to as the shooter) turns away from the caller to go to another room, hurling invectives at the caller as he moves away. The caller now informs the operator that the other person has become abusive and asks for help. The distant person can be heard, if only with marginal intelligibility, speaking in a distant room. At this point footfalls are heard followed by two loud impulsive sounds. There are then more footfalls fading away and more marginally intelligible speech as the speaker leaves the room. At trial, the shooter is being accused of first degree murder. The evidence used by the prosecution to justify first degree murder, as opposed to a lesser charge, is found in the transcript of the 911 call. The transcript attributes marginally intelligible utterances to the shooter that suggest the shooting was premeditated. The prosecution asks the Court to admit the transcript into evidence against the defendant. The prosecution argues that the 911 recording has been enhanced and transcribed by an expert and, therefore, should be admitted.

2 SCIENTIFIC ISSUES

The scenario described in the above paragraph poses two questions that have been raised in scientific circles. The first relates to the question of whether audio enhancement techniques can alter the audio signal in a way that might affect how it is understood. In other words, in the case of garbled speech, could enhancement cause even subtle changes in the spectral energy that might lead a listener to perceive a sound differently when compared to the unenhanced recording. If so, how does one determine which perception reflects ground truth. Although there is not yet scientific evidence which bears on this question, it would be imprudent to entirely dismiss the issue.

The second question, which is more relevant to the issues addressed in this paper, relates to whether or not the process of transcribing difficult to understand audio recordings constitutes a scientific technique or procedure. Although Speech Understanding is a well researched scientific area, the purpose of its study is to learn more about the way humans perform the task of understanding speech, not to attempt to teach people how to understand marginally intelligible speech. This knowledge is sought in order to enable researchers to intelligently design automatic speech recognition systems, and the expertise required by these researchers primarily includes Artificial Intelligence, Linguistics, Acoustic Phonetics, Articulatory Phonetics and Computer Science. Scientists in these areas do not profess greater speech understanding abilities than a layperson, nor do they claim to be able to teach laypeople or colleagues to more accurately produce transcripts from marginally intelligible speech. There is, in fact, no reported scientific methodology that would enable a person to become expert, and therefore superior to a layperson, in understanding speech from a recording in which the speech is either too weak or too noisy or too garbled to be easily understood.

3 LEGAL ALTERNATIVES

If, in the hypothetical case described above, the prosecution attempts to have the transcript admitted as the work product of a qualified expert, the defense should object on the grounds that there is no scientific underpinning to support the claim of expertise in this area. Lacking scientific evidence that would contradict this premise, the prosecution's attempt to have the transcript admitted into evidence should be denied.

The prosecution might then argue that the transcript should simply be used to assist the trier of fact while the recording is being played in the courtroom, and need not be admitted as evidence. By removing the question of expertise and admissibility from the argument, the prosecution hopes to have the trier of fact still be influenced by the words in the transcript, albeit without the aura of authority that an expert would lend them.

We believe that the overriding issue is not whether the transcript is admitted as evidence, but whether the trier of fact will be exposed to someone else's interpretation of difficult to understand words in the 911 call. In other words, we feel that the more important problem is the pejorative affect that reading the transcript will have on the ability of the trier of fact to independently assess the meaning of the speech heard in the recording. In fact, once the trier of fact reads somebody else's rendition of marginally intelligible speech, he or she will no longer be able to bring their own interpretation to what was said free of the biasing effect of the supplied transcript.

Under the circumstances described above, some Courts have applied, what we will refer to as the "even playing field" decision. We choose this designation because we believe that the Court is trying to find a fair solution that will not exclude the use of a transcript by the trier of fact. Since the defense argues that some of the words in the prosecution's transcript are prejudicial and should not be seen by the trier of fact, and the prosecution argues that the transcript is accurate, the Court decides that it will allow each side to submit a transcript and will ensure that the recording is played at least twice, once for each version of the transcript.

While the Court may believe that this arrangement has "evened the playing field," it has, in fact, done no such thing. It hasn't had the desired affect because the difference of opinion about the prejudicial words in the transcript is usually manifested by the defense's

transcript showing "(Unintelligible)" in the places that the prosecution's transcript shows the words that are objected to by the defense. Consequently, the trier of fact gets to see the objectionable words during one playing of the recording, and then sees "(Unintelligible)" during the other playing. In terms of prejudicing the trier of fact with some other person's rendition of the marginally intelligible words in the recording, there might just as well have been only one transcript.

