VOICE IDENTIFICATION AND ELIMINATION USING AURAL-SPECTROGRAPHIC PROTOCOLS

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

[1] *Poza Consulting, 250 Princeton Road, Menlo Park, California 94205*
Tito@Poza.net
www.poza-consulting.com

[2] *Audio Forensic Center, Charles M. Salter Associates, 130 Sutter St., Suite 500, San Francisco CA 94104*
Durand.Begault@cmsalter.com
www.audioforensics.com

The use of spectrographic pattern matching and aural comparison in forensic voice identification requires careful control of examiner bias and an awareness of the principles of signal detection theory. This paper briefly reviews the history of the aural-spectrographic method and the experimental results of Oscar Tosi, and then summarizes the voice elimination and identification protocols addressed by Gruber and Poza (*American Jurisprudence* 54, 1995) and other authors. Given suitable voice exemplars, and prudent examination protocols, voice examiners can provide useful probative findings in the forensic setting. Opinions expressed in reports and in court room testimony should set forth the limitations of the technique so that the trier of fact can properly evaluate the examiner's findings

1 THE BASIS OF THE AURAL-SPECTOGRAPHIC METHOD FOR VOICE IDENTIFICATION

The "aural-spectrographic" method of voice identification relies on a trained examiner's ability to make a discrimination judgment as to whether a known and an unknown speech exemplar were produced by one speaker or by two different speakers. Even when more than one known exemplar is involved, the fundamental task is a form of an A-B discrimination task. Typically, one or more known voice exemplars must be compared against an unknown exemplar. As indicated by the name, in this method of voice identification the examiner uses both spectrographic and aural information to assist in making the discrimination judgment.

Figure 1 shows a screen shot made from a digital audio workstation for five utterances of the same phoneme. The wideband speech spectrograms ("voicegrams") shown in the Figure reveal the characteristic formant patterns found in vocalic speech sounds. The visual component of the comparison involves an assessment of the pattern similarity between the known and unknown samples.

As to the aural comparison, using the digital workstation, the examiner listens to each sample uttered by the known and unknown played in close temporal juxtaposition, thus enabling him to gauge the similarity of their aural patterns. There is no scientific data that would indicate that a trained examiner can aurally discriminate voices more accurately than a layman, but the data does indicate that aural voice identification, in general, can be quite accurate.

Both aural and visual comparisons are involved in forming an opinion based on the similarity or dissimilarity of the totality of the observed patterns. This is to say that rather than trying to somehow rate individual characteristics, such as formant positions, phoneme durations, etc., the examiner allows experience in spectrographic pattern matching and knowledge of acoustic phonetics to guide him in evaluating the aggregate of appropriate patterns at his disposal. Such pattern matching is known in cognitive psychology as a "gestalt".

Gestalt pattern matching is only effective if the patterns being matched are comparable. In other words, when spectrographic patterns from the known and the unknown are compared to each other, it is crucial that these patterns correspond to the same phonetic utterance. This requirement may seem obvious, but the authors have observed sufficient real life instances of improper comparisons to believe that it represents a significant potential for error. Therefore, simply having a known speaker utter the text of a transcription of an unknown's words does not assure that all of the utterances thus obtained will be usable in carrying out the comparison.

