Harry Hollien
University of Florida*

UDK 81'342:343.98

ABOUT FORENSIC PHONETICS

INTRODUCTION

Forensic Communication is a discipline within the Forensic Sciences. It is made up of three related sub-disciplines, all of which are structured to meet the relevant needs of Criminal Justice, Judicial and Intelligence agencies. The three areas are:

1) **Forensic Linguistics**. This area, including Psycholinguistics, is the one with which you are most familiar. It targets language (written or spoken) which is analyzed to determine authorship, the intent of the individual, deception and so on. The area also includes speech/language decoding, a task shared with Forensic Phonetics.

2) **Forensic Psychoacoustics** is the second area; it relates to human hearing or audition. In this case, the analyses involve heard signals and their effect (acoustic, perceptual, neural) on individuals and their behaviors.

3) **Forensic Phonetics** focuses on the analysis of spoken communication for the purposes cited above. It includes speaker identification, enhancing and decoding spoken messages, analysis of emotions in voice, authentication of recordings and related. Forensic Phonetics is what this article is all about. Hence, it would appear useful to start with definitions.

Definition

First off, it is important to provide a *general* description of Phonetics, and just why professionals of this sort are able to carry out the functions/operations to be described below. Obviously an Experimental Phonetician is interested in, and grounded in, the acoustical, physiological and perceptual study of human speech. To conduct research, to teach and to practice in these areas, he or she must be trained in both the central issues (including Phonology to some extent) plus a number of complimentary disciplines. Most notable among them are Acoustics and relevant Engineering specialties. Phoneticians must also have fundamental expertise in human behavior (ranging from Psychology to Physiology) and in hearing (especially Psychoacoustics and Hearing Neurophysiology). So too must they be able to develop specialized processing procedures and/or adapt available, but relevant, systems to their needs. Finally, a *Forensic* Phonetician is one who is both interested in, and trained in, a number of areas within the Forensic Sciences (Hollien 2008).

* *Author's address:* Institute for Advanced Study of the Communication Processes, University of Florida, Gainesville, Florida 32611, USA. Email: hollien@ufl.edu

The Forensic Phonetics area itself consists of two major elements. One involves the analysis (usually electro-acoustical) of those speech signals which have been transmitted and stored; the other is that of analyzing the communicative acts themselves (Hollien 1990). The first of these two domains addresses the enhancement of speech intelligibility, speech decoding (including accuracy of transcripts), the authentication of recordings and the like. The second area involves issues such as recognition of speakers from their voices, identification of the health, emotional or psychological states of the talker and the analysis of speech for evidence of deception. Of course, Forensic Phonetics also interfaces with a number of other specialties. Forensic Linguistics is one (re: language and its analysis), as is Forensic Psychoacoustics (for auditory-perceptual issues) and Audio Engineering (especially with respect to nonhuman signal analysis). Since these are tangential, they will not be discussed in this brief review. What will be featured here is 1) the integrity of captured utterances (and related), 2) the accuracy/completeness of messages, 3) the identification of the human producing the utterances and 4) the identification of the various behavioral states (including stress and deception) that they are experiencing.

1 SIGNAL ANALYSIS

The *analyses* of electro-acoustical transmissions or stored speech signals are carried out for a number of purposes. That is, good recordings can be of great importance to groups such as law enforcement agencies, the courts and the intelligence services. Indeed, the effectiveness of most of these agencies would be reduced by a magnitude if suddenly they could no longer utilize recorded spoken intercourse and related events for surveillance, interrogation and/or other operations. Indeed, it is quite possible that analysis of such information ranks among the more powerful of the tools investigators have at their disposal. Yet the detection, storing, and analysis of messages, events and information has become so common that many investigators, attorneys and agents tend to overlook the myriad of problems associated with them.

Moreover, even currently, the recordings obtained during surveillance and from related activities are rarely of studio quality and the utterances found on many of them will have to be enhanced before they can be understood. Other than the problems induced by speakers (examples: overlapping talkers, effects of stress, drug/alcohol usage, etc.), the primary sources of speech degradation are related to input noise and distortion. Both can result from inadequate equipment, poor recording techniques, operator error, or the acoustic environment in which the recordings were made. In turn, these problems create challenges. The first of these involves understanding what the recordings contain.

2 SPEECH ENHANCEMENT

As has been implied, one of the important areas within Forensic Phonetics is that of enhancing intelligibility of recorded speech – and also, of course, the development of speech decoding techniques (Dean 1980; Hollien 1990; 1992). As stated above, recordings made for these purposes are often of poor quality, with such degradation resulting from: 1) reduction of frequency bandwidth, 2) addition of noise, 3) reduction

of the energy level, 4) spectral or harmonic distortion, 5) inadequate transmission links or coupling, 6) inadequate pickup transducers (microphones, "bugs", telephones) and so on. Masked or degraded speech also can result from environmental factors such as "hum", the wind, vehicle movement, fans/blowers, clothing friction, other talkers, music, etc. (cf. Dallasarra et al. 2010 re: complications). Whatever the causes, all detrimental events must be identified and compensatory actions taken. The remedies here include several types of electronic filtering, noise elimination (computer) programs (examples: Lee 1998; Wan 1998) and speech decoding.

The initial step in the enhancement process is to protect the original recording by making a digital copy. It is not desirable to work on an original because repeated playing could result in signal deterioration or other damage. It is not difficult to imagine the problems which would arise if important evidence were thusly destroyed or compromised. Second, the examiner should listen to the recording a number of times before attempting to process it. This procedure permits development of a log plus a good working knowledge of the recording's contents – plus, of course, information about interference and degradation. It is by this process that various problems and remedial techniques can be identified. One useful analysis approach is to digitize problem areas and apply software which can permit "visualizations" of the relationships within the signal (spectra or waveforms, for example). These graphs also can be used to obtain a great deal of quantitative information about the signal, and thus aid in identifying relevant speech sounds, words, or phrases.

The first of the remedies cited involves filtering by use of relevant analog or digital equipment and/or computer software. If there is a substantial amount of noise at either (or both) the extreme lows and highs, speech may often be enhanced by band-pass filters (frequency range 300–3500 Hz). Second, if spectrum analyses show that a relatively narrow band of energy exists at a specific frequency, a notch filter may be employed. Third, comb filters (especially programmable ones) can provide an assist when noise exists *within* the frequency range of the speech. They can be used to continuously modify the spectrum of a signal by selectively attenuating the contaminating frequency bands. Of course, filtering here must be carefully applied so as not to remove necessary speech elements along with the unwanted sounds. Finally, these procedures are best carried out with systems that have been expressly developed to compensate for noise problems.

Since binaural listening is particularly helpful, the equipment employed should permit stereo listening (Bronkhorst 2000). Further, use of variable-speed recorders can provide an assist. One type involves a manual increase or decrease of recorder speed while it compensates for parallel distortion of the internal speech signal. Finally, other useful techniques include filling in short dropouts with thermal noise, deconvolution, bandwidth compression, and so on.

Speech Decoding

The second phase of the process of extracting utterances and messages from difficult recordings involves speech decoding. That is, if the still degraded speech is to be

comprehended, a remedial program involving formal decoding procedures must be initiated. This is an instance where Linguists and Phoneticians often work together.

As stated, Forensic Phoneticians are among those specialists who can deal effectively with problems in this area. They can be especially adept at dealing with things such as 1) voice disguise, 2) dialects (and/or foreign languages), 3) very fast speech, 4) multiple speakers and 5) the effects on intelligibility of talker stress, fatigue, intoxication, drugs, etc. (Hollien 1992).

To initiate the decoding process, it is necessary to repeatedly monitor the entire recording and apply any intelligibility enhancing techniques still necessary. Next, it is best to first focus on the relatively easy-to-decode sections. As is well known, speech/language is quite redundant, words have an effect upon one another (coarticulation), context aids intelligibility and determination of one word (or several) in a multi-word series (context analysis) often can aid in the correct identification of the other words. Hence, it is wise to start with the easier elements and work up to the more difficult.

Decoding also can be aided by the use of graphic displays of the speech signal. Particularly useful here are sound spectrograms of the time-frequency-amplitude class. Displays of this type can be used to estimate the vowel used (from its formants), the consonant used by analysis of its acoustic patterns, or by comparison of their energy distributions to identifiable patterns, and so on. Other techniques also are available.

Basic knowledge about, and skills in applying, many of the phonetic and linguistic realities of language and speech are also quite helpful. For example, a good understanding of the nature and structure of vowels, consonants and other speech elements is fundamental to good decoding. So too is an understanding of word boundaries, word structure, dialect, linguistic stress patterns and the paralinguistic elements of speech (i. e., fundamental frequency, vocal intensity, prosody, duration, etc.). Especially useful here is a thorough understanding of coarticulation and its interactive effects on words and phrases. In short, speech/language systems can be effectively employed in support of the decoding process.