A more fair approach that results in the production of a transcript that is unlikely to offend the defense in the case hypothesized above would be to have two or more court reporters transcribe the recording and let the final transcript reflect only those portions on which there is agreement. The remaining portions can be marked as unintelligible and the trier of fact can decide for him or herself whether he or she can understand those portions. This makes sense because court reporters represent average listeners with advanced stenographic skills but, with no contextual knowledge going in to the task.[1]

4 BIAS AND SIGNAL DETECTION THEORY

In order to fully understand the scientific basis behind the defense's objection to allowing the trier of fact to use a transcript as an aid, one must refer to some of the principles that derive from the study of Signal Detection Theory (SDT) [1]. (A discussion of SDT in the context of the related field of voice identification and elimination is found in [2, 3]). SDT provides a methodology for analyzing decision making in the presence of uncertainty. Attempting to transcribe marginally intelligible recordings is the epitome of a task that requires decision making in the presence of uncertainty. Anyone attempting such a transcription brings many personal "biases" to the task. The term "bias" is not used here in a pejorative sense but rather to describe a, sometimes unconscious, predisposition to make a particular choice when faced with a difficult decision.

In a general sense, one's linguistic background and personal experiences provide biases in difficult transcription, but more significant are the content specific biases that inevitably affect the transcriber's decision making. The content specific biases come from the knowledge the transcriber has acquired about the context of the recording. Even if the transcriber has been told nothing about the events related to the recording, it is inevitable that some context will be acquired from listening to the intelligible portions of the recording itself. Any and all such information, possibly including inputs from one of the participants in the recording, presented to the transcriber, will have an important influence on whether a specific set of words is chosen to represent a particular segment of speech, or whether the transcriber decides to refer to that segment as unintelligible.

For example, the transcriber may use the intelligible portions of the recording to "reasonably" transcribe the unintelligible utterances. The most benign example of this effect is when a speaker in a recording asks a partially unintelligible question, but the intelligible answer disambiguates the question sufficiently to allow the transcriber to "transcribe" it. In actuality, of course, the question is still unintelligible, but external factors have altered the transcriber's criterion to the point that he or she may be convinced that they actually understood the unintelligible portion. The example is benign because, in all likelihood, the unintelligible portion was probably correctly transcribed. In most situations, however, the transcriber does not have such direct knowledge regarding the possible content of the unintelligible portions of a recording. In these cases the transcriber attempts to decode the unintelligible utterances in a manner that is consistent with the general tenor of the intelligible portions of the recordings or, worse yet, by using information from people knowledgeable about the nature of the recording but not included in the exchanges in the recording itself. In either case, the transcriber is likely to decode the unintelligible portions in a "reasonable" manner, but not in an unbiased manner, and by no means necessarily accurately.

The above paragraphs have tried to make clear the fact that any transcription, whether by a Forensic Audio Expert, a dedicated paralegal or a dedicated layperson, will reflect the criteria used by a particular individual while listening to the questioned recording. When a person, such a the trier of fact in a legal proceeding, is to listen to the recording only a few times, compared to the dozens or even hundreds of times the transcriber may have listened to it (or at least to portions of it,) one realizes that marginally intelligible portions of the recording will probably not be understood by the trier of fact. This is especially true if the trier of fact does not have the benefit of using headphones during the audition process.

So, one can argue, why not let the fact finder use a transcript to enable him or her to better "understand" the recording? The answer is that Signal Detection Theory tells us that in such a circumstance, the fact finder is very likely to "hear" what he or she reads in the transcript because the exposure to the transcript has shifted the listener's criterion to a point where he or she will accept the printed words as a "reasonable"

[1] See Appendix A.1 for further discussion re presentation of audio material in court.

interpretation of the speech. In other words, without the transcript, the outcome of the listener's criteria upon listening to marginally intelligible speech is that it is understandable, or it is unintelligible. Regardless of his or her choice, one can be certain it is unbiased by outside suggestion. With a transcript, the listener's criteria will most likely be shifted by having to consider a specific alternative provided by some unknown person. The criterion shift reflects a new awareness that this unknown person may have some knowledge, unfamiliar to the listener, that has enabled the unknown person to decipher the marginally intelligible speech. The very fact that the Court has allowed him or her to use the transcript may affect the trier of fact's criterion. Even cautionary instructions by the Court to only consider the transcript as an aid will not remove bias; as the proverb states "you can't un-ring a bell."