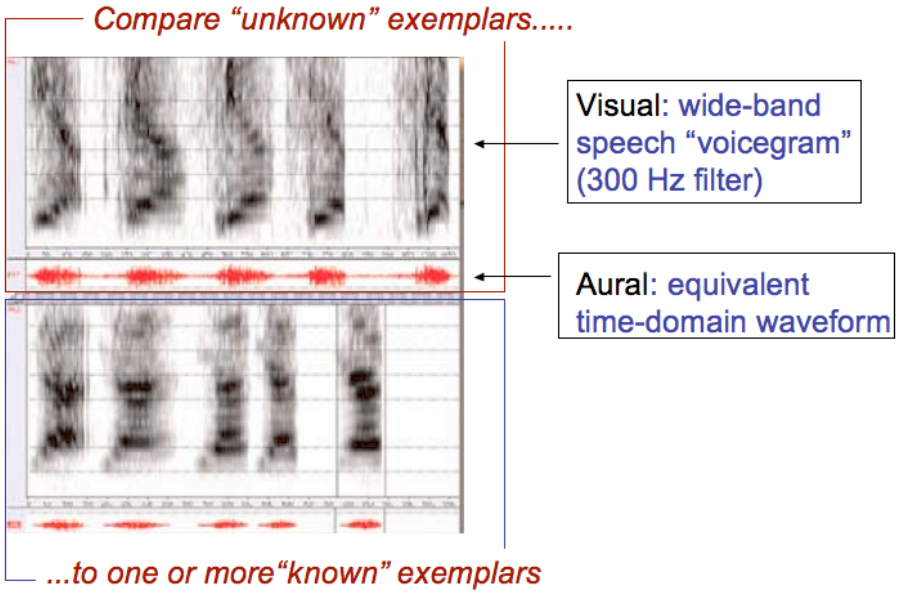

Figure 1: Layout of spectrogram and time-display of waveforms for an aural-spectrographic comparison of an unknown (top) versus a known (bottom) exemplar, using a digital audio workstation. The examiner can alternate between different sections of the exemplars, both aurally and visually. (Software: GW Instruments Soundscope).

The basis for the use of speech spectrograms in the current application of speaker identification in forensic settings is derived from the work of Lawrence Kersta and Oscar Tosi [1, 2]. Interestingly, Kersta, who was the first person to testify as a voice identification "expert" in 1966, did not utilize aural comparisons in his work. Kersta claimed that the accuracy of voiceprints was comparable to that of fingerprints in the forensic setting, which at the time were incorrectly considered to be infallible.

His oft-cited study published in *Nature* was based on novice subjects (high school girls) who had a 0-3% error rate when matching single words spoken in isolation from 12 different speakers. The test was unrealistic in comparison to actual forensic circumstances for many reasons, including the experimental design, the use of individual words in isolation, and the quality of the recording. Error rates were minimized by the use of a *closed set* design for the presentation of the twelve exemplars. Subjects were allowed to make free comparisons among the exemplars until the "best" matches were made. Error rates would have increased had their been a 'forced choice' match made during a sequence of exemplar presentations in an A/B format, from an *open set*, where it was unknown how many matches might be made [3].

Kersta's credibility was further tarnished by the fact that he manufactured and sold sound spectrographs and also profited from the training and "certification" of voice identification examiners. The controversy that Kersta's work caused in the scientific community led to the funding of a large research grant in 1968 by the Justice Department to the Michigan Department of State Police. This grant was overseen by Professor Oscar Tosi of Michigan State University.

Tosi et al. examined the performance of lay observers as voice identification judges under a variety of laboratory conditions, and reported 2-6% false identification and 5-12% false elimination rates [2]. Although Tosi et al. argued that in the forensic setting, error rates would actually improve over the experimental results (the so-called "Tosi Extrapolation"), they presented no scientific justification for such an eventuality. Real-world conditions, which introduce imperfect channel transmission and higher intra-speaker variability, will almost certainly increase error rates, as will inherent pre-screening for similar-sounding voices [4].

The crux of the "Tosi Extrapolation" argument was that professional examiners would, based on their experience, decline to make judgements when the data was so poor that errors were more likely to occur. Unfortunately, the decision to proceed with a given examination is based on a subjective evaluation of the quality of the speech data in any given case. Without objective criteria to guide the examiner, no amount of experience can assure that his or her decision to proceed or not, based on a subjective evaluation of the data, will result in fewer false identifications. This is not to say

that professional examiners should not utilize their experience in their efforts to avoid making judgements as to whether the data is too weak to proceed, but they should not misguidedly assume that such caution will necessarily guarantee a reduced error rate.