The reporting structure and the mechanics of decoding must be as highly organized as is the decoding process itself (P. Hollien 1984). That is, it is not appropriate to simply write down the heard words and phrases, it is also necessary to explain what went on during the recording period and what may be missing. Codes are useful here, but they must be easily understood and used consistently and systematically. Example: inaudible and/or unintelligible sections must be described in detail. Note the following: “But I will... (plus) 10 sec. unintelligible speech by talker A” or “6–8 words of inaudible response, the two center of which were ‘hit me’”. Another requisite is to include descriptions of all the events which occur; i. e.: “Talker C: Help me. footsteps, door closing, two gunshots, loud thump.” Once the decoding basics have been completed and confusions resolved, the text can be structured into final form, carefully re-evaluated for errors, and submitted.

3 AUTHENTICATING RECORDINGS

As is well known, certain specialists exist who can make any individual appear to say almost anything they (i. e., the operators) wish – that is, if a quantity of the target

person's recorded speech, plus good equipment and the appropriate technical knowledge all are available. This possibility can be serious (re: investigations, trials) when someone involved claims that a particular recording was "doctored" and, hence, some or all of the information it contains is false. Thus, before a recording can be considered "authentic", it must be shown NOT to have been modified in some manner. The integrity of a recording also may be examined for other reasons.

A definition of authenticity follows. To be valid, a recording must include all of the events that occurred, and nothing can have been added, changed or deleted, during the recording period or subsequently (Hollien 1977; 1990). It also must be stressed that, no matter how pristine its source (individual or agency), if the recording is to be properly authenticated, it must be analyzed thoroughly, impartially and ethically (Aperman 1982; Hollien 1977). To do this, the Forensic Phonetician must apply knowledge about the intricacies of the speech act, technical information about the modern processing/storing of speech and appropriate electronic/computer analysis technologies. Moreover, since a number of different classes of equipment can be used for speech (included are analog, digital and video recorders – plus computers), evaluations must include physical examinations and signal assessment.

Authentication procedures commence with a log being established, high quality copies made and the recording listened to for familiarization purposes. Chain-of-custody and identification marks should be checked.

The Physical Examination

The first procedure is to determine if duration is correct. Any external timing information available should be employed as should the length of the tape (if that medium used) and/or the time of the recording. Timing signals also are useful; they can be found on many types of recordings. Announcements of time, images of clocks (on video) and related, should be checked. The second examination is focused on housing – boxes, cassettes, reels – which are examined for modifications, damage, etc. Tapes must be examined for splices; discs (and related) for interrupts. If any are found, editing may be present and, hence, the recording must be appropriately examined. Next, the unit upon which the recording purportedly was made must be examined and tested for evidence of originality. The procedures here involve making a series of test recordings (of the cited unit) under conditions of quiet, noise and speech. They should be repeated while all equipment switches are operated serially. If the recording is an original, the electronic signatures found on it will be the same as those produced by the tested unit. Also, information of stop-and-start activity, over-recordings and so on must be obtained. Finally, the "tracks" made by the recorder drive mechanism, the magnetic recording patterns or codes all should be checked. These processes can be complex and they vary among various classes/types of recordings (Brixen 2007).

Laboratory Examination

This phase involves an intensive analysis of the signals captured on the recording. To do so, high quality copies of the "originals" must be made and used in laboratory

tests. First, the recording is listened to *in detail* and all questionable events logged. Questionable events include clicks, pulses, thumps, rings, crackles, etc. They also may involve loss of signal, abrupt shifts in ambient noise, inappropriate noises as well as breaks in (or difficulty with) context, word boundaries and/or coarticulation. Analysis protocol for coarticulation is of significance since it transcends operation of different types of recorders. That is, discontinuities resulting from breaks in the coarticulatory stream can often be heard. Such observation leads to application of spectral analysis techniques which, in turn, will permit the study of the relevant phonemes and their neighbors (in order to resolve the possibility of a modification).

Next, it is necessary to systematically identify and/or explain the causes of these (observed) events. Some will be innocuous. Examples: a telephone disconnect, a door closing, operation of interface equipment. Such events can be "suspicions" but may not reflect modifications. On the other hand, any of them could signal an attempt to alter the recording.

The procedures by which questionable events are evaluated will vary; time-amplitude and/or spectral analyzes are examples.

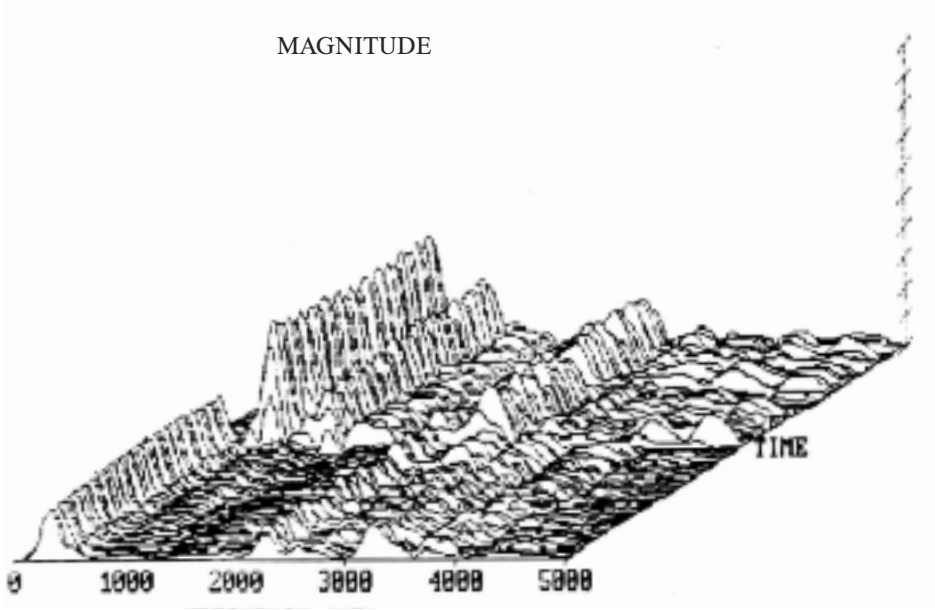


Figure 1: A continuous frequency-amplitude spectra (i. e., a waterfall plot) of a rerecorded splice or digital interrupt.

In Figure 1, an abrupt change can be seen in a waterfall type spectrum. It signals that the recording probably has had a section removed. In any case, every questionable event must be identified and explained. This process is a challenging one and, while

most of the suspicious events will be found to be "innocent", it is important not to miss evidence of manipulation.

However, please be advised that, even if it is demonstrated that the recording has been altered, the reason for modification cannot be specified. It could have been either accidental or intentional; the examiner ordinarily cannot know which is the case without external evidence.

Second, it has been suggested that the authentication of digital recordings is more difficult than that for tapes. In response, certain specialized techniques have been developed (Brixen 2007) and many of the approaches described above have been successfully adapted. For example, a physical examination still can be conducted. Moreover, dialog and/or coarticulation evaluations can be effective ways to analyze speaker/speech characteristics, as can (linguistic) topic analysis. In short, while the examinations here usually involve quantitative signal analysis, many should still parallel those used with analog recordings.

Third, it can be said that, in some ways, the evaluation of video recordings can be a little easier than that for audio-only. For one thing, the audio portion can be assessed (as above) in parallel with the video but, even more importantly, modifications (if made) often can be "seen" on the video. Nonetheless, the initial phase here should parallel techniques used for audio evaluations – starting with the physical and signal evaluations described above. The next stage would be to directly assess the video channels. A successful evaluation here requires the availability of appropriate equipment and the use of, at least, some specialized techniques. First, it is important to examine and test the equipment on which the original video recording was made as there is at least some evidence that electronic signatures unique to each unit can be recovered. In addition, if the noise that can be "seen" on a previously unrecorded video is different from that which is generated by the target unit's erase system, a spectrum analysis can be employed to quantify and assess these "impressions". In addition, when a second video sequence is recorded over the initial one, a series of (very brief) irregular flashes may occur. And ... abrupt "turn ons" and "turn offs" (of the video picture) also suggest that editing may have taken place. Perhaps the most relevant evidence of all is that the visual sequence often will shift abruptly when modifications have been made. Finally, a specialized device that can be used to evaluate video recordings is easily procured. It has a capability which somewhat parallels the frame-to-frame viewing systems utilized in motion picture film processing. While a video is not created in a strict frame-to-frame mode, the observation of "stills" – coupled with very small advances of the sequence – often can reveal interruptions to the video portion of the recording.

CONTENT ANALYSIS

4 SPEAKER IDENTIFICATION

Speaker Identification (SI) should not be confused with Speaker Verification (SV). The second of these two processes (verification) results when a known and cooperative person *wants* to be recognized from his or her speech. Thus, they will provide speech

samples in order to permit construction of "reference sets". The identity of the target individual is later verified when a new speech sample is compared to these referents and the decision made: correct speaker or imposter? While speaker verification techniques are useful (in industry, government, prisons), they are only occasionally employed in law enforcement. Ironically, however, since speaker *identification* is a far more difficult task, any technique that will work for identification purposes will work even better for verification (Hollien/Harnsberger 2010).