The essence of the argument is: When the task at hand is one that the trier of fact can execute as well as any other person, *it is precisely the trier of fact's unbiased criterion that our judicial system seeks*— not the trier of fact's criterion that has been shifted by some other person who is no more qualified at the task than he or she. Triers of fact are routinely asked not to read newspaper accounts or other forms of media exposure of their trials for exactly the same reason. In other words, if, as we contend, there is no scientific basis to claim an expertise in transcription, the precepts of Signal Detection Theory tell us that *any* transcript used by the trier of fact preempts the province of the trier of fact. Such a pre-emption is, of course, especially egregious if the transcript of an "expert" is admitted. In such a case, it is very likely that the trier of fact will have their criteria shifted even further by an impressive curriculum vitae, and therefore would accept the validity of all the words in the transcript, no matter how unintelligible.

Having made the point that the use of transcripts by the trier of fact when listening to marginally intelligible recordings will almost certainly result in interpretations of the recordings that are highly biased by outside sources, we are not discounting the usefulness of transcripts as an aid to attorneys. We do, in fact, believe that Forensic Audio Experts can assist attorneys in cases where such recordings are important to civil and criminal proceedings. We believe that properly rendered transcripts in these cases can be of significant use to attorneys in preparing such cases. By "properly rendered," we mean that such transcripts should be prepared with an eye to reminding the reader that in recordings that contain segments of speech that are only marginally intelligible, the transcript should reflect that fact. It is not uncommon to have one side or the other in such legal proceedings, not only present a transcript to aid the trier of fact (whether prepared by an expert or others,) but to have such a transcript contain only plain text showing no differentiation between words and phrases in terms of the relative difficulty of understanding them. In other words, the implication of such a transcript is that, unless the speech segment was transcribed as "Unintelligible," then every word transcribed was as unambiguously understood as every other word. In the experience of the authors, such a possibility is extremely unlikely.

5 CONCLUSIONS

In conclusion, it is our opinion that there is no scientific evidence to justify the admissibility of expert opinion in transcribing recordings for court. Further, if the trier of fact is allowed to read *anybody's* transcript while listening to a recording containing marginally intelligible speech, his or her interpretation of the recording will be unduly influenced by the transcript.

We understand that imposing the restriction of disallowing the use of transcripts as aids may result in a loss of information to the trier of fact due to the poor listening conditions that exist in most courtrooms. This can be avoided by having the triers of fact use headphones when listening to marginally intelligible recordings (see Appendix A.1).

We believe that Forensic Audio Experts can, however, provide useful transcription services to attorneys who are preparing cases involving marginally intelligible recordings. And, in situations where, in spite of the admonitions presented in this paper, Courts do allow the trier of fact to use transcripts while listening to such recordings, we believe such transcripts should be prepared using a methodology that incorporates a textual representation that reflects the relative uncertainties involved in understanding some portions of the recording relative to others.

APPENDICES

A1. PROBLEMS INHERENT IN COURTROOM PLAYBACK OF NOISY RECORDINGS

Unfortunately, under many circumstances it has occurred that triers of fact are unable, for whatever reason, to have the opportunity to audition a tape with the same advantages available to the expert. For instance, a judge will allow playing a noisy recording in the courtroom once or twice over a loudspeaker while court is in session, without a transcript. Jurors may have hearing impairments, court rooms often have excessive reverberation that affects intelligibility, and audio playback systems in courtrooms are usually inferior to what is available to an expert; yet, it does occur that such presentation becomes the only opportunity for the

jury to review audio material.

If the jury or judge is unable to audition the evidentiary recording multiple times, using headphones, and perhaps with the advantage of 'non-destructive enhanced versions' in addition to the unenhanced version of the recording, then information that may either help exculpate or convict a defendant has been effectively hidden. The outcome of a trial may be biased in that the evidence cannot be brought to light, due to insufficient exposure of the evidence to the jury.

The lack of exposure to audio evidence often contrasts the exposure available to the trier of fact for the inspection of visual evidence. For example, a jury is usually allowed to take high-quality photographs to the jury room for subsequent inspection without time constraint. An expert can enhance a photograph to bring out details that would otherwise be obscured, although it is incumbent on the expert to show that the details are indeed not artifacts of their enhancement process, and is not suggesting to the jury what they should 'see'. Similarly, we suggest that triers of fact should be allowed to have audio playback capabilities that meet a minimum standard of quality and that allow repeated listening over headphones (headphone listening is widely accepted as superior to loudspeaker listening for forensic recordings- see for instance, [4]).