In his frequent appearances as an expert witness in courts of law, Tosi postulated, as part of his "extrapolation," that error rates for identification judgments where both aural and spectrographic data were evaluated, would be lower than when either aural or spectrographic data were used alone (Poza, personal communication). In any case, in the forensic setting the various identification modalities are not applied independently of each other, thus raising the question of whether one modality may bias the examiner to the point where the use of additional modalities may not significantly increase accuracy.

Furthermore, approaches that derive a mean 'score' for exemplar similarity or dissimilarity based on the notion of averaging a series of vocal features along a continuous dimension are not scientifically justified. A study by Clark et al. showed that overall gestalt listening proved to be superior to identification based on listening to individual physical or psychophysical measures of speech [5].

2 CURRENT STATE OF VOICE IDENTIFICATION

Forensic speaker identification, as currently practiced, is grounded on fundamentals of speech science, and its use can be justified on the basis of experimental results, but is not infallible. The audio forensic community must be aware of, and must deal with, the history of overstated rates of accuracy and reliability attributable to Kersta and Tosi. Given suitable quantity and quality of voice exemplars, combined with conservative criteria, forensic voice examiners can provide opinions that can be probative in the legal setting. It is important, however that the trier of fact be made aware of the limitations of the technique. In particular, the examiner must resist any possible coercion, on the part of a client, to make claims regarding proven error rates under forensic conditions. Unfortunately, there is currently no scientific data available upon which to make such claims.

In a large-scale review conducted in the late 1970's, the National Academy of Sciences concluded:

> ....estimates of error rates now available…do not constitute a generally adequate basis for a judicial or legislative body to use in making judgments concerning the reliability and acceptability of aural-visual voice identification in forensic applications [6].

That statement continues to accurately reflect the state of research in this area.

Nevertheless, as mentioned above, in spite of its limitations, forensic voice analysis can provide useful information with probative value in many circumstances. Although examiners who use the aural-spectrographic method of voice identification are more accurately described as "forensic speech scientists" than "forensic phoneticians", Peter Ladefoged's description of the way forensic voice identification is carried out is quite insightful. In a recent Acoustical Society of America newsletter, he commented:

> Forensic phoneticians are like medical doctors giving prognoses. They make many tests that provide useful clues, but their opinions are inevitably based on their own experience....they have evidential value, but they are not established scientific truth [7].

3 IMPORTANCE OF CONTROL OF BIAS

In the forensic setting, an awareness of inherent examiner bias and the adoption of a protocol to mitigate its effects is essential. It is equally important to convey the significance of this issue, and of the examiner's rationale for a given protocol, to the trier of fact. This is especially important because the forensic examiner performs a discrimination task (A/B) between exemplars of known and unknown individuals. *Bias* is defined as "Inclination; bent; a preconceived opinion; a predisposition to decide a cause in a certain way, which does not leave the mind perfectly open to conviction. To incline to one side" (*Black's Law Dictionary)*. This can be especially true for any forensic examiner who performs a discrimination task between known and unknown exemplars when the person giving the known exemplar is known to the examiner to be suspected of a crime.

An understanding of the manner in which bias can influence the outcome of a subjective decision task can be had via an introductory explanation of the signal detection matrix (hit-miss-false alarm-correct rejection) and the receiver operator curve (ROC curve). Figure 2 illustrates the possible outcomes of an examiner's decision when faced with a two-alternative forced choice regarding a known and unknown exemplar being from the same person or from two different persons. The matrix shows not only "correct identification" and "correct rejection", but also two types of error: "false alarms" (an incorrect identification, or Type I error) and "misses" (incorrect elimination, or Type II error).