Speaker identification (SI) is a process where attempts are made to identify an individual from his or her speech when that person's identity is *not* known and when anyone within a relatively large, open-ended population could have been the talker (Hollien 2002). This problem, while of considerable importance, is a difficult one to solve. First, in the forensic (or intelligence) situation, all sorts of system and speaker distortions are present. Samples are rarely contemporary; speakers are usually uncooperative; equipment often is poor, and decision criteria are difficult to establish. Moreover, the "sets" of talkers always are open-ended (i. e., one is never totally sure that the criminal is among the suspects).

Considerable research has been carried out on SI with investigations focused on three areas: 1) aural/perceptual approaches (including earwitness identification), 2) "voiceprints" and 3) machine-computer approaches. Only the first and third of these will be reviewed here as "voiceprints" have been shown to be totally inadequate and, thus, have been discarded. Indeed, it would not be appropriate, in this short review, to describe an invalid technique.

Aural-perceptual SI approaches take two forms; the first of these involves ear witnesses. That is, some courts permit a lay witness to make an identification but only if they are able to satisfy the jurist that they "really know" what the speaker sounds like. The more common form here is where lay individuals, who can identify the person from a "voice parade", are permitted to testify. In the second instance, qualified experts are permitted to render opinions. In such cases, a sample of the unknown talker's speech (evidence recording) must be available – so too must an exemplar of the suspect's voice. An examination is carried out by the expert where it is determined if the two recordings contain, or do not contain, the voice of but a single person.

Earwitness Identification

Basically, an earwitness lineup or voice parade is defined as a process where a person who has heard, but not seen, a perpetrator attempts to pick his or her voice from a group of voices. It ordinarily is conducted by law enforcement agents. They have the witness listen to the suspect's exemplar embedded in a group of 3–6 similar samples produced by other people. Good speaker identification accuracy can be possible if: 1) the listener remembers the talkers' speech patterns, 2) reasonably good samples are available, and 3) negative emotions do not interfere (Hollien 2002; Yarmey 1995). As with eye-witness identifications, the witness is required to observe the "group" and choose which one is the perpetrator. However, this approach — especially when pat-

terned after eyewitness lineups – has come under fire (Broeders 1996; Hollien/Majewski 2009). Although, there is little question but that both these types of identification can be quite important to the conduct of criminal investigations and/or trials, *neither* of them is as robust as would be desired. For example, there is a large literature about *eyewitness* examinations and it has been found that witnesses with poor eyesight (Loftus 1979), inadequate memory, and so on can make serious mistakes. Also, poor line-of-sight or bad lighting, plus very brief encounters, create problems. Perhaps of most concern here, is where DNA evidence has demonstrated that certain individuals who were convicted on eyewitness evidence actually were innocent (Schuster 2007). On the other hand, upgrading factors *can* include: 1) race, 2) sex, 3) attractiveness, 4) distinctive features, 5) age, 6) nature of presentation and so on (Cross et al. 1971; Wells 1993). Thus, it is recognized that reasonably accurate assessments often can occur if conditions are favorable (Wells/Olsson 2003).

Secondly, eye and ear witness approaches, while seemingly similar, are quite different in many ways; mostly with respect to: 1) how different types of memory is processed, 2) how a *voice* analysis is different from visual approaches, 3) the differing ways by which fear or arousal can affect the processes and so on. Yet other problems also exist for both: 1) latency is always present, 2) witnesses' emotions may be a factor, 3) the witness may feel the suspect is guilty simply because he is being tested, and so on (Aarts 1984; Hollien 2002; Yarmey 1995). However, if competent personnel are used, if the lineup is well organized and appropriately structured, experiments have shown that the results can be both robust and reasonably accurate (Hollien et al. 1983).

Basically, two different earwitness lineups can be employed (Hollien 1996). The first, the "Multiple Choice" approach, is one where the suspect's voice, plus samples of foil talkers, are heard as a "set" and the procedure is repeated a number of times. Specifically, a speech sample produced by the suspect is embedded within a group of samples (each set in a different order) spoken by a group of four foils or "distracters". The sets are played 20–25 times and the witness attempts to pick out the perpetrator from each set. The results are then checked to see if the witness has made an "identification".

The second, or "Comprehensive Review" procedure, was first proposed by Major (Major/Komulainen 1989); it has also been refined and used by others (Hollien 2002; Nolan 2003). It can be best understood by consideration of Figure 2.

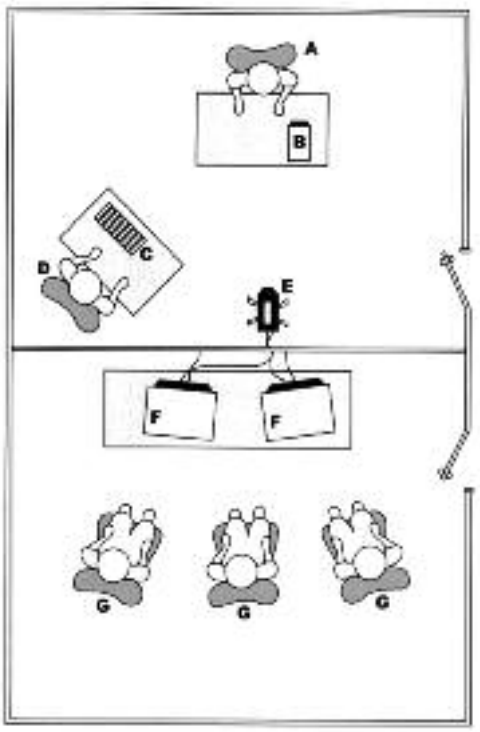


Figure 2: The "Comprehensive Review" approach to Earwitness Identification.

This graphic display demonstrates the positions of: **A**. the witness, **B**. the CD playback, **C**. the recordings, **D**. the administrator, **E**. a videocam, **F**. the TV monitors and **G**. the observers. Here, the witness is provided numbered recordings containing voice samples produced by the suspect and each of the foils. The witness calls out a number and the administrator (who does not know which recording contains the suspect's voice) gives that recording to him or her. The witness plays it; then hears, in turn, the rest of the recordings. Once all of them have been heard, he or she is permitted to re-listen to any of them. Ultimately, the witness is asked if any of the voices is identifiable and if so, which one. Of course, he or she is not required to make a selection if unable to do so.

The standards for administration of these earwitness lineups must meet the highest of standards. They are summarized as follows:

1. The Comprehensive Review approach is recommended.

2. A clear and complete set of instructions must be provided all parties.

3. Comprehensive records must be kept.

4. Only high quality recordings are to be used. However, protocol should take into account how the suspect was heard (free field, telephone, etc.; Nolan et al. 2008).

5. Samples and their presentation should be carefully controlled. They all should be structured/presented in the same manner and checked by means of mock trials.

6. While a witness should be briefed about the procedure, no information about his or her performance should be provided until all aspects of the lineup are completed.

7. No partiality toward either the witness or the suspect should occur. The entire lineup should be observed (remotely) by individuals representing the witness, the suspect and any relevant agency(s).

8. The witness should be asked – but not required – to make an identification.

Aural-Perceptual SI

A substantial amount of research re: aural-perceptual SI has been carried out. Much of it supports strategies employed by Forensic Phoneticians who conduct structured, speaker-specific, aural/perceptual identifications of talkers. Indeed, it can be argued that aural/perceptual approaches by humans – or, at least, by trained specialists – are the most accurate of all available methods. However, it is important that these individuals 1) are of high talent, 2) assess "natural" speech characteristics, and (most important) 3) are qualified Phoneticians cross-trained in Forensics. Indeed, a number of these specialists have independently demonstrated valid analysis techniques and excellent identification rates (Hollien 2002). However, their approaches vary. Many attempt to determine a talker's identity by assessing speech segmentals. Such procedures are usually successful if a robust set of processing criteria are established and they are well trained/experienced. But, an even more powerful technique is one that is also rigorously structured but is focused on the suprasegmental assessments of voice and speech (i. e., prosody, fo, intensity, and voice quality) which are supplemented by segmental analyses.

It also should be noted that a rather substantial number of the AP SI studies have been directed at evaluating the human auditory and cognitive processing of heard signals for forensic purposes. Almost all of them have been based on subjects having but brief encounters with very limited stimuli. Consider how difficult it is for a listener to identify a person "known" to them only from hearing a brief utterance (by the speaker) which is then embedded within a group of voice samples spoken by other individuals. Many of the experimental tasks upon which research of this type was based are as challenging as that, few are markedly easier. Yet it is clear that many of these Forensic type experiments demonstrate just how discriminative the auditory mechanism is and how sensitive it can be in identifying speakers – even within the sharp limitations of the Forensic model.

An experiment by the present author and his associates (Hollien et al. 1982) provides pivotal information here. Three groups of listeners were used. Group 1 consisted of individuals who knew the talkers very well, Group 2, others who did not but who received at least two hours of training in recognizing their voices and Group 3, one that

neither knew the speakers nor the language spoken (but who also were briefly trained). The talkers were 10 adult males who produced phrase-length samples under three speaking conditions; 1) normal speech, 2) speech uttered during shock-induced stress and 3) disguised speech. The listeners heard a recording of 60 samples of the 10 subjects presented randomly (each of the three speaking conditions included twice) and had to name each talker. The results are best understood by consideration of Figure 3.