A2. CHARACTERISTICS OF FORENSIC SPEECH SIGNALS

Forensic speech signals are those speech recordings that an expert is asked to enhance or otherwise analyze. This discussion excludes 'identification', i.e., determination of who has made a specific recorded utterance. The source context of the recordings can include 'hidden microphones' that are not optimally placed with reference to the sound source; inadvertent pickup of undesired voices or non-speech background noise; 'two-party' contexts where one voice is completely audible while the other is not; etc.. Some common characteristics of forensic speech signals are as follows:

- Nearly all exemplars have band-limited frequency range (esp. telephone, wiretap, hidden mic)

- Nearly all exemplars have limited dynamic range (particularly those involving digital compression, e.g. telephonic codecs such as g.729).

- Many exemplars have low speech transmission index (STI) levels due to low signal-noise and high reverberant-direct sound ratios

- Untrained talkers: articulation, enunciation, 'proper English' are frequently not features of private conversation in forensic contexts.

Figure 1 shows the difference between a 'low quality' signal recorded with a laptop computer microphone versus a typical forensic recording (here, a telephonic recording). Although the laptop computer mic picks up ambient noise, it has a sufficient signal-noise ratio and frequency bandwidth to allow 3rd and even 4th formant frequencies to be visible.

More typical in the forensic world is the telephonic recording shown at bottom of Figure 1. Although the signal-noise ratio (overall) is better than +5 dB, the recording is far noisier and there is a lack of spectral energy in the 3rd and 4th formant frequencies. Nevertheless, the presence of even only the first two formant frequencies allows for some level of intelligibility in distinguishing most speech sounds, except for fricatives (typically, *f* versus *s*; *v* versus *z*).

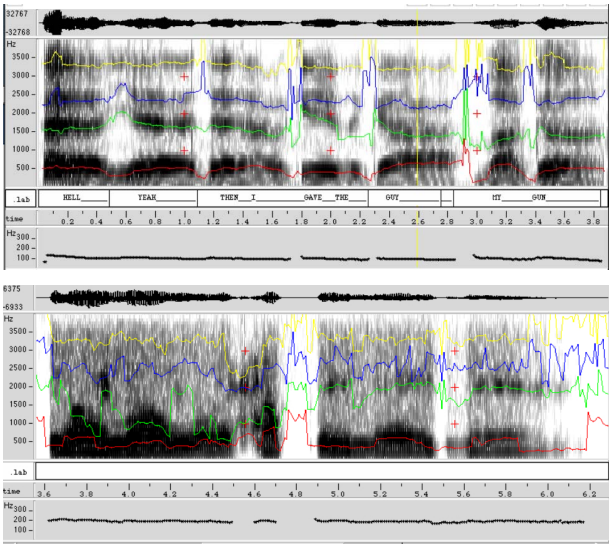

Figure 1: Top- laptop "low quality" recording with high signal-noise ratio and clear presence of 3rd formant; Bottom- more typical forensic recording having low signal-noise ratio and 3rd formant nearly invisible

A.3 OPTIONS IN THE TRANSCRIPTION OF MARGINALLY INTELLIGIBLE SPEECH

More specifically then, what options does an attorney have when seeking to obtain a transcription of recording in a civil or criminal case? What role can the Forensic Audio Expert play to assist attorneys in such situations?

It is possible to classify those whose job is to transcribe legal recordings in terms of their practitioners. Here are

the differences between traditional legal transcribers, such as court reporters, and translators; and a Forensic Audio Expert who practices 'Forensic Transcription'.

The context of traditional legal transcription involves either real-time or recorded speech to text conversion, where the spoken word is for the most part intelligible and presumably unambiguous. There are varying levels of certification that are principally based on speed. For example, a certified court reporter is a transcriber who is 'certified' primarily in terms of number of words that can be accurately typed in legal contexts such as depositions and trials. Qualification can extend to expertise at the practice of 'real-time' transcription from intelligible speech[2]. Required qualities of a court reporter include sufficient command of the language being spoken, attention to detail, and the ability to focus for long periods at a time.