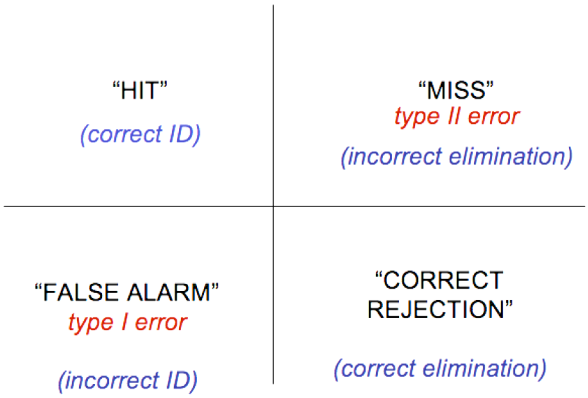

Figure 2: Decision matrix.

Under the most difficult identification conditions described in reference [2], Tosi observed a 11.8% type II error (miss) and a 6.4% type I error (false identification rate). The relatively higher Type II error rate is acceptable, in light of the fundamental precept of the U.S. Justice system that it is more important to not convict an innocent person than it is to convict a guilty one. The phrase, "beyond a reasonable doubt" implies that the trier of fact should err on the side of acquittal rather than guilt. In signal detection terms this means the examiner should attempt to minimize type I errors, even if at the cost of more type II errors.

Figures 3 and 4 illustrates the implementation of the matrix shown in Figure 2 into a descriptive set of curves for describing bias known as receiver operator curve (ROC). These curves were originally developed by psychophysical experts for the military to determine how the role of a *criterion shift* (bias) can influence the relative proportion of misses versus false alarms (type II versus type I errors, as in Figure 2).

The following is a very simplified explanation of how bias can create a criterion shift in a military setting. If you had a gun and were staring into pitch black darkness while on guard duty, and no one had attacked the fort in the 50 years since it was built, your criterion (bias) might be shifted towards committing more type II errors; if an enemy suddenly attacked, you might be taken by surprise, since you weren't expecting the event to occur. On the other hand, if there had been a recent attack, you might be "trigger happy" and fire at anything that moved (make many false alarms.)

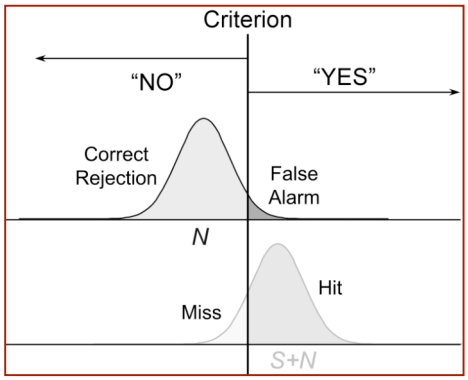

Figure 3*:* Effect of criterion shift on a distribution of correct identifications and eliminations.

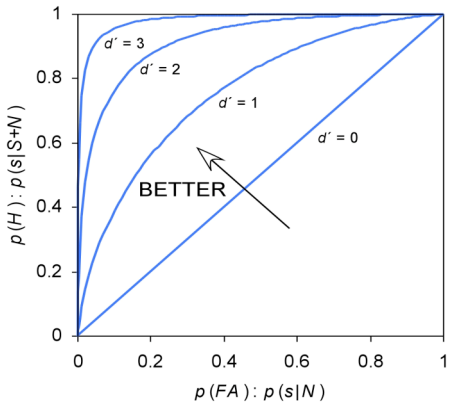

Figure 4: Receiver-operator curves (ROCs) for different values of *d'* (increasing sensitivity). The probability of false alarms (*p(FA)*) is plotted versus probability of hits (*p(H)*). *See text*.

Figure 3 shows a Gaussian distribution for both correct rejections ("noise", or N) and hits ("signal in noise", or S+N). The voice identification equivalent for S+N is an "identification" and for N, "elimination". There is an overlap between the curves that yields the proportion of false alarms (incorrect identifications) to misses (incorrect eliminations). The correct rejection and hit values on the *x* axis represent the level of accuracy possessed by a particular examiner; this is a measure of sensitivity known as "d prime", *d'*. Note how the vertical criterion line, when shifted to the right, increases the number of misses while decreasing the false alarms, and vice-versa when the criterion line is shifted to the left. This represents a shift in criteria. As with the soldier on guard duty, one's bias can cause the criteria to shift and thereby affect the number of Type I and Type II errors.