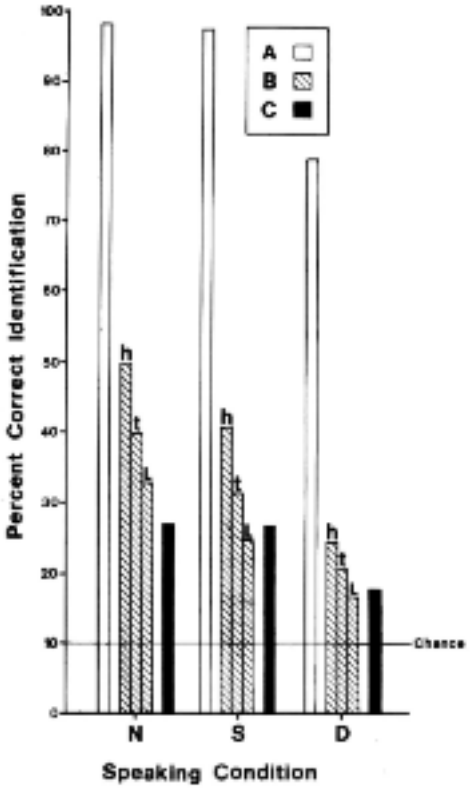


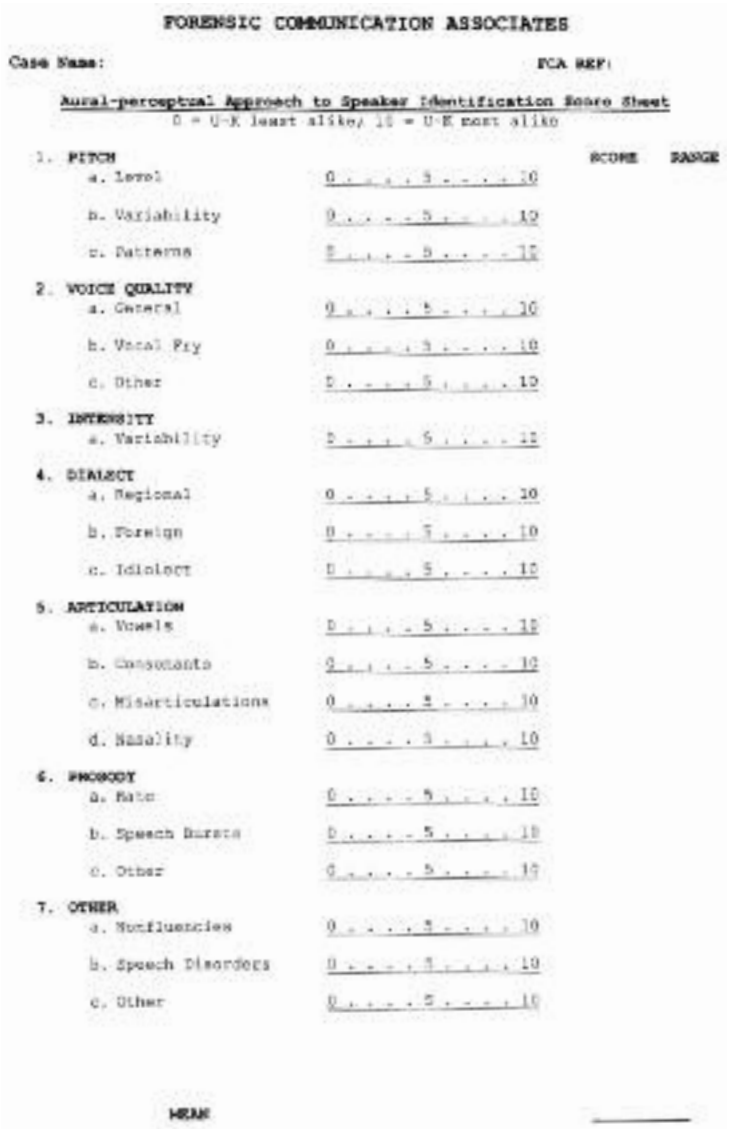
Figure 3: Speaker Identification (SI).

This figure provides data relative to the correct identification of 10 talkers speaking under three conditions: normal (N), stress (S), and disguise (D). Listener group **A** knew all of the talkers very well; groups **B** and **C** did not know the talkers but were trained to task; group **C** did not know English.

As may be seen, the accuracy of the listeners who knew the speakers approached 100% – and for both normal and stressed speech. Further, they could tell (about 80% accuracy) who the talker was even when disguise was attempted. The university students did not do as well, but their native processing abilities and the limited training they received, allowed them to succeed in correct identifications from double chance to four times better. Even a general population of non-English speaking Poles ex-

ceeded chance – and for all conditions. These data can be considered yet more impressive when it is remembered that the presentation procedure was extremely challenging. That is, all 60 samples were presented in a single trial and listeners had to rapidly identify the talker, find him on a list of names and, then, place a corresponding number on an answer form before the next sample was presented.

This, and many other studies, have led Forensic Phoneticians to develop structured aural-perceptual SI systems. One (see Figure 4) will be described here (Hollien 2002).



FORENSIC COMMUNICATION ASSOCIATES

Case Name: FCA REF:

Aural-perceptual Approach to Speaker Identification Score Sheet

0 = U-K least alike; 10 = U-K most alike

		SCORE	RANGE
1. PITCH			
a. Level	0 . . . . 5 10		
b. Variability	0 . . . . 5 10		
c. Patterns	0 . . . . 5 10		
2. VOICE QUALITY			
a. General	0 . . . . 5 10		
b. Vocal Fry	0 . . . . 5 10		
c. Other	0 . . . . 5 10		
3. INTENSITY			
a. Variability	0 . . . . 5 10		
4. DIALECT			
a. Regional	0 . . . . 5 10		
b. Foreign	0 . . . . 5 10		
c. Idiolect	0 . . . . 5 10		
5. ARTICULATION			
a. Vowels	0 . . . . 5 10		
b. Consonants	0 . . . . 5 10		
c. Misarticulations	0 . . . . 5 10		
d. Nasality	0 . . . . 5 10		
6. PROSODY			
a. Rate	0 . . . . 5 10		
b. Speech Bursts	0 . . . . 5 10		
c. Other	0 . . . . 5 10		
7. OTHER			
a. Nonfluencies	0 . . . . 5 10		
b. Speech Disorders	0 . . . . 5 10		
c. Other	0 . . . . 5 10		

MEAN _____

Figure 4: Response form for recording scores resulting from a structured aural-perceptual speaker identification evaluation.

In Figure 4, each parameter is rated on a continuum ranging from 0 (definitely two individuals) to 10 (production by a single person). The resulting scores are summed and converted to a percentage which, in turn, can be viewed as a confidence level estimate.

This SI system involves obtaining speech samples by the unknown speaker (evidence recording) and the known talker (exemplar) and placing them in pairs on good quality recordings. The pairs are played repeatedly and comparisons are made of one speech parameter at a time. Each characteristic (pitch patterns, for example) is evaluated continually until a judgment is made. The next parameter is then assessed and the process replicated until all possible comparisons have been completed. Subsequently, decisions are made and a confidence level generated. The entire process is repeated, independently and 1–2 more times.

If the overall range of the scores resulting from the cited procedure falls between 0 and 3, a match cannot be made and the samples probably were produced by two different people. If the scores fall between 7 and 10, a reasonably robust match is established. Scores between 4 and 6 are generally neutral. Incidentally, if foils are used and the mean scores polarize, the resulting confidence level is enhanced. Data are now available (Hollien 2002) about Phoneticians' abilities to make good AP-SI analyses. That is, if the task is highly structured and the auditors are well trained professionals.

Machine/Computer SI Approaches

The speaker recognition issue changes radically when efforts are made to apply modern technology to the problem. However, while it would appear that solutions are now easily possible, such is not the case. Further, most of the research effort here has gone into studying speaker verification (VI) not identification (SI). Worse yet, while some of the VI approaches are based on research (Hautamaki et al. 2010) many others (Beigi 2011) are not. Concern has been expressed about this problem (Campbell et al. 2009; Hollien/Majewski 2009).

So, where does one start with respect to computer processed SI? First, it is necessary to develop a procedure. Then, to *test* it. One of the more successful efforts here is the SAUSI (Semi-Automatic Speaker Identification) program being carried out at the University of Florida (Hollien 1990; 2002; Jiang 1995).