The interpretation by transcribers of speech sounds into text is limited by the specific transcribers' vocabulary; they cannot transcribe a particular dialect or language unknown to them. Professional transcribers are trained to notate and subsequently verify spelling of unknown words heard, particularly if they are present during the spoken word (e.g., immediately after a deposition). However, bias in the form of 'presumed knowledge' can alter the words that are said, particularly if the transcriber is working from a recording without the benefit of being able to confirm the truth of what was said. In fact, particularly in non-real time circumstances, some legal transcribers will overlook the 'errors' in the articulation of speech or will effectively 'translate' dialect in order to provide 'the meaning' in written text; one must specially request a 'literal' transcription of the actual words said. For instance, an interpretation of what is said in terms of a two-alternative forced choice 'yes' or 'no' might originate as follows:

As heard ('literal'):
P: Did you see the suspect?
A: Umm.....uh....well, yeah....
As written:
P: Did you see the suspect?
A: Yes

In the written version, one misses the uncertainty inherent in the affirmative response that may or may not be important. The other side of the coin is that 'literal' transcription can veer into literary stylization, and at worst can supply extraneous cues to the sound quality of speech such as capitals, and explanation marks that can unwittingly bias the reader (trier of fact). For example consider the following two approaches to transcription of an incident recorded on a 911 tape.

'Literary' stylization:
A: Let GO of me! PLEASE.....oooh....
LET GO OF ME!!!!!! owwwwch.

Preferred:
A: Let go of me. Please. Let go of me.
(non-verbal vocalizations).

In the latter preferred version, the transcriber made an effort to indicate that the non-verbal vocalizations that occurred, without using suggestive or 'visual' language (e.g., *"sounds of pain"*). The semantic content is preserved while simultaneously supplying a note that may or may not be useful to an interested party who wishes further information by listening to the recording.

The audio forensic expert who practices forensic transcription may produce a more "error free" transcript than the traditional transcriber in that they are aware of the role of bias; both within their own process of transcription, and with a consequent effort to indicate their level of uncertainty via the written word. The audio forensic expert also has access to, and is aware of the potential invasive influence, of specialized signal processing software and hardware, and high-quality listening facilities (waveform editors; audiophile-grade headphones; etc.). Finally, the audio forensic expert is typically more expensive in terms of hourly rate compared to a traditional transcriber, and presumably would have a higher criteria for notational accuracy on selected, 'difficult to hear' passages, with an understanding of the limitations typically available to a traditional legal transcriber. Nevertheless, two different experts may reasonably disagree where vocalizations are ambiguous or uncertain, regarding their 'best estimates' of what is said.

Different audio forensic experts who transcribe marginally intelligible recordings may attempt to accomplish the goal of 'properly documented transcription' in different ways. A protocol involving a complex key is frequently involved; such a key allows the client to assess the degree of confidence in a transcription of the speech. Figures 2 and 3 contrast two different approaches used by each of the authors.

The use of such protocols in no way contributes to the accuracy of the transcription, but it does make the reader aware of the range of difficulty a listener might have in understanding the words spoken in a recording. We believe such information is important generally, and might be even more important were the transcript at

[2] For example, the National Court Reporters Association awards the title Registered Professional Reporter (RPR) to those who pass multiple examinations and participate in continuing education programs. Additional certifications can be earned that demonstrate higher levels of expertise, such as Certified Real-time Reporter (CRR).

some point to end up in the hands of a trier of fact. In such an instance the transcript would at least alert the trier of fact to the reality that some, perhaps crucial, words had been more difficult for the transcriber to understand than others.

KEY:

[-] = unintelligible
[example] = 'example' is best guess for what was said
[example1/example2] = example1 and example2 are alternatively best guesses
[*italics*] = sounds like
{*laughs}* = extraneous sounds (beeps, laughing, noise, etc.)
(*NOTE)* = note from transcriber

EXAMPLE:

Curly:	If I have to, I'll [deliver] [-]. I hope we won't get to that.
Moe:	Listen, I want you to [*fill*] the guy
Curly:	[*nyuck nyuck nyuck*]

Figure A.2: Transcription key and example, DB

- Words sufficiently intelligible to be understood by native listeners without the need of contextual or syntactic information:

  I'll have the soup

- Words marginally intelligible when heard in isolation, but understandable and reasonably unambiguous when heard in context :

  Pass the salt and (pepper)

- Words or phrases ambiguous but which "fit" a particular interpretation that seems to make contextual sense, are surrounded by braces {}.