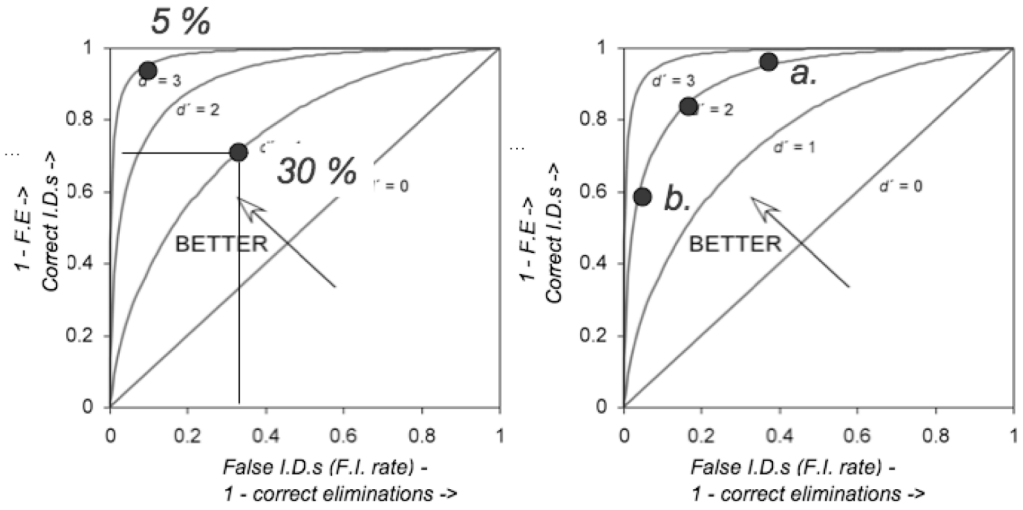

Figure 5. Left: criteria shift on a distribution of correct identifications and eliminations. Right: ROC curves with d' = 0, 1, 2, 3 (increasing sensitivity). *See text.*

Figure 4 shows a graph of several ROC curves for increasing levels of sensitivity (ability to make correct identifications and eliminations), plotted against different ratios of hits (*y* axis) versus false alarms (*x* axis). The diagonal line (*d' = 0)* represents guessing, or chance. The interaction of bias and criterion shift is used to illustrate their effect on false identification and false elimination error rates.

A completely unbiased examiner would have an equal proportion of false eliminations and false identifications, as illustrated by the two filled circles in the graph shown on the left side ROC curve plot in Figure 5. Note that these circles lie directly on the middle of the line (on the leftward diagonal). The filled circle labeled "30%" is an examiner with fairly poor sensitivity; and for 70% of his correct identifications, he makes 30% false identifications and by implication 30% false eliminations. The filled circle labeled "5%" is an examiner with excellent sensitivity; he makes correct identifications 95% of the time, and 5% false identifications.

The ROC curve on the right side of Figure 5 shows how bias influences an examiner with an equal level of accuracy, had there been no bias. Filled circle 'a' shows a criteria shift towards false identifications, where the examiner "loosens" their criteria to make a greater number of correct identifications but at the expense of increasing the false identification rate. (This is akin to the soldier in the previous example shooting 'wildly' into the night multiple times. There'll be more hits and more misses). Filled circle 'b' shows a criteria shift where the examiner adopts "tighter" criteria before declaring positive identification. There are fewer correct identifications (and correspondingly, more false eliminations) but there are also far fewer false identifications. Importantly, these curves indicate that, for a given sensitivity level, shifting towards "looser" criteria (circle A) yields slightly more correct identifications (10% improvement) but *doubles* the false identification rate (20 – 40%), while shifting towards tighter criteria *halves* the false identification rate, albeit with a reduction in correct identifications.