Our initial step was to identify and evaluate a number of parameters. It was discovered early on that traditional approaches to signal processing were inadequate, hence, speech features were adopted. This decision was supported by early experiments, the AP-SI research literature and the realization that humans use procedures of this type for everyday SI. Our research led to four vectors, each made up of a number of related parameters. They are speaking fundamental frequency or SFF (Jiang 1995; LaRiviera 1975; Wolf 1972), voice quality or LTS (Gelfer et al. 1989; Hollien/Majewski 1977; Jiang 1995), vowel formants or VFT (Jiang 1995; Kovoor et al. 2009; Wolf 1972) and temporal patterning (prosody) or TED (Gelfer et al. 1989; Jacewicz et al. 2010; Jiang 1995). A full description of them is available (Hollien 2002). Since no single SAUSI vector seemed to provide appropriately high levels of correct identification for all types of speech encountered, the four vectors were normalized, combined into a single unit

and organized as a two-dimensional continuum or profile. This procedure also addressed the forensic limitations imposed on the identification task. That is, one referent, one test sample embedded in a field of competing samples in a procedure which employs forced matches (or non-matches) within a large population. After the profile is generated, the process is replicated several times. Figure 5 shows that the system correctly "picked" U and K as the same talker (they both were the same person).

Unknown Reference	C:\SAUSI\E3
Unknown Test	C:\SAUSI\E3
Known	C:\SAUSI\E3
Foil 1	C:\SAUSI\E3
Foil 2	C:\SAUSI\E3
Foil 3	C:\SAUSI\E3
Foil 4	C:\SAUSI\E3
Foil 5	C:\SAUSI\E3
Foil 6	C:\SAUSI\E3
Foil 7	C:\SAUSI\E3
Foil 8	C:\SAUSI\E3
Foil 9	C:\SAUSI\E3
Foil 10	C:\SAUSI\E3
Foil 11	C:\SAUSI\E3

	LTS	TED	SPF	VPT	SUM
Unknown test	1.0000	1.4818	1.0000	1.0000	1.0000
Known	1.5323	1.0000	1.2836	1.2292	1.1686
Foil 1	3.8862	7.5851	3.9469	2.6156	5.1448
Foil 2	6.1144	4.4177	9.5805	3.1292	6.6102
Foil 3	9.1714	5.4474	5.2633	3.4191	6.6367
Foil 4	9.0549	5.5074	10.0000	10.0000	10.0000
Foil 5	5.5006	3.4713	7.0671	2.0946	5.2058
Foil 6	6.1824	10.0000	9.5805	4.6620	8.7621
Foil 7	9.7969	3.5349	9.5805	3.6224	7.5982
Foil 8	7.6665	8.2456	7.4505	4.4635	7.9845
Foil 9	10.0000	5.4801	9.5805	4.7024	8.5640
Foil 10	8.6318	7.4416	5.8762	4.4418	7.5553
Foil 11	4.0598	4.7911	6.3774	2.7540	5.0393

Figure 5: Printout of Normalized SAUSI Data involving an unknown talker, a known talker and 11 controls (foils).

The final continuum usually consists of the data from 2–3 replications and includes a summation of all vectors. Hence, any decision made about identity is based on several million individual comparisons (factors, parameters, vectors, rotations).

A practical illustration of how this procedure works may be found in Figure 6.

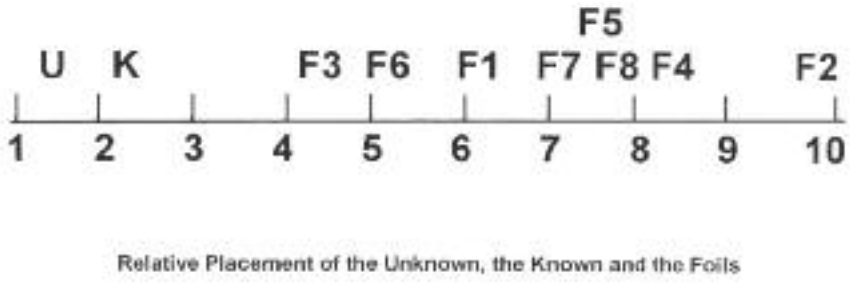


Figure 6: An example taken from a real-life investigation.

The evaluation here (a real life case) involved unknown (U) and known (K) talkers plus a number of foils (F). As can be seen, the unknown speaker (U) was found to place close to himself (a test of validity); so did the known talker (K) with the foils occupying in other parts of the continuum; hence, a match could be made. Case outcome provided (nonscientific) support for this judgment.

5 VOCAL BEHAVIORS

Many of the behaviors felt or exhibited by humans can be detected by analysis of that person's speech and voice. Included are 1) emotions (stress, anger, fear, contempt, sadness, depression, elation, happiness), 2) states induced by external conditions (ethanol intoxication, drugs), 3) certain intentional behaviors (deception, untruths, disguise, insolence, avoidance) and 4) health states (cold/flu, illness, fatigue). Indeed, there are too many of them to review here. Hence, a major issue from each of the first three categories will be described. Two are psychological stress and intoxication; a third – deception – also will be considered but more within the context of commercial devices.

Stress

Stress is a condition that denotes a negative psychological emotion. A definition: It is a psychological state which results from a response to a perceived threat and is accompanied by the specific emotions of fear and/or anxiety (Hicks/Hollien 1981; Lazarus 1993; Scherer 1981). Further, it has been shown that listeners can accurately identify stress from speech samples alone (Scherer 1981). How do they do this? First, increases in pitch or speaking fundamental frequency appear to correlate with stress increments. Second, frequency variability can do so also even though it is a less robust predictor. Third, vocal intensity is another acoustic parameter that correlates with psychological stress and, while the data here are a little "mixed", the best evidence is that it tends to increase with stress in most people. A fourth relationship is that of a tem-

poral pattern based on speech bursts – they are reduced when the speaker is stressed. Finally, an important recent finding is that increases in speaker nonfluencies also appear to correlate with stress increments. It is on these bases that a predictive model of the vocal correlates of stress was developed; it may be found in Figure 7.

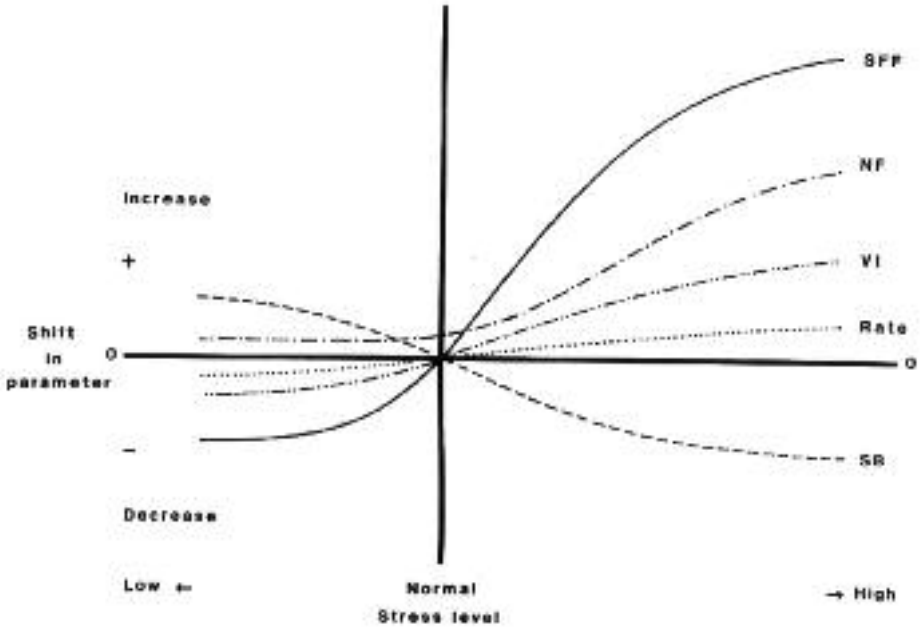


Figure 7: Model of the most common shifts in the voice and speech of individuals who are experiencing psychological stress.

In Figure 7, **SFF** is speaking fundamental frequency, **NF** is nonfluencies, **VI** is vocal intensity, and **SB** refers to the number of speech bursts per unit of time. As may be seen, changes occur for all the listed elements. Nonetheless, it should be noted that information of this type will be of greatest value when it can be contrasted with reference profiles for a person's normal speech.

Alcohol-Speech Relationships

Almost anyone who is asked to do so probably will describe the speech of an inebriated talker as "slurred", "misarticulated", or "confused". But, what are its actual correlates? Is it possible to determine a person's sobriety solely from analysis of his or her speech?

The basic rationale for an intoxication-speech interface has been established. A review of the general intoxication biometrics would appear relevant here as it will provide a foundation for understanding the speech-alcohol relationships that follow. First, it has been demonstrated that the consumption of even moderate amounts of alcohol

can result in impaired cognitive function (Arbuckle et al. 1994; Hindermarch et al. 1991) as well as reduced sensory motor performance (Abrams et al. 2004; Hill et al. 1990). Since the speech act represents the output of a number of high level integrated systems (sensory, cognitive, motor), it is legitimate to assume that this process also is susceptible to the influence of extraneous factors such as alcohol consumption (Goldstein 1992; Walgren/Barry 1970).