  Do you have a {spoon} for my soup

- Words or phrases that are totally unintelligible or are too ambiguous to resolve in the immediate context, are indicated by an estimate of the perceived number of syllables.

  I think this soup has [3]

Figure A.3: Transcription key and examples, FP

A4. COMMON MISCONCEPTIONS

It is both characteristic and important that the forensic audio expert has expertise using specialized audio hardware and software. But triers of fact must realize that no system can magically restore 'missing' signal information having insufficient level– it must be present in the original recording. The phenomenon of auditory masking indicates that noise in one critical frequency band can influence detectability (and therefore intelligibility) in adjacent critical bands, and so strategically mitigating the noise while amplifying the signal in terms of frequency analysis can sometime be helpful. Level conditioning (compression) can also be helpful. But changing the spectral and amplitude characteristics of the recording usually has more to do with avoiding listener fatigue and minimizing masking than with any direct manipulation of the signal *itself* that is useful for improving intelligibility. In particular, signal processing that involves 'smearing' of the time signal can have onerous results on the recognition of consonants. It is often prudent to compare, often, unprocessed and forensically 'cleaned' recordings.

The notion that *"an audio forensic expert uses a **scientific approach** to enhancing and transcribing noisy recordings"* is problematic in that the normal standards of the scientific method, wherein experimental hypothesis testing and statistical analysis is employed, are not part of the expert's process in making a transcript. It is true that experts employ technical skills and experience, and can make use of techniques that are based on scientific principals; but transcription and enhancement *in itself* is not bound to the scientific method nor is it 'proven' in any study. (The notion of 'scientific approach' is incorrectly applied in many forensic disciplines within and outside of audio).

Perhaps the most problematic misconception is that of the expert who asserts to have a *golden ear*. At the beginning of the paper, the notion of laypersons versus experts at what is essentially *listening and transcribing* was discussed. Nevertheless, the notion that a forensic audio expert is a superior listener compared to lay listeners is commonplace. In fact, an audio forensic expert who transcribes may be more careful, or have better technology, than a layperson; but no scientific study bears out what can termed the *golden ear* hypothesis.

Analysis by a person with a "golden ear" refers in the audio industry to experts who are presumed to be able to hear features in an audio recording, typically music, that 'normal' listeners would not hear. In the past, such golden ear experts were used more frequently in the past by loudspeaker manufacturers, acousticians, etc.; today larger companies use more scientific approaches based on statistical response of a panel of trained listeners [5]. The golden ear mystique has unfortunately found its way into court as well, where the 'expert' can supply the court with information that might not be otherwise discerned by laypersons.

> …[the audio expert] enhanced garbled or faint recordings after other experts, including those at the FBI, were unable to do so…..the US attorney's office in Tampa, Fla., hired [the audio expert] to enhance recordings that were the key evidence against a couple suspected of killing their baby daughter. [The audio expert] said he heard incriminating utterances by the parents, including a comment by the mother that the child "died real bad". But after listening to the tapes, a federal judge said they were worthless as evidence. "I heard none of it" said US District Judge Steven D. Merryday, who later awarded the couple nearly $3 million in attorney's fees after the federal government conceded that charges never should have been file. [6]

We do not know if there were any particular difficulties for the judge to hear what the audio expert must have heard. But clearly, the difference between what one expert might report as fact and what another expert might report with an effort towards communicating their level of uncertainty can be dramatic.

REFERENCES

[1] D. M. Green and J. A. Swets *Signal Detection Theory and Psychophysics. New York: Wiley (1966)*

[2] F. Poza and D. R. Begault, "Voice Identification and Elimination Using Aural-Spectrographic Protocols", *Audio Engineering Society, Proceedings of the 26th International Conference, Audio Forensics in the Digital Age*, Denver, USA, pp. 21-28, July 7th-9th 2005.

[3] National Academy of Science, Committee on Evaluation of Sound Spectrograms. *On the theory and practice of voice identification* (1979).

[4] B. E. Koenig, D. S. Lacey and S. A. Killion, "Forensic Enhancement of Digital Audio Recordings" *Journal of the Audio Engineering Society* vol. 55, no. 5, pp. 252-371 (2007).

[5] S. Bech, S. and N. Zacharov *Perceptual Audio Evaluation. Theory Method and Application*, Wiley (2006).

[6] *Los Angeles Times* Feb. 1, 2004.