It is important to realize that "sensitivity" *d'* not only reflects an examiner "ability", but can also vary as a function of the quality of the data examined. In other words, one examiner with a highly consistent ability to make discriminations will make more errors if the quality of voice exemplars is diminished.

It should also be noted that by moving along a curve with a given *d'*, one does not observe more or fewer errors, but simply different ones. The only way to decrease the error rate is to increase *d',* in other words, improve the quality of the data and/or improve the ability of the examiner. In the interests of brevity, we have not expounded here on the issues of a priori probability or values and costs. Suffice to say that these issues, and their interaction with the criterion shift question described above, further reinforce the need for examiner protocols that mitigate inherent bias as much as possible. The following section will include references to these concepts as we address the rationale underlying such protocols.

4 PROPOSED VOICE ELIMINATION AND IDENTIFICATION PROTOCOLS

The following summarizes protocols detailed in reference [8]. As discussed above, examiner protocols for both identification and elimination comparisons should reflect an understanding of inherent bias.

4.1 Voice identification protocols

Historically, voice identification protocols have not appeared to recognize the issue of inherent bias. Although eyewitness identification protocols now routinely include some form of line-up, many voice identification practitioners continue to carry out examinations using only a single known (the suspect) to compare to the unknown.

It has been argued that the use of line-ups is financially impractical due to the increased cost of having to compare more exemplars than would be required without a line-up. While this is undoubtedly true, we find the alternative to be potentially significantly more costly in human terms. Another reason some examiners may object to a line-up is that it puts them at risk of making a proven false identification. Such an event is generally considered career ending. The notion that a human being who routinely makes subjective judgments as a part of his job, must never make an error if he is to keep his job, is ludicrous and reveals an inherent lack of understanding of the subjective decision making process.

Sadly, this unsound reasoning pervades our legal system. Perhaps spurred on by aggressive prosecutors, expert witnesses have often given testimony that implies that there is no chance that they could commit an error. Even fingerprint comparisons, once thought to be infallible, have been shown to have a strong subjective component when carried out using latent prints. There are very few forensic comparison techniques that do not include a subjective component when used in real life situations.

Consequently, we propose that forensic voice identification examiners require some form of line-up whenever it is possible. If a line-up is, for some reason, not possible, the examiner should inform the client that any result so obtained should be considered "investigational" and should not be used to advance a prosecution.

Generally, there are three types of line-ups that can be employed. The first, and least desirable, will be called a *closed set line-up*. Here the examiner is given the recording of the unknown voice and a recording with N known voices saying the same words and phrases as the unknown. The size of N should be at least 4 or 5, and care must be taken to assure that the exemplar made by the suspect does not stand out as different in any way from those made by the other known voices.

Although this protocol is far better than the simple A-B comparison, it does not totally eliminate the potential for bias. The examiner inevitably has some reason to believe that he is aware of the *a priori* probability that the guilty party is one of the N known voices. Earlier it was indicated that the notion of *a priori* probability would interact with the problem of eliminating bias, and now it can be seen that the very fact that a law enforcement organization has apprehended the suspect is bound to affect the examiner's opinion as to the likelihood that the guilty party is present in the line-up. This opinion, no matter how weak, is bound to have an affect on the examiner's criterion, perhaps causing him to match a known voice to the unknown with a slightly looser "similarity" tolerance than he might otherwise have. On the other hand, the cost of a mistake would hopefully have the opposite affect on his criterion, therefore pushing his criterion back to a more conservative position. The key to insuring that this protocol has the proper effect is making sure that the suspect's exemplar in no way stands out from the rest. If this is done, it is unlikely that an examiner will be consciously or unconsciously inclined to risk a "public" false identification unless he is very strongly convinced of the soundness of his decision. In other words, the examiner's criterion will be conservative.