Prior to the work of the present author, the relationships between speech and intoxication appeared to be that: 1) speaking fundamental frequency (SFF) level was often lowered and SFF variability sometimes increased, 2) speaking rate was reduced, 3) the number and length of speech pauses often increased and 4) speaking intensity levels sometimes were lowered (Chin/Pisoni 1996; Pisoni/Martin 1989). Again, however, virtually all of these alcohol-speech relationships were quite variable. One group of investigators (Klingholz et al. 1988) attempted to account for the inconsistencies found in the literature by arguing that they resulted from inadequate and/or differing research designs. This group also observed that variation probably was due to nonobjective measurements of blood alcohol level (BAL), excessively high BAL, the use of too few (intoxicated) subjects and/or use of analyses which were only qualitative. While most of their observations undoubtedly were correct, they did not include all of the problems found in the research they reviewed. Specifically, but few of the investigators they cited controlled for drinking habits, intoxication level, increasing vs. decreasing BAL, effort, or comparisons of sober vs. intoxicated speech.

An extensive program of research was conducted at the University of Florida and a number of reports have been published (examples: Hollien et al. 2001; 2009). Since other behavioral states (stress, fatigue, emotions) could have complicated attempts to determine intoxication level from speech analysis, the investigators here structured new and innovative protocol. They were designed to induce acute alcohol intoxication under highly controlled conditions. Subjects were administered doses of 80-proof liquor mixed with both a soft drink and Gatorade while drinking at their own pace. Breath concentration levels (BrAC) were measured at 10–15 minute intervals. This approach was efficient with nausea and discomfort sharply reduced; it also permitted serial measurements which, in turn, allowed intoxication level to be accurately tracked (cf. Figure 8).

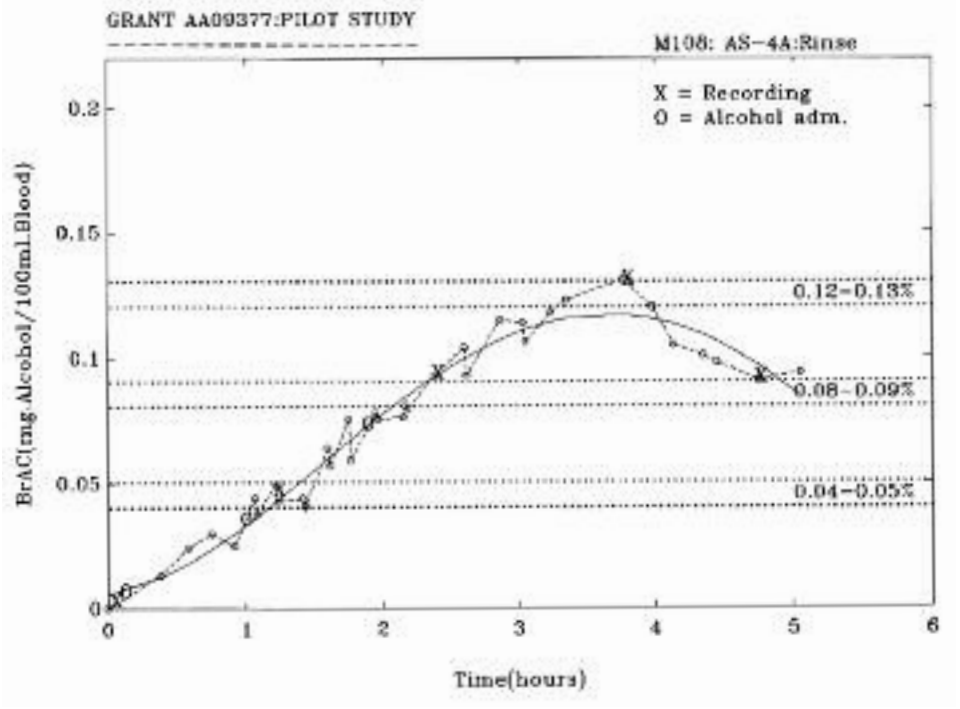


Figure 8: Changes in intoxication level as a function of ingesting alcohol.

The curve demonstrates a subject's progress relative to increasing and decreasing levels of intoxication from the beginning to the completion of the experimental portion of a trial. The speech samples were recorded when the speaker was in each of the desired "windows".

Hence, large drinker-class groups were studied with subjects participating in all procedures related to their experiment; i. e., data were taken serially at sober-to-severely intoxicated "windows".

Subjects were carefully selected on the basis of 27 behavioral and medical criteria. After training, they were required to produce four types of speech at each intoxication level: 1) oral reading and an extemporaneous passage, 2) responses to articulation and diadochokinetic tests. The many experiments completed included auditory processing by listeners (drunk-sober, intoxication level, etc.), acoustic analysis of the signal, and various classification/sorting (behavioral) tests. A large base population, plus actors and binge drinkers were studied.

A number of relationships emerged (Hollien et al. 2001; 2009). Auditors tended to overestimate speaker impairment for individuals who were only mildly (to moderately) intoxicated and, then, underestimate the involvement level of the severely intoxicated (Figure 9).

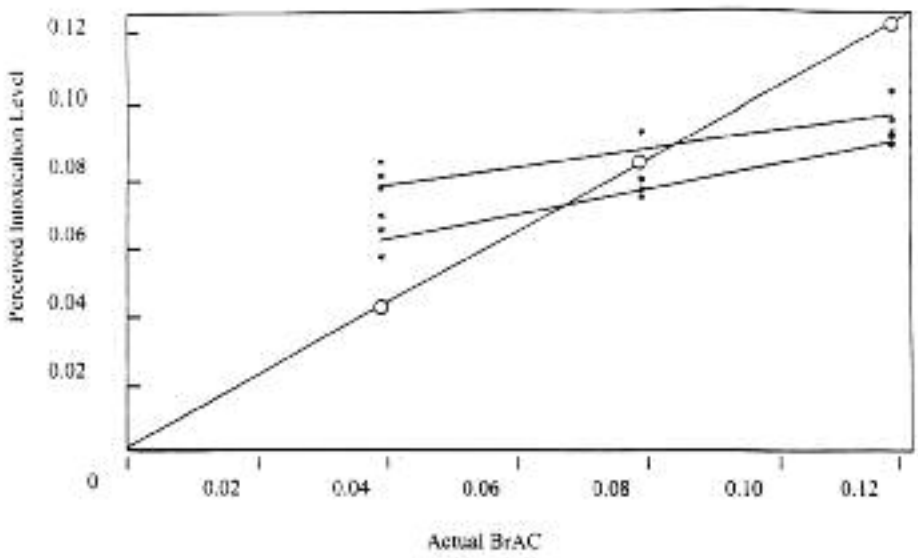


Figure 9: Perceived intoxication level contrasted to the physiologically measured levels (45 degree line with circles) from sober to severely intoxicated (BrAC 0.00 to 0.12). Four studies are combined for the top set (35 talkers, 85 listeners) and two for the lower (36 talkers, 52 listeners). Second, it appears possible to accurately mimic rather severe levels of intoxication by speakers who actually were sober and, conversely, to reduce the percept of intoxication if inebriated individuals attempt to sound sober. Moreover, there appeared to be only minor gender differences and few-to-none for drinking level (light, moderate, heavy). The voice shifts somewhat paralleled those for stress; that is excepting those for vocal intensity. More important, it was found that, as intoxication increases, speaking fundamental frequency (heard pitch) is raised, *not lowered* and speech is slowed. Finally, (as expected) a very high correlation was found between nonfluencies and intoxication level.

Deception

Relatively few of the experiments carried out in this area have been on basic speech-deception relationships; most have focused on assessing "truth machines". While some relationships have been established (Anolli/Ciceri 1997), this area is not well developed – a condition primarily due to the lack of support for basic research but, especially, because it has been dramatically overshadowed by the controversy about the voice stress/lie detecting devices. At present, a number of such devices exist (i. e., VSA, PSE, CVSA, LVA). In all cases, their manufacturers claim that they can detect stress and lying by their analysis of a subject's voice. Of course, if these devices actually were able to detect when a person was lying, they would be of inestimable value to legal, law enforcement, and intelligence agencies – as well as to all of us! For example, consider what would happen if it were possible to determine the beliefs and intent of politicians simply from hearing them speak. Moreover, there would be no

need for trials by jury as the guilt or innocence of anyone accused of a crime could be determined by simply asking them: "Did you do it?"

But, to qualify as a "lie response", the observed behavior would have to be validly measurable and every person would have to exhibit that particular response whenever they lied. In this regard, it was Lykken (1981) who seems to have best articulated the key concept here. He argues that, if lies are to be detected, there has to be some sort of a "lie response", – a measurable physiological or psychological event which always occurs when a person lies. Lykken correctly suggested that, "until a lie response has been identified and its validity and reliability have been established, no one can claim to be able to measure, detect and/or identify falsehoods on anything remotely approaching an absolute level."

But, upon what criteria or theories are the cited "voice analyzers" based? Unfortunately, it is almost impossible to answer this question as their maker's explanations are almost always quite vague. Some indicate that their devices access the micro-tremors in the laryngeal muscles. Such micro-tremors do exist in the body's long muscles (Lippold 1971) but data confirm that they do not exist in the larynx (Shipp/Izdebski 1981) and, even if they did, that they would not affect the actions of the complexly interacting respiratory, laryngeal and vocal tract motor units during speech. Other manufacturers (i. e., LVA) claim that they use an individual's thoughts or intent as their foundation – and nearly all of them appear to rely on the presence of stress. But, is stress somehow equivalent to lying in the first place? A myriad of such questions can be asked but, at present, the manufacturers offer no explanations or valid support for their claims. Furthermore, of the scores of studies (examples: Hollien et al. 1987; Horvath 1982) that have been carried out by independent investigators, nearly all refute the claims of validity.

By 2000, many professionals believed that, since the "voice analyzers" had been discredited they would disappear. Unfortunately, that was not true. At least two devices: the National Institute for Truth Verification's (NITV) Computer Voice Stress Analyzer (CVSA) and Nemesysco's Layered Voice Analyzer (LVA) were being sold in even greater numbers than previously. Thus, since neither had been the subject of extensive and comprehensive, but fair, assessment, it appeared timely to do just that. Accordingly, The U.S. Dept. of Defense requested the present author, and his associates, to test the ability of both the CVSA and LVA to identify people when they were 1) speaking the truth, 2) telling a falsehood at high jeopardy, 3) talking while highly stressed and 4) producing unstressed speech (plus combinations of these and other types of speech). Both systems were tested in large double-blind laboratory experiments which, in no instance, permitted the examiners to directly observe the on-scene events or the human subjects who were providing the speech materials. It was only through the use of these controlled approaches that their characteristics could be rigorously evaluated.

It also should be noted that these experiments were designed to compensate for prior criticism of research on products of this type. Thus, in this case, speech samples were recorded by subjects who systematically varied normal utterances plus intensely

deceptive and stressed speech. To do so, they had to hold very strong views about an issue and were required to make sharply derogatory statements about them while believing that they would be observed doing so by colleagues and friends. The stress levels (further enhanced by the addition of electric shock) were externally measured for all subjects to ensure that the deception was produced at very high levels of jeopardy. Each system was then evaluated – using a double blind protocol – by 1) a UF team trained/certified by NITV – and later, by the LVA school. A second set of teams also provided data. One was a group of three senior CVSA personnel provided by NITV and, for LVA, a pair of their senior instructors. All evaluators worked at their own pace.

The analysis results for the sets of 300 samples each were subjected to a variety of statistical procedures. All were negative with the most telling being the sensitivity measures of d' (d prime). Here, as with all others, none of the results – see Figure 10 – can be seen to even *approach* significance (i. e., the +1.00 level).

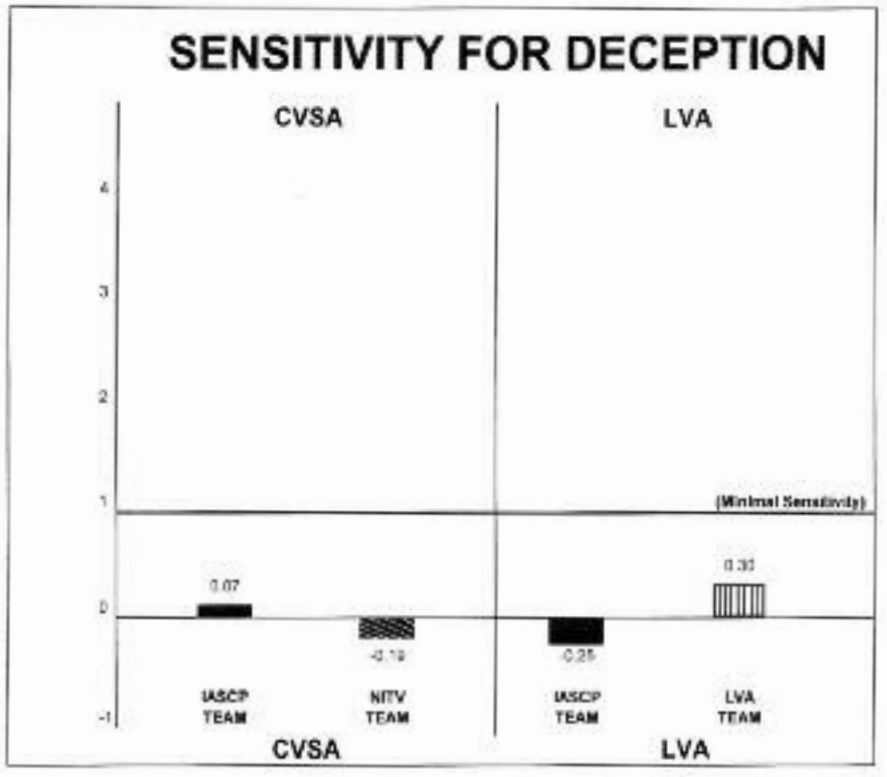


Figure 10: Sensitivity measures (d') for the two devices studied (CVSA on left and LVA on right).

In Figure 10, one assessment team (IASCP, UF) was staffed by the same two scientists; the other two teams (CVSA; LVA) were provided by the respective manufacturers. Note that all scores are well below the cutoff for *minimal* sensitivity, meaning that the devices performed only at chance levels.

These data held for all procedures, systems and teams (Harnsberger et al. 2009; Hollien et al. 2008). Other research (Damphousse et al. 2007), plus our field studies (Hollien/Harnsberger 2006), confirmed these findings. In short, it is clear that the CVSA and the LVA devices were not able to detect deception or stress; indeed, both operated only at chance levels. To conclude: the only approach practitioners can use to identify deception in speech is to employ those vectors that have been generated from *basic research*.

PROLOGUE

The discussions found in this chapter attempt to provide descriptions of the burgeoning discipline of Forensic Phonetics. As you know, you have discovered which topics clearly fall within its scope; they are: 1) the enhancement and decoding of speech on audio recordings, 2) their authentication, 3) speaker identification and 4) the detection of a number of behavioral states from voice/speech. Rapid future progress is expected.

Acknowledgements

The author thanks the U.S. Dept. of Defense (ONR, ARO, CIFA), the Institutes of Health, U.S. Justice Dept., Nat. Science Found. and the Dreyfus Foundation for their research support.

References

Aarts Nancy L. (1984) *Effects of listener stress on perceptual speaker identification*. Florida: University of Florida.

Abramos, Ben/Mark Fillmore (2004) "Alcohol induced impairment of inhibitory mechanisms involved in visual search." *Exp. Clin. Psychopharmacol.* 12, 243–250.

Anolli, Luigi/Rita Ciceri (1997) "The voice of deception: Vocal strategies of naïve and able liars." *J. Nonverbal Behav.* 21, 259–284.

Aperman, A (1982) "Examination of claims of inauthenticity in magnetic tape recording." In: R. W. de Vore/J. S. Jackson (eds), *Proceed., Carnaham Conf. Crime Contermeasures.* Lexington: KY, 63–71.

Arbuckle, Tannis/June Chaikelson/Dolores Gold (1994) "Social drinking and cognitive function revisited." *J. Studies Alcohol* 55, 352–361.

Beigi, Homayoon (2011) *Fundamentals of speaker recognition.* Secausus (NJ): Springer.

Brixen, Eddy B. (2007) "Techniques for the authentication of digital audio recordings." *Proc., Audio Engineer. Soc., 122nd Convention*, Vienna,

Broeders, A.P.A. (1996) "Earwitness identification: Common ground, disputed territory and uncharted areas." *Forensic Linguistics* 3, 1–13.

Bronkhorst, Adelbert W. (2000) "The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions." *Acustica* 86, 117–128.

Campbell, Joseph/Wade Shen/William Campbell/Reva Schwartz/Jean-François Bonastre/Driss Matroof (2009) "Forensic speaker recognition." *IEEE Signal Process Mag.* 2009/3, 95–103.

Chin, Steven B./David Pisoni (1997) *Alcohol and Speech*. San Diego: Academic Press.

Cross, John F./Jane Cross/James Daly (1971) "Sex, race, age and beauty as factors in recognition of faces." *Perception and Psychophysics* 10, 393–396.

Dallasarra, Cassy/Aericka Dunn/Shanna White/Al Yonovitz/Joe Herbert (2010) "Speaker identification: Effects of noise, telephone bandwidth, and word count on accuracy." *J. Acoust. Soc. Am.* 128, 2393.

Damphousse, Kelly R./Laura Pointond/Deidre Upchurch/Rebecca K. Moore (2007) "Assessing the validity of voice stress analysis tools in a jail setting." *Report No. 219031*. U.S. Department of Justice.

Dean, J.D. (1980) "The work of the home office tape laboratory." *Police Research Bull.* 35, 25–27.

Gelfer, Marylou P./K. P. Massey/Harry Hollien (1989) "The effects of sample duration and timing on speaker identification accuracy by means of long-term spectra." *J. Phonet.* 17, 327–338.

Goldstein, Dora (1992) "Pharmacokinetics of alcohol." In: J. H. Mendelson/N. K. Mello (eds), *Medical Diagnosis and Treatment of Alcoholism*. New York: McGraw-Hill, 24–54.