The second protocol is known as a *closed sequential line-up.* Using this protocol there are once again N known voices, and once again the suspect's exemplar should not stand out from the other knowns. However, because the protocol is *sequential*, the examiner will not be able to simultaneously compare all of the known to unknown exemplars before making a decision. In this protocol the examiner is presented with the first of the N knowns (in random order) to compare to the unknown, and he is required to make a match/no-match decision on that pair before he examines the second known. This procedure continues until he either makes a match or runs out of knowns. Although he knows how many knowns are coming, the examiner is forced to treat each comparison as a simple A-B discrimination task, and he can make no reasonable inference concerning the *a priori* probability that the suspect is present in any given pair. Even if one of the known exemplars seems to stand out as different from the rest, it is much less likely

that such differences will be well remembered when previous exemplars can not be referred to at will.

The last protocol, known as an *open sequential line-up*, is really only slightly different from the second. As indicated in the name, the only difference is that the examiner doesn't know the value of *N*. Although this protocol represents a more "pure" version of the sequential line-up, we don't believe that its effectiveness will be sufficiently greater to justify the added effort required to carry it out.

In summary, voice identification protocols should take into account the fact that bias in subjective tasks can only be avoided through careful design. When performing such tasks, an examiner's expertise is not a shield against bias. Only by assuring that each fundamental discrimination task is presented to the examiner in total isolation from the facts of the case, especially of the knowledge of whether the known is the suspect, can the trier of fact be assured, not that the resulting decision is correct, but that the examiner has imparted his expertise in an impartial manner. The weight to be given to such a decision will depend on the expert's ability to explain the basis of his decision and on the other side's ability to effectively critique that testimony.

4.2 Voice elimination protocols

With voice elimination comparisons, the use of a lineup does not mitigate the possible effect of bias. In an identification scenario an examiner is being asked by law enforcement individuals or by an attorney to determine whether a message of unknown origin was made by a specific known individual. In that context, the examiner "succeeds" if he provides evidence that the known (suspect) did indeed make the unknown message; hence, the need for a line-up. In an elimination scenario an examiner is being asked by an attorney to determine whether or not his client uttered an incriminating message. Here, the examiner "succeeds" if he provides evidence that the client did not make the incriminating message. In this case, a line-up would be of no value, since the effect of bias, e.g., knowledge of the facts, desire to please the hiring attorney, etc., would move the examiner's criterion to a point that would tend to have him reject all the line-up voices. Such a result could be easily attacked by the opposing counsel as being obviously self-serving.

In order to address this issue, a protocol has been designed which, while not eliminating the potential for bias, at least enables the examiner to carry out a 'hypothesis testing' approach to the task that may provide credible evidence of value in the client's defense.

The strategy behind this protocol, is to compensate for inherent bias by having the examiner obtain from the suspect, exemplars that allow for exceptional comparability between the spectrographic patterns of the known and unknown exemplars. In order to execute this protocol properly, the examiner must take full advantage of the (usually) cooperative nature of the suspect in this kind of situation. This design is meant specifically for an attorney who could utilize potentially exculpatory evidence, and it is premised, for reasons that will become obvious, on the understanding that the possible decision outcomes are elimination or non-elimination, but not identification.

As with any protocol that involves subjective decision-making, the examiner should be careful to avoid acquiring and knowledge of the facts of the case in order to minimize the bias that can result from such knowledge.

In order to produce an appropriate exemplar for this protocol, the examiner should:

(1) Create a computer file of recorded items, each item consisting of a phrase excerpted from the unknown message.
(2) During the exemplar creating session, play the first excerpted phrase and have the suspect repeat the phrase in a manner that conforms to the rate, intonation and emphasis patterns of the unknown, while still ensuring that the suspect stays within his or her normal speech repertoire.
(3) Monitor the suspect's utterances closely and, if necessary, "coach" the suspect until the desired result is obtained. It should be remembered that whether or not the suspect actually made the unknown message, what he or she is now being asked to do is a form a form of "acting," and as such may be difficult for most speakers to carry out. Patience is often required.
(4) When the suspect has repeated the phrase in an acceptable manner, steps 2 and 3 should be repeated for the rest of the phrases.
(5) If the suspect seems to be consistently unable or unwilling to meet the examiner's standards for repeating the phrases, the session should be discontinued and the requesting attorney should be informed.