Harnsberger, James D./Harry Hollien/Camilo A. Martin/Kevin A. Hollien (2009) "Stress and deception in speech: evaluating layered voice analysis." *J. Forensic Sciences* 54, 642–650.

Hautamäki, Ville/Toni Kinnunen/Mohaddeseh Nosratighods/Kong-Aik LEE/Bin MA/Haizhou LI (2010) "Approaching human listener accuracy with modern speaker verification."*INTERSPEECH-2010*, 1473–1476.

Hicks, James W. Jr./Harry Hollien (1981) "The reflection of stress in voice-1: understanding basic correlates." In: R. W. de Vore/J. S. Jackson (eds), *Proceed. Carnahan Conf. Crime Countermeasures*. Lexington: KY, 189–194.

Hill, John/Glenda Toffolon (1990) "Effect of alcohol on sensory and sensorimotor visual functions." *J. Stud. Alcohol.* 51, 108–113.

Hindermarch, Ian/J. S. Kerr/N. Sherwood (1991) "The effects of alcohol and other drugs on psychmotor performance and cognitive function." *Alcohol and Alcoholism* 26, 71–79.

Hollien, Harry (1977) "Authenticating tape recordings." In: R. W. de Vore/J. S. Jackson (eds), *Proceed. Carnahan Conf. Crime Countermeasures*. Lexington: KY, 19–23.

Hollien, Harry (1990) *Acoustics of Crime*. New York: Plenum.

Hollien, Harry (1992) "Noisy tape recordings in forensics." *ESCA Proceed., Speech Processing in Adverse Conditions*, Nice, 167–170.

Hollien, Harry (1996) "Consideration of guidelines for earwitness lineups," *Forensic Linguistics* 3, 14–23.

Hollien, Harry (2002) *Forensic Voice Identification*. London: Academic Press Forensics.

Hollien, Harry (2008). "Forensic Phonetics." In: C. H. Wecht (ed), *The Forensic Sciences.* New York: Matthew Bender Co., 1–149.

Hollien, Harry/G T Bennett/Marylou P. Gelfer (1983) "Criminal identification comparison: aural/visual identifications resulting from a simulated crime." *J. Forensic Sciences* 28, 208–221.

Hollien, Harry/Gea De Jong/Camilo A. Martin/R Schwartz/K Liljegren (2001) "Effects of ethanol intoxication on speech suprasegmentals." *J. Acoust. Soc. Amer.* 110, 3198–3206.

Hollien, Harry/Laura Geisson/James W. Hicks Jr. (1987) "Voice Stress Evaluators and Lie Detection." *J. Forensic Sci.* 32, 405–418.

Hollien, Harry/James D. Harnsberger (2006) "Voice stress analyzer instrumentation evaluation." *Final Report. US Dept. Defense. Contract: FA-4814–04-0011.* Florida: CIFA.

Hollien, Harry/James D. Harnsberger (2010) "Speaker identification: the case for speech vector analysis." *J. Acoust. Soc. Amer.* 128, 2394(A).

Hollien, Harry/James D. Harnsberger/Camilo A Martin/Rebecca Hill/Alan G. Alderman (2009) "Perceiving the effects of ethanol intoxication on voice." *J. Voice* 23, 552–559.

Hollien, Harry/James D. Harnsberger/Camilo A. Martin/Kevin A. Hollien (2008) "Evaluation of the CVSA voice stress analyzer." *J. Forensic Sciences* 53, 183–193.

Hollien, Harry/Wojciech Majewski (1977) "Speaker identification by long-term spectra under normal and distorted speech conditions." *J. Acoust. Soc. Amer.* 62, 975–980.

Hollien, Harry/Wojciech Majewski (2009) "Unintended consequences: due to lack of standards for speaker identification and other forensic procedures." *Proceed. Sixteenth Internat. Congress Sound/Vibration 866*, Krakow, 1–6.

Hollien, Harry/Wojciech Majewski/Thomas E. Doherty (1982) "Perceptual identification of voices under normal, stressed and disguised speaking conditions." *J. Phonetics* 10, 139–148.

Hollien, Patricia A. (1984) "An update on speech decoding." *Proceed. Inst. Acoustics, Part I: Police Applic. Speech, Tape Record. Analysi*s, 33–40.

Horvath Frank (1982) "Detecting deception: the promise and the reality of voice stress analysis." *J Forensic Sci.* 27, 340–351.

Jacewicz, Ewa/Robert A. Fox/Lai Wei (2010) "Between-speaker and within-speaker variation in speech tempo of American English." *J. Acoust. Soc. Amer.* 128, 839–850.

Jiang, M (1995) "Experiments on a speaker identification system." Florida: University of Florida.

Klingholz, F/R Penning/E Liebhardt (1988) "Recognition of low-level alcohol intoxication from the speech signal." *J. Acoust. Soc. Amer.* 84, 929 - 935.

Kovoor, Binsu C./M H Surpiya/K P Jacob (2009) "Parametric study on speaker identification biometric system using formant analysis." *J. Acoust. Soc. Amer.* 125, 2530–2530.

Lariviere, Conrad L. (1975) "Contributions of fundamental frequency and formant frequencies to speaker identification." *Phonetica* 31, 185–197.

Lazarus, Richard S. (1993) "From psychological stress to the emotions – a history of changing outlooks." *Ann Rev Psychol.* 44, 1–21.

Lee, Ki Yong (1998) "Speech enhancement based on neutral predictive hidden markov models." *Signal Process.* 65, 373–381.

Lippold, Olof (1971) "Physiological tremor." *Scientific American* 224, 65–73.

Loftus, Elizabeth (1979) *Eyewitness Testimony*. New York: Cambridge University Press.
Lykken, David (1981) *A Tremor in the Blood*. New York: McGraw-Hill Inc.
Mayor, D./Eeva Komulainen (1989) *Subjective voice identification*. Calgary: Calgary Police Service.
Nolan, Francis (2003) "A recent voice parade." *Forensic Linguistics* 10, 277–291.
Nolan, Francis/Kristi Mcdougall/Toby Hudson (2008) "Voice similarity and the effect of the telephone: a study of the implications for earwitness evidence (VoiceSim)." *Final Report RES-000–22–2582, ESRC Swindon, UK.*
Pisoni, David B./Christopher S. Martin (1989) "Effects of alcohol on the acoustic-phonetic properties of speech: perceptual and acoustic analyses." *Alcoholism: Clinical Exper. Res.* 13, 577–587.
Scherer, Katalin R. (1981) "Vocal indicators of stress." In: J. Darby (ed), *Speech Evaluation in Psychiatry.* New York: Grune and Stratton, 171–187.
Schuster B. (2007) "Police lineups: making eyewitness identification more reliable." *NIJ Journal* 258, 1–6.
Shipp, Thomas/Krzysztof Izdebski (1981) "Current evidence for the existence of laryngeal macrotremor and microtremor." *J. Forensic Sciences* 26, 501–505.
Wallgren, Henrik/Herbert Barry (1970) *Actions of Alcohol.* Amsterdam: Elsevier.
Wan, Eric A. (1998) "Removal of noise from speech using the dual EKF algorithm." *Proceed. IEEE, ICASSP*, Piscataway, NY, CH36181, 381–384.
Wells, Gary (1993) "What do we know about eyewitness identification?" *Amer. Psychol.* 48, 553–571.
Wells, Gary/Elizabeth Olsson (2003) "Eyewitness testimony." *Ann. Rev. of Psych.* 54, 277–295.
Wolf, Jared (1972) "Efficient acoustic parameters for speaker recognition." *J. Acoust. Soc. Amer.* 51, 2044–2055.
Yarmey, Daniel A. (1995) "Earwitness speaker identification." *Psychol. Public Policy Law 1*, 792–816.

Abstract

ABOUT FORENSIC PHONETICS

This article sets forth the goals and content of Forensic Phonetics and its *major* elements. Considered are 1) the processing and analysis of spoken utterances, 2) enhancement of speech intelligibility (re: surveillance and other recordings), 3) authentication of recordings, 4) speaker identification, and 5) detection of deception, intoxication, and emotions in speech. Stress in speech, and the psychological stress evaluation systems that some individuals attempt to use as lie detectors also will be considered.

Keywords: forensic phonetics, speaker identification, detection of deception, emotions in speech.

Povzetek
FORENZIČNA FONETIKA

Prispevek opredeljuje cilje in vsebino forenzične fonetike in njenih ključnih komponent. To pomeni, da obravnava postopke 1) obdelave in analize govornih izjav, 2) izboljševanja razumljivosti govora na zvočnih posnetkih, 3) preverjanja pristnosti zvočnih posnetkov, 4) identifikacije govorcev ter 5) ugotavljanja prevar, intoksikacije in čustev v govoru. Navaja tudi različne evalvacijske sisteme za ugotavljanje stresa v govoru, ki jih ponekod uporabljajo za detektiranje laži v govoru.

Ključne besede: forenzična fonetika, identifikacija govorcev, laž v govoru, čustva v govoru.