The examiner then compares the unknown utterances to the ones cooperatively produced by the suspect. If the comparison of the known and unknown exemplars shows some compelling similarities or does not show consistent and widespread dissimilarities, the decision

outcome is that the voice could not be eliminated on this basis, and the procedure is ended. If, on the other hand, the reverse is true, there is yet one further comparison to be made to strengthen confidence that an elimination decision is warranted. After the passage of about a week, the subject should be asked to read the same unknown's phrases from a written transcript.

By comparing the two known exemplars that were taken a week apart, and spoken in two different ways (albeit the reading from a transcript a week later can be expected to retain some of the "flavor" of the coached rendition given earlier), the examiner can make an estimate of the intra-speaker variability inherent in the suspect's voice. If the two known exemplars exhibit low intra-speaker variability, then the examiner can be somewhat reassured that the dissimilarity between the unknown and the suspect's first exemplar was not a result of inherently large intra-speaker variability. On the other hand, if substantial variability is exhibited by the two known exemplars, the decision outcome must be considered to be inconclusive.

Although the credibility of evidence reported using this protocol may ultimately depend upon the perceived integrity of the examiner, the protocol does produce a substantial amount of evidence to support the findings, some in a form that can be appreciated by the layman's ear.

Some speech experts in the scientific community have shown support for the power of this type of elimination protocol.

> ...it would certainly count as a **cogent argument for exclusion** to be able to say that, even though a suspect mimicked the questioned speech, unaccountable differences persisted (P. Rose, *Forensic Speaker Identification* (2002)).

Manfred Schroeder, famous for many developments at Bell Laboratories, has stated perhaps more bluntly that

> In forensic applications, voiceprints are especially useful in eliminating a suspect because, with a given vocal apparatus, some suspects could not have possibly produced the recorded utterance. But the general applicability of voiceprints in criminal trials remains doubtful. (M. Schroeder, *Computer Speech* (1999)

5 CONCLUSION

Forensic speaker identification, as currently practiced, is grounded on fundamentals of speech science, and its use can be justified on the basis of experimental results, but is not infallible. Reliable error rates for its use in the forensic setting are not yet available. The audio forensic community must be aware of, and must deal with, the history of overstated and unproven rates of accuracy and reliability attributable to Kersta and Tosi. Given suitable quantity and quality of voice exemplars, combined with conservative criteria that results from scientifically appropriate protocols, forensic voice examiners can contribute useful probative data in legal situations if limitations of the technique involving error rates and control of bias are openly addressed.

REFERENCES

[1] L. G. Kersta, "Voiceprint Identification", *Nature* 196 (1962).

[2] O. Tosi, H. J. Oyer, W. Lashbrook, C. Pedney, J. Nichol, W. Nash, "Experiment on voice identification", *Journal of the Acoustical Society of America* 51:2030-43 (1972).

[3] R. H. Bolt, F. S. Cooper, E. E. David, P. B. Denes, J. M. Pickett, K. N. Stevens, "Speaker identification by speech spectrograms: a scientists' view of its reliability for legal purposes", *Journal of the Acoustical Society of America* 47:57-612 (1970).

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

[5] F. R. Clark, R. W. Becker, J. C. Nixon, "Characteristics that determine speaker recognition", SRI (Stanford Research Institute) contract report AF-19 (628)-3303 (1966).

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

[7] P. Ladefoged, in *Echoes,* newsletter of the Acoustical Society of America (2004).

[8] J. S. Gruber, F. T. Poza "Voicegram Identification Evidence", *American Jurisprudence. Trials,* 54 (1